

Article

Early Bearing Fault Evolution Modeling with a Dynamics-Constrained Graph Neural Network

Enxu Liu, Zhan Wang , Rui Zhao, Shiyu Xing * and Zinan Wang * 

School of Mechanical Engineering, Shenyang Jianzhu University, Shenyang 110168, China; 13020319741@163.com (E.L.); juven1126@163.com (Z.W.); zzz1138326976@163.com (R.Z.)

* Correspondence: xsy900717@163.com (S.X.); wzn591672678@126.com (Z.W.)

Abstract

During the incipient localized-fault stage, impact signatures in ball bearings are weak. Accurate degradation labels are difficult to construct. Purely data-driven models often lack physical consistency. To address these issues, a dynamics-constrained graph neural network (DC-GNN) was developed for early fault evolution modeling in ball bearings. Firstly, a bearing dynamic model was established based on nonlinear ball–raceway contact mechanics. A fault evolution parameter was then introduced. The incipient localized fault was represented by contact stiffness degradation. It was also described through increased local additional displacement. This formulation captured the continuous transition from a healthy condition to slight localized damage. Secondly, the bearing was represented as a graph structure. The inner ring, outer ring and balls were defined as nodes. Ball–raceway contacts were treated as edges. Contact deformation, contact stiffness, contact force, and a fault-region indicator were embedded into node or edge features. A DC-GNN was then constructed. Finally, dynamic simulations were performed under different fault evolution levels. Bearing vibration data were also used for validation. The results show that contact-force fluctuations intensified as the fault evolution parameter increased. The time-domain impact amplitude also increased. Fault characteristic components became more prominent in the envelope spectrum. The degradation indicator exhibited a consistent evolution trend. These variations agreed well with incipient localized fault progression. This study provides an effective approach for early fault evolution characterization in rolling bearings.

Keywords: data-driven models; bearing dynamic model; DC-GNN; early fault

1. Introduction

Rolling bearings are critical support components in rotating machinery. Their operating condition directly affects the safety and reliability of equipment systems [1]. During long-term service, bearing raceways and balls are susceptible to cyclic loads, unstable lubrication, assembly errors, or impact loading. These factors can induce microcracks, pitting, or localized spalling [2,3]. Compared with advanced faults, incipient localized defects usually exhibit a small damage size, weak impact energy, low characteristic-frequency amplitudes, and high sensitivity to noise. Conventional time-domain, frequency-domain, or time-frequency methods often produce unstable features during weak-fault identification [4]. Therefore, accurately characterizing the transition from a healthy state to incipient localized damage is essential for early warning degradation prediction.

From a mechanistic perspective, incipient localized faults in bearings essentially result from coupled changes in surface geometry, local load-carrying capacity, and contact state



Academic Editor: Kambiz Vafai

Received: 12 August 2026

Revised: 5 September 2026

Accepted: 9 September 2026

Published: 11 September 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.

at the ball–raceway interface [5]. Under healthy conditions, periodic nonlinear contact occurs between each ball and the raceway. The resulting contact deformation and force depend on radial load, radial clearance, and contact angle. They are also strongly governed by ball position, cage motion, and inner-ring speed [6]. According to Hertzian contact theory, contact force exhibits a pronounced nonlinear dependence on local elastic deformation [7]. As each ball enters or leaves the load zone, the load distribution among balls changes periodically with time. The contact state also varies, providing an important internal excitation for bearing vibration [8]. After a localized defect forms, a ball passing through the defect zone typically undergoes local unloading, defect entry, edge impact, recontact, and load recovery. These processes induce transient fluctuations in contact deformation, equivalent contact stiffness, and contact force. They further excite structural resonance, periodic impacts, and amplitude–frequency modulation [9]. Previous studies have developed bearing dynamic models with different degrees of freedom [10]. These models are based on nonlinear dynamic equations. Localized defects have been represented in several forms [11]. Common representations include local geometric depressions and time-varying displacement excitation. Time-varying contact stiffness or equivalent impact forces have also been used. Defect width and depth strongly affect collision intensity [12]. Edge geometry also plays an important role. The defect position within the load zone further influences collisions during ball entry or exit. Radial load, speed, and clearance alter load distribution [13]. They also affect contact duration and the intrinsic properties of the system. These changes further influence impact amplitude, characteristic frequencies, and nonlinear response patterns. These studies have provided an important basis for explaining the vibration mechanism of localized faults. However, most existing models treat a defect as a fixed geometric structure with prescribed width and depth. In practice, an incipient fault does not instantly develop from a healthy state into pronounced spalling. It usually progresses through several continuous stages. These stages include microcrack initiation, crack propagation, local material removal, reduced contact capacity, and slight spalling. Therefore, a continuous fault evolution variable with clear physical meaning is required. This variable can bridge the gap between fixed-defect models and actual progressive degradation.

In recent years, deep learning methods have been widely applied to bearing fault diagnosis, health assessment, and remaining useful life prediction. Common methods include Convolutional Neural Network (CNN) [14], Long Short-Term Memory (LSTM) [15], Transformer [16], transfer learning [17], and autoencoders [18]. These studies have greatly improved automatic feature extraction from complex vibration signals. However, most models still use an end-to-end framework. They directly map input signals to fault classes, health indicators or remaining useful life. In these frameworks, a bearing is treated as a single signal source [19]. The contact topology among the inner ring, outer ring, and balls is rarely modeled explicitly. Local load distribution and fault propagation paths also receive limited attention. When training samples are limited, degradation labels are often constructed from empirical indicators. Test conditions may also deviate from the training distribution. Under these conditions, models tend to rely on statistical correlations in the data. The extracted features may therefore lack physical consistency with changes in contact state [20]. For a ball bearing, the inner ring, outer ring, and individual balls can be represented as graph nodes. Contacts between balls and raceways can be represented as graph edges. Contact deformation, contact stiffness, contact force, and fault-zone state can serve as node or edge features. This graph representation explicitly captures internal structural relations and load-transfer paths [21]. However, a graph structure alone cannot ensure dynamic equilibrium. It also cannot guarantee compliance with nonlinear contact constitutive laws. Pure dynamic models rely on idealized parameters and boundary

assumptions. They cannot fully reproduce assembly errors, structural transmission, or random noise in real operation. Purely data-driven models lack physical constraints. Their predictions may therefore conflict with actual contact states. A unified graph neural network (GNN) is still needed. It should integrate continuous early fault evolution, bearing contact topology and nonlinear dynamic constraints. This integration is essential for improving generalization, physical consistency, and interpretability.

Therefore, an early bearing fault evolution method based on a dynamics-constrained graph neural network (DC-GNN) is proposed. Firstly, a bearing dynamic model is established from the nonlinear ball–raceway contact mechanism. Secondly, a fault evolution variable is introduced. An incipient outer-ring fault is represented by reduced contact stiffness and increased local additional displacement. This representation describes the continuous transition from a healthy state to slight localized damage. The bearing system is then represented as a graph. Its nodes include the inner ring, outer ring, and balls. Ball–raceway contacts form the graph edges. A DC-GNN is further constructed. The loss function includes data error and dynamic-equation residuals. It also contains a Hertzian contact-force constraint and a fault-evolution smoothness constraint. These constraints allow the model to learn vibration degradation features while obeying bearing contact dynamics. The proposed method is validated using dynamic simulations and experimental bearing vibration data. Its effectiveness is evaluated through contact force, time-domain responses, envelope spectra, phase portraits, and shaft-center orbits. Degradation indicators and health-indicator predictions are also analyzed.

2. Dynamic Model of Early Failure Bearing

2.1. Dynamic Model of the Healthy Bearing

A ball bearing is mainly composed of the inner ring, outer ring, balls, and cage. Its structure is shown in Figure 1. During operation, complex nonlinear contact occurs between the balls and raceways. The bearing vibration response is mainly governed by contact deformation, radial load, radial clearance, and ball position. To highlight the influence of an incipient localized fault on the bearing dynamic response, the following assumptions are made:

- (1) The outer ring is fixed, while the inner ring is rotated with the shaft.
- (2) Hertzian contact is assumed between each ball and the raceways.
- (3) Complex factors are neglected, including temperature rise and severe lubrication fluctuations. Intense cage impacts are also ignored.

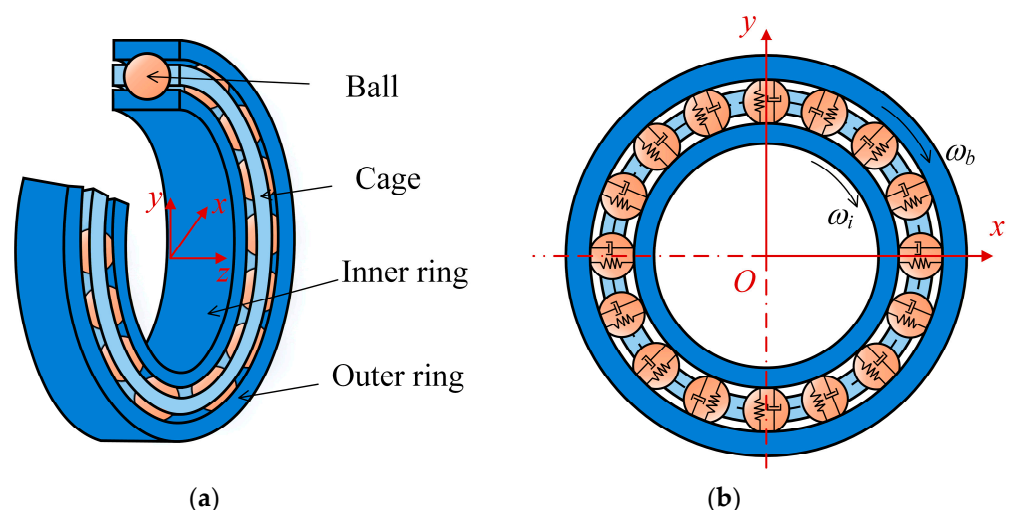


Figure 1. Bearing structure: (a) three-dimensional structure; (b) radial cross-sectional diagram.

Let the number of balls in the bearing be denoted by N . The angular position of the z -th ball at time t can be expressed as [22]

$$\theta_z = \omega_c t + \frac{2\pi(z-1)}{N} + \theta_0 \quad (1)$$

where θ_0 denotes the initial angular position of the z -th ball. The cage angular velocity is denoted by ω_c and can be expressed as

$$\omega_c = \frac{1}{2} \left[\omega_i \left(1 - \frac{D_b}{D_m} \cos \alpha \right) \right] \quad (2)$$

where ω_i denotes the inner ring angular velocity. The ball diameter and pitch diameter are denoted by D_b and D_m , respectively. The contact angle is denoted by α . Therefore, the contact deformation of the z -th ball can be expressed as

$$\delta_z = (x_i - x_o) \cos \theta_z + (y_i - y_o) \sin \theta_z - c_r \quad (3)$$

where x_i and y_i denote the inner ring displacements in the x and y directions, respectively. The outer ring displacements are denoted by x_o and y_o . The initial bearing clearance is denoted by c_r . According to Hertzian contact theory, the contact force between the z -th ball and the raceway can be expressed as [23]

$$Q_z = \lambda K \delta_z^{1.5} \quad (4)$$

where K denotes the equivalent contact stiffness of the bearing. The contact-state coefficient λ determines whether contact occurs between the ball and the raceway. It can be expressed as

$$\lambda = \begin{cases} 1, & \delta_z > 0 \\ 0, & \delta_z \leq 0 \end{cases} \quad (5)$$

A nonlinear relationship exists between the bearing contact force and contact deformation. Contact force is generated only when a ball is under compressive contact. This nonlinear contact behavior is a major source of the vibration response of bearings. Considering the equivalent mass, damping, radial load, and eccentricity of the system, the bearing dynamic equations can be written as

$$\begin{cases} m\ddot{x} + c\dot{x} + \sum_{z=1}^N Q_z \cos \theta_z = F_x + F_e \cos \omega_i t \\ m\ddot{y} + c\dot{y} + \sum_{z=1}^N Q_z \sin \theta_z = F_y + F_e \sin \omega_i t - mg \end{cases} \quad (6)$$

where m denotes the equivalent mass of the bearing, and c denotes the equivalent damping coefficient. The external radial load components in the x and y directions are denoted by F_x and F_y , respectively. The system eccentric force is denoted by F_e .

2.2. Characterization of Incipient Fault Evolution

To characterize the evolution of incipient localized faults in ball bearings, a fault evolution variable H is introduced to quantify the progression of a localized fault from a healthy state to a slight spalling state. The variable satisfies $0 \leq H \leq 1$, where $H = 0$ indicates a healthy bearing and $H = 1$ indicates a bearing with slight localized spalling. During incipient fault evolution, changes in the local contact state are mainly reflected in two aspects. Firstly, the contact stiffness within the defect zone gradually degrades. Secondly, an additional local displacement is introduced when a ball passes through the

defect zone. Accordingly, the fault evolution variable is incorporated into both the contact stiffness and the additional local displacement induced by the defect.

Considering the gradual reduction in contact stiffness caused by an incipient localized fault, the equivalent Hertzian contact stiffness is expressed as

$$K_H = K_0(1 - \xi H) \quad (7)$$

where K_0 denotes the equivalent Hertzian contact stiffness in the healthy state. Since bearings of the same type were used in this study, the bearing material was not treated as an independent variable, and its elastic effect on ball–raceway contact was implicitly represented by the equivalent Hertzian contact stiffness K_0 . ξ denotes the contact stiffness degradation coefficient. Considering the additional displacement generated when a ball passes through the localized defect zone, the defect-induced additional displacement is expressed as

$$h_H = h_0 + \beta H \quad (8)$$

where h_0 denotes the initial local additional displacement, and β denotes the growth coefficient of the defect-induced additional displacement. Therefore, the increase in defect depth is represented by the evolution of h_H , which directly modifies the ball–raceway contact deformation and the resulting Hertzian contact force.

The crack propagation process of the outer ring defect is illustrated in Figure 2. Let θ_f denote the angular position of the center of the defect, φ_f denote the angular span of the defect zone, and ω_b denote the angular velocity of the ball. Accordingly, the time-varying displacement excitation can be expressed as

$$h_H(t) = \begin{cases} h_H, & |\theta_z - \theta_f| \leq \frac{\varphi_f}{2} \\ 0, & \text{others} \end{cases} \quad (9)$$

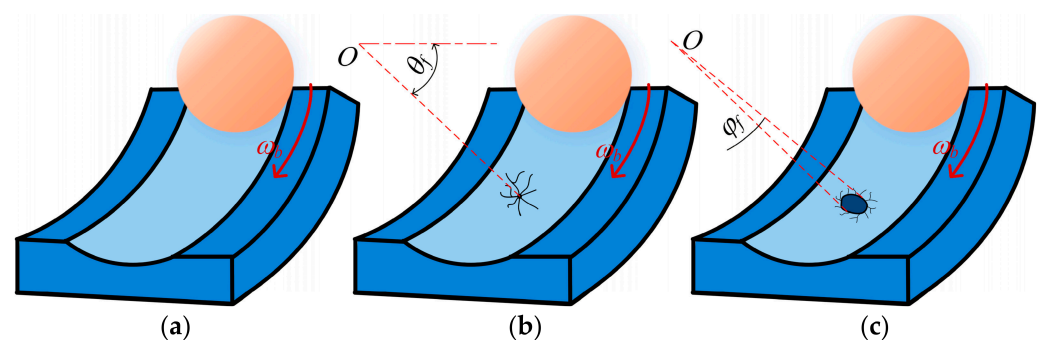


Figure 2. Raceway crack propagation process: (a) healthy raceway; (b) weak crack; (c) minor fault.

The angular span φ_f characterizes the circumferential extent of the localized defect. In the present model, the detailed edge profile of the defect is simplified and is not independently varied. Therefore, the model considers the equivalent defect depth and extent. After accounting for the defect-induced displacement, the equivalent contact deformation at the z -th ball can be expressed as

$$\delta_z^f = x \cos \theta_z + y \sin \theta_z - c_r - h_H(t) \quad (10)$$

The local additional displacement affects the contact deformation only when a ball passes through the fault zone. This effect induces a transient fluctuation in the contact

force. According to the Hertzian contact relationship, the contact force of the z -th ball, with incipient fault evolution taken into account, can be expressed as

$$Q_z^f = \lambda K_H \left(\delta_z^f \right)^{3/2} \quad (11)$$

3. Dynamics-Constrained Graph Neural Network

GNNs can model systems with explicit topological relationships and are therefore well suited to describing the internal coupling within a bearing. Accordingly, the bearing system is represented as a graph. The inner ring, outer ring and balls are treated as nodes. The ball–raceway contact relationships are defined as edges. Dynamic information, including contact deformation, contact stiffness, contact force, and the fault evolution variable, is incorporated into the graph structure. On this basis, a dynamics-constrained loss function is constructed using the bearing dynamic equations and the Hertzian contact relationship. The DC-GNN can thus learn data-driven features while satisfying the governing laws of bearing contact dynamics.

3.1. Construction of the Bearing Graph Structure

Conventional neural networks typically take vibration signals directly as input and extract signal features using convolution, recurrent, or attention mechanisms. However, these approaches generally ignore the internal structural relationships within the bearing, making it difficult to explicitly characterize the contact coupling among the inner ring, outer ring, and balls. To address this limitation, the bearing system is represented as a graph, in which the bearing components and their contact relationships are incorporated into the neural network model. The bearing graph can be expressed as

$$G = (V, E) \quad (12)$$

where V and E denote the node set and edge set, respectively. Since the present study focuses on the contact dynamics between the balls and raceways, the inner ring, outer ring, and individual balls are defined as graph nodes. Accordingly, the node set is expressed as

$$V = v_i, v_o, v_1, v_2, \dots, v_N \quad (13)$$

where v_i and v_o denote the inner ring and outer ring nodes, respectively, while $v_1, v_2, \dots, v_N$ denote the ball nodes. The edge set is used to represent the contact relationships between the balls and the raceways. The contact edge between the z -th ball and the inner ring is denoted by $e_{z,i}$, while the contact edge between the z -th ball and the outer ring is denoted by $e_{z,o}$. Accordingly, the edge set can be expressed as

$$E = e_{z,i}, e_{z,o} \mid z = 1, 2, \dots, N \quad (14)$$

For the inner ring and outer ring nodes, the displacement, velocity, and acceleration responses of the system are selected as node features. For each ball node, the angular position, contact deformation, and contact state are used as node features. Accordingly, the node feature matrix can be expressed as

$$X_z = [\theta_z, \delta_z, Q_z] \quad (15)$$

Edge features characterize the contact relationships between the balls and raceways. Based on the bearing dynamic model, contact stiffness, contact deformation, contact force,

and the fault-zone indicator are selected as edge features. For the contact edge between the z -th ball and the outer ring, the edge feature vector can be expressed as

$$A_{z,o} = [\delta_z, K_H, Q_z, S_z] \quad (16)$$

where S_z denotes the fault-zone indicator and can be expressed as

$$S_z = \begin{cases} 1, & |\theta_z - \theta_f| \leq \frac{\varphi_f}{2} \\ 0, & \text{others} \end{cases} \quad (17)$$

For the z -th ball node, the state update depends not only on its own motion state but also on the features of its contact edges with the raceways. The message-passing process can be simplified as

$$\begin{cases} m_z = \phi_e(X_z, X_i, A_{z,i}) + \phi_e(X_z, X_o, A_{z,o}) \\ X_{z,new} = \phi_v(X_z, m_z) \end{cases} \quad (18)$$

where m_z denotes the contact information aggregated by the z -th ball node; ϕ_e is implemented as a gated MLP, in which a sigmoid gate adaptively weights the transmitted contact message; ϕ_v is implemented using a GRUCell followed by layer normalization. The hidden feature dimension is 128, and three message-passing layers are employed, and $X_{z,new}$ denotes the updated feature vector of the z -th ball node. At the output layer, a readout function is used to aggregate the features of all nodes and edges, yielding a global representation of the bearing system state. The readout process can be expressed as

$$g = \phi_r(X_z^{new}, A_z) \quad (19)$$

where g denotes the graph-level representation. The readout function ϕ_r combines attention, mean, and maximum pooling of the updated node features, followed by the MLP to obtain the final graph representation. The pooling operations are invariant to the ordering of the nodes.

3.2. Design of the Dynamics-Constrained Loss Function

To improve the physical consistency and interpretability of the model, the bearing dynamic equations, Hertzian contact relationship, and fault evolution smoothness are incorporated into the loss function to construct a DC-GNN. The DC-GNN takes the bearing graph-structured samples as input and outputs the predicted vibration response and fault evolution variable. Accordingly, the overall loss function consists of the data loss, dynamic equation residual loss, contact force constraint loss, and fault evolution smoothness loss, and is expressed as

$$L = L_{data} + \lambda_1 L_{dyn} + \lambda_2 L_{contact} + \lambda_3 L_{smooth} \quad (20)$$

where L_{data} denotes the data error loss, L_{dyn} denotes the dynamic equation residual loss, $L_{contact}$ denotes the contact force constraint loss, and L_{smooth} denotes the fault evolution smoothness loss. The coefficients λ_1 , λ_2 , and λ_3 are weighting factors that regulate the contribution of each constraint term during model training. The data error loss constrains the discrepancy between the model output and the ground-truth samples. When the model is used to predict the bearing vibration response, the mean squared error can be adopted as the data loss

$$L_{sig} = \frac{1}{M} \sum_{j=1}^M \|\hat{a}_j - a_j\|^2 \quad (21)$$

where M denotes the number of samples, $\hat{a}_j$ denotes the predicted vibration response of the j -th sample, and a_j denotes the corresponding measured or simulated vibration response. When the model output is the fault evolution variable, its prediction loss can be expressed as

$$L_{deg} = \frac{1}{M} \sum_{j=1}^M |\hat{H}_j - H_j|^2 \quad (22)$$

where $\hat{H}_j$ denotes the predicted fault evolution variable for the j -th sample, and H_j denotes the corresponding degradation level of the sample. Accordingly, the data error loss is expressed as

$$L_{data} = L_{sig} + \lambda_H L_{deg} \quad (23)$$

where λ_H denotes the weighting coefficient associated with the fault evolution variable. To ensure that the model predictions satisfy the dynamic equilibrium of the bearing system, the dynamic equation residual loss is introduced. Based on the nonlinear bearing dynamic model, the predicted displacement, velocity, acceleration, or equivalent state variables obtained from the network are used to construct the dynamic residual, which is expressed as

$$\begin{cases} R_x = m\hat{\ddot{x}} + c\hat{\dot{x}} + \sum_{z=1}^N Q_z^f \cos \theta_z - F_x - F_e \cos \omega_i t \\ R_y = m\hat{\ddot{y}} + c\hat{\dot{y}} + \sum_{z=1}^N Q_z^f \sin \theta_z + mg - F_y - F_e \sin \omega_i t \end{cases} \quad (24)$$

The corresponding dynamic equation residual loss is expressed as

$$L_{dyn} = \frac{1}{M} \sum_{j=1}^M (|R_{x,j}|^2 + |R_{y,j}|^2) \quad (25)$$

To enforce the contact relationship between the balls and the raceways, the Hertzian contact force constraint loss is introduced

$$L_{contact} = \frac{1}{M} \sum_{j=1}^M \sum_{z=1}^N |\hat{Q}_{z,j} - Q_{z,j}^f|^2 \quad (26)$$

Because incipient fault evolution is generally continuous and progressive, the fault evolution variable should not exhibit abrupt changes between adjacent samples. Accordingly, a fault evolution smoothness loss is introduced

$$L_{smooth} = \frac{1}{M-1} \sum_{j=1}^{M-1} |\hat{H}_{j+1} - \hat{H}_j|^2 \quad (27)$$

Because the individual loss terms have different numerical scales, normalization was performed before constructing the overall loss. The weighting coefficients were determined using the validation subset of Bearing1_3. A one-factor-at-a-time sensitivity analysis was conducted by varying each coefficient from 0.25 to 4 times its reference value. Considering both prediction accuracy and physical consistency, the final coefficients were selected as $\lambda_H = 127.633$, $\lambda_1 = 0.252$, $\lambda_2 = 3.297$, and $\lambda_3 = 4000$. Representative sensitivity results are summarized in Table 1, while Table 2 reports the relative magnitudes of the weighted loss components during training.

Table 1. Sensitivity analysis of the loss-weighting coefficients.

Parameter	Factor	Weight	H-RMSE	Signal NRMSE	Dynamic Residual	Contact Residual	Smooth
λ_H	0.25	7.977	0.386	1.050	1.235	0.471	0.0104
	1	31.908	0.274	1.049	1.035	0.317	0.0107
	4	127.633	0.236	1.045	0.963	0.227	0.0098
λ_1	0.25	0.252	0.268	1.037	0.948	0.200	0.0111
	1	1.008	0.274	1.049	1.035	0.317	0.0107
	4	4.034	0.318	1.042	1.234	0.521	0.0081
λ_2	0.25	0.206	0.290	1.061	1.311	0.605	0.0075
	1	0.824	0.274	1.049	1.035	0.317	0.0107
	4	3.297	0.271	1.054	0.950	0.208	0.0121
λ_3	0.25	250	0.278	1.053	1.330	0.773	0.0070
	1	1000	0.274	1.049	1.035	0.317	0.0107
	4	4000	0.338	1.030	0.949	0.203	0.0122

Table 2. Relative magnitudes of the weighted loss components during training.

Stage	L_{sig}	$\lambda_H L_{deg}$	$\lambda_1 L_{dyn}$	$\lambda_2 L_{contact}$	$\lambda_3 L_{smooth}$
Initial	26.0%	20.9%	6.6%	33.0%	13.6%
Intermediate	53.9%	7.9%	13.3%	8.1%	16.8%
Converged	53.9%	7.6%	13.9%	8.2%	16.4%

The results indicate a trade-off between prediction accuracy and physical consistency. At convergence, the data-related terms accounted for approximately 61.5% of the total loss, whereas the three physics-related constraints retained approximately 38.5%, confirming that the physical constraints remained active during model optimization.

The bearing system is represented as a graph in which the inner ring, outer ring, and balls are treated as nodes, while the contact relationships between the balls and the raceways are modeled as edges. Contact deformation, contact stiffness, contact force, and the fault-zone indicator are incorporated as key graph features. The overall training procedure is illustrated in Figure 3. This graph representation explicitly characterizes the internal contact topology and load transmission paths within the bearing. The main implementation and training settings of the DC-GNN are summarized in Table 3.

Table 3. Main implementation and training settings of the DC-GNN.

Item	Setting	Item	Setting
Physical ball–raceway contacts	16	Directed message-passing edges	32
Hidden dimension	128	Message-passing steps	3
Edge update	Gated MLP	Node update	GRUCell + LayerNorm
Graph readout	Attention + mean + max pooling	Temporal encoder	2-layer bidirectional GRU
Attention heads	4	Dropout	0.12
Optimizer	AdamW	Gradient clipping	5.0
Fine-tuning epochs	1200	Simulation-pretraining epochs	700
Pretraining learning rate	1.0×10^{-3}	Fine-tuning learning rate	6.0×10^{-4}

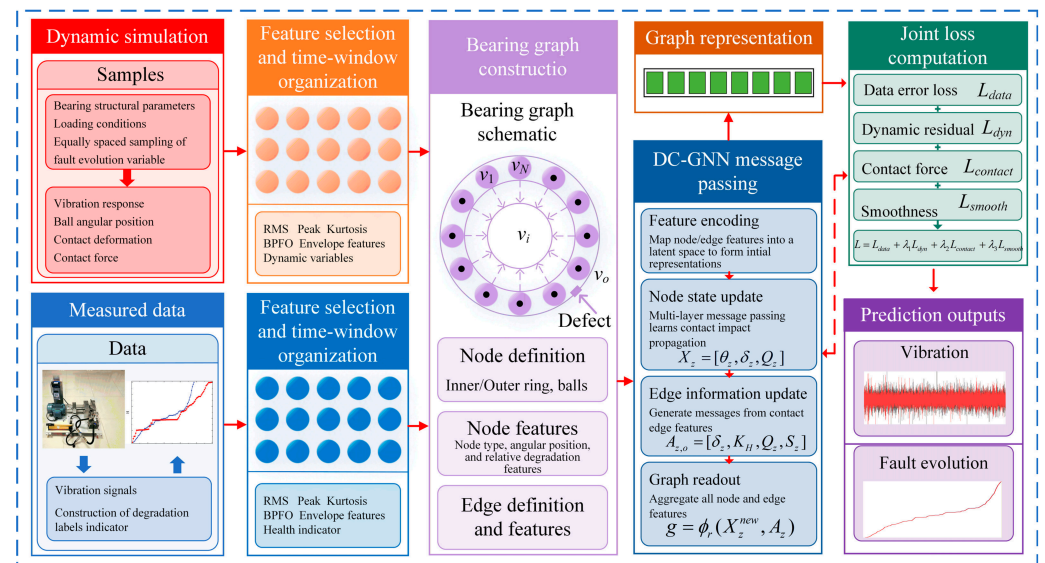


Figure 3. Training procedure of the DC-GNN.

4. Result Analysis

The detailed computational procedure is illustrated in Figure 4. During model training, dynamic simulation samples are first generated based on the bearing structural parameters, loading conditions, and different fault evolution variables. These samples provide the vibration responses, ball angular positions, contact deformations, contact forces, and other state information corresponding to different degradation levels. The inner ring, outer ring, and balls are then represented as graph nodes, while the ball–raceway contact relationships are modeled as graph edges. Contact stiffness, contact force, contact deformation, and the fault-zone indicator are incorporated as graph features and fed into the model. During training, the DC-GNN updates the node and edge features through message passing and employs a readout function to obtain a graph-level state representation of the bearing system. This representation is subsequently used to predict either the vibration response or the fault evolution variable. During backpropagation, the model parameters are updated not only according to the data error loss but also, under the joint constraints of the dynamic equation residual, Hertzian contact force relationship and fault evolution smoothness. Through iterative optimization, the overall loss gradually decreases, enabling the model to learn the contact coupling relationships among the balls, inner ring, and outer ring and to predict the evolution state of incipient faults. In practical condition monitoring, an accelerometer can be installed on the bearing housing to acquire vibration signals at predefined sampling intervals. For each newly acquired vibration record, wear-sensitive features are extracted and compared with the healthy-state baseline. These features are then mapped to the corresponding bearing graph. For a given bearing, the contact topology among the inner ring, outer ring, and balls remains unchanged. Therefore, only the state-related node and edge features need to be updated for each new vibration record. The trained DC-GNN can then directly estimate the current fault evolution variable H . The resulting H sequence can be used to track the bearing degradation trend. It can also provide a basis for condition warning and maintenance decisions.

To validate the effectiveness of the proposed dynamic model for incipient localized faults in ball bearings, numerical simulations of the vibration responses at different fault evolution levels were conducted based on the established nonlinear dynamic equations. The LDK UER204 bearing was considered, and its detailed parameters are listed in Table 4. During the simulations, the fault evolution variable was used to characterize the progression of an outer ring localized fault from slight degradation to slight spalling.

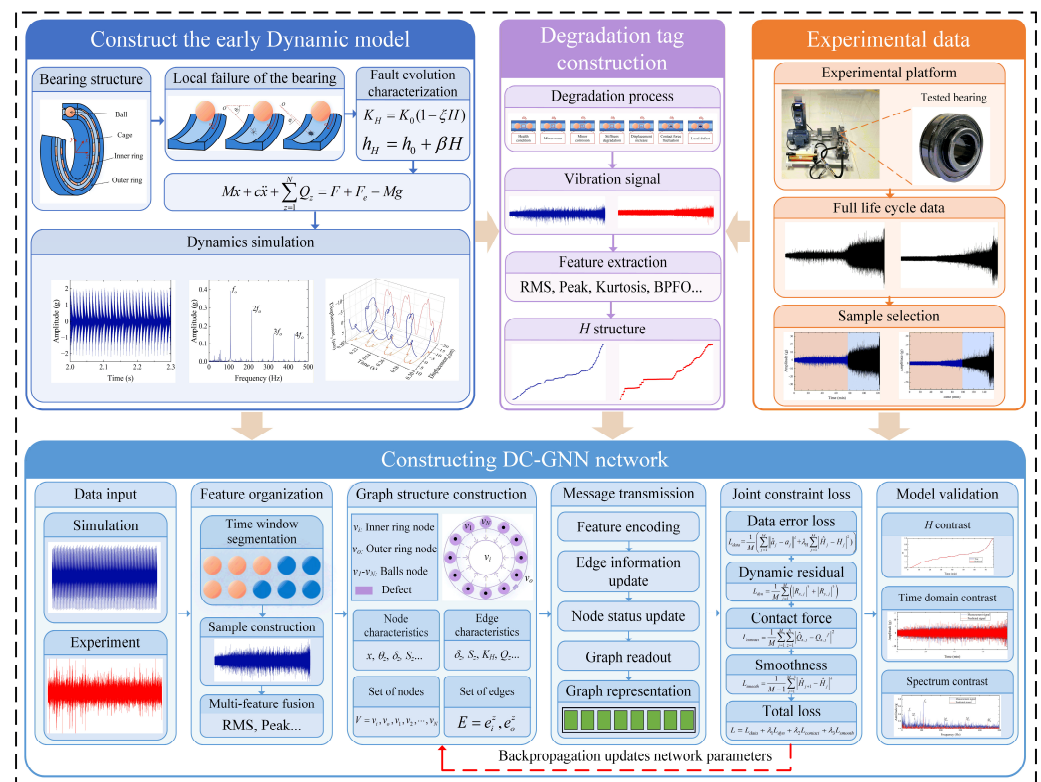


Figure 4. Computational flowchart.

Table 4. Parameters of the LDK UER204 bearing.

Parameters	Value	Unit
Inner raceway diameter	29.30	mm
Outer raceway diameter	39.80	mm
Bearing pitch diameter	34.55	mm
Basic rated dynamic load	12,820	N
Diameter of ball	7.92	mm
Number of balls	8	-
Initial contact angle	0	deg
Basic rated static load	6.65	kN
Speed	2100	r/min
Radial force	12	kN

4.1. Vibration Analysis of Incipient Fault Evolution

To investigate the influence of incipient localized fault evolution on the bearing vibration response, three representative fault evolution states were selected for comparison. Comparing these states provides a clear illustration of the changes in the dynamic response as the bearing progresses from slight degradation to slight localized spalling.

Figure 5 presents the contact force between a ball and the outer raceway under three different fault evolution levels. After the ball enters the load zone, the contact force exhibits a pronounced periodic loading–unloading pattern. Under the healthy or slightly degraded condition, the contact force curve remains relatively smooth, indicating that the ball–raceway interface is dominated by normal Hertzian contact. As the fault evolution variable increases, the local additional displacement reduces the contact deformation when the ball passes through the outer ring fault zone, resulting in localized attenuation or a slight depression near the peak of the pure contact force.

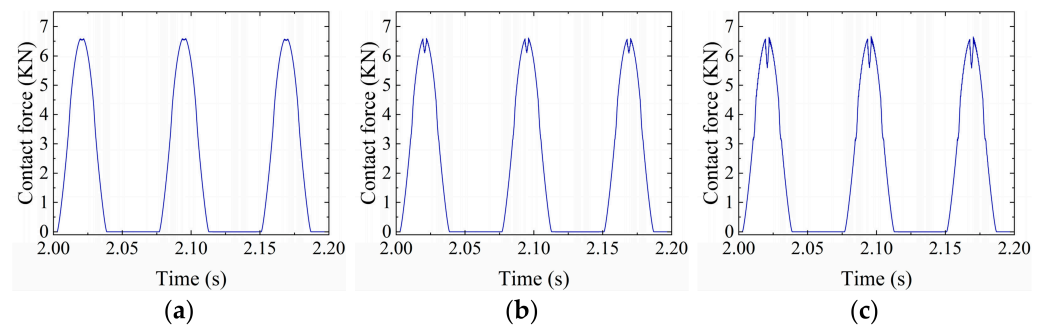


Figure 5. Contact force under three different fault evolution levels: (a) 0.1; (b) 0.5; (c) 1.

Figure 6 presents the time-domain responses, envelope spectra, phase portraits, and Poincaré sections under three different fault evolution levels. The time-domain responses show that the impact amplitude of the vibration acceleration gradually increases with H . During the slight degradation stage, periodic impact components are already present. However, the overall amplitude remains low, and the impact intervals are relatively stable. This indicates that the local defect has not yet strongly disturbed the overall system response. As the fault progresses, sharp impacts become more pronounced. The waveform amplitude increases significantly, and the high-frequency oscillatory components become stronger. This behavior results from the periodic passage of the balls through the outer ring fault zone. The reduction in local contact stiffness and the increase in additional displacement cause short-duration fluctuations in the contact force. These fluctuations further excite the high-frequency vibration response. The envelope spectra show that the outer ring fault characteristic frequency f_o and its harmonics can be observed under all fault evolution states. As H increases, the amplitudes of the characteristic frequency and its harmonics increase consistently. This indicates that the periodicity of the fault-induced impacts becomes more pronounced.

Figure 7 shows the shaft orbits under three different fault evolution levels. As the fault evolution variable increases, not only does the vibration amplitude increase, but the spatial motion pattern also changes. At a low fault evolution level, the shaft orbit remains relatively regular and confined within a small region. This indicates that the system response is primarily governed by the steady radial load and the periodic passage of the balls through the load zone. As H increases, the shaft orbit becomes progressively more complex. Local fluctuations and irregular distortions appear along the orbit boundary. These results indicate that fault-induced impacts cause the shaft motion to deviate from simple periodic behavior and exhibit stronger nonlinear modulation characteristics.

Figure 8 presents the degradation indicators under different fault evolution levels. These indicators further confirm the trends described above. The peak and root mean square (RMS) values generally increase with H . This indicates a progressive increase in fault-induced impact energy. The peak value is highly sensitive to localized impacts. It therefore directly reflects the increase in impact intensity. The RMS value represents the overall vibration energy level. Its continuous increase with H indicates that fault evolution enhances both transient impacts and the overall vibration energy. Unlike the peak and RMS values, kurtosis first increases and then decreases. At the early fault stage, the impacts are sparse and impulsive, leading to an increase in kurtosis. As the fault progresses, the impacts become more frequent. The signal energy is then distributed over a broader range, reducing the prominence of individual peaks and causing kurtosis to decrease.

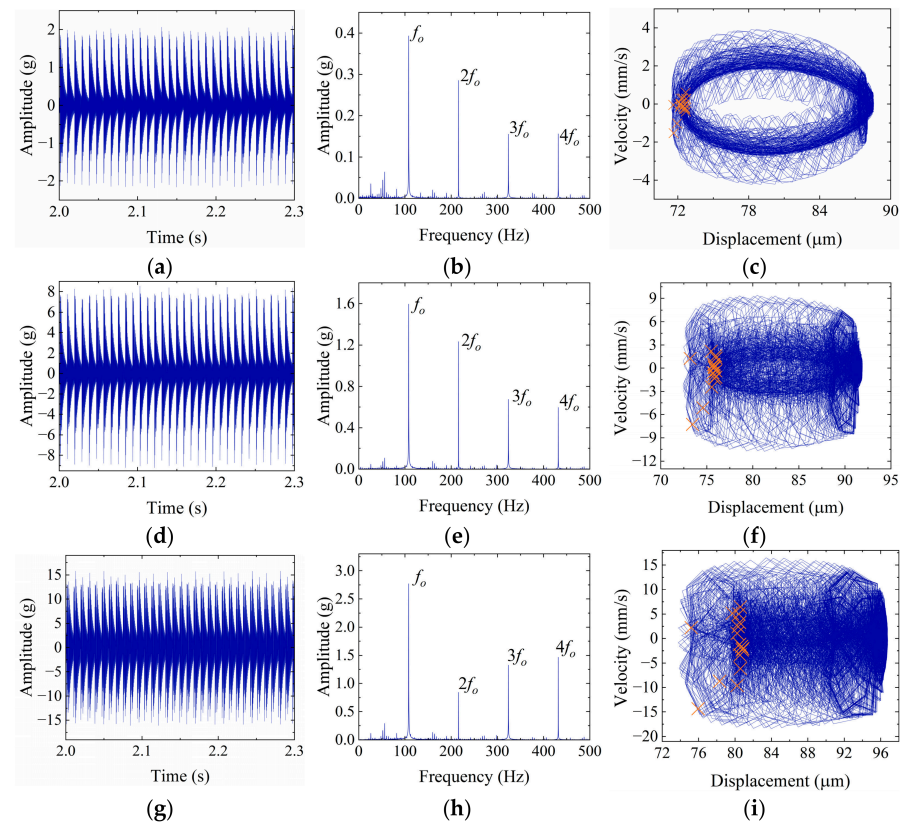


Figure 6. Time-domain responses, envelope spectra and phase portraits under three fault evolution levels: (a–c), 0.1; (d–f), 0.5; (g–i), 1. (a) Time domain. (b) Spectrum envelope. (c) Phase diagram and Poincaré map. (d) Time domain. (e) Spectrum envelope. (f) Phase diagram and Poincaré map. (g) Time domain. (h) Spectrum envelope. (i) Phase diagram and Poincaré map.

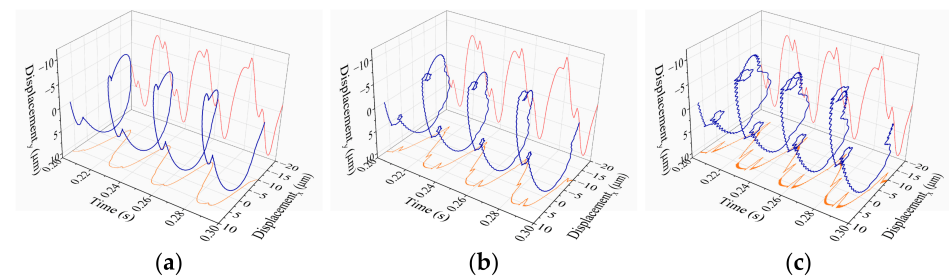


Figure 7. Shaft orbits under three fault evolution levels: (a) 0.1; (b) 0.5; (c) 1.

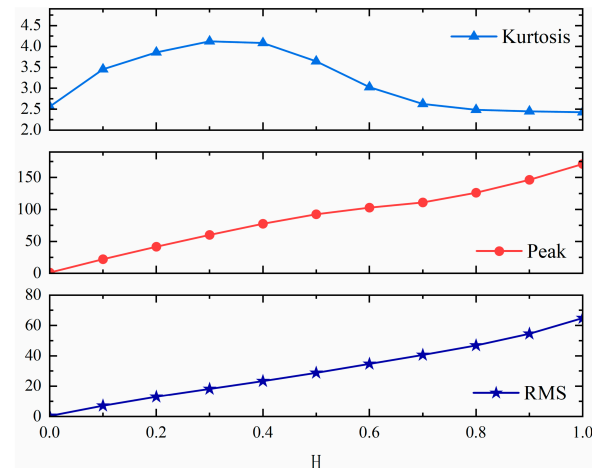


Figure 8. Degradation indicators under different fault evolution levels.

4.2. Prediction Results' Analysis

To validate the proposed model using real degradation data, the XJTU-SY bearing dataset [24] was employed for prediction and comparative analysis. The experimental test rig is shown in Figure 9. The full-life vibration signals of Bearing1_1 and Bearing1_3 were analyzed in the time domain. The results are presented in Figure 10. For both bearings, the vibration amplitudes remain low during the initial operating stage. The signals are dominated by weak random vibration components, indicating that the bearings are in a healthy or slightly degraded state. As the operating time increases, the vibration amplitude gradually rises. Pronounced impact responses and sudden amplitude changes appear during the later stage of operation. Bearing1_1 enters a clear degradation stage at approximately 78 min, whereas Bearing1_3 does so at approximately 99 min. Therefore, 78 min for Bearing1_1 and 99 min for Bearing1_3 are regarded as the onset of incipient localized faults. The operating-life information is reflected by the chronologically ordered vibration sequences during the bearing degradation process. In this study, operating time was used to characterize the temporal evolution of the fault state rather than as an independent input for predicting absolute fatigue life or remaining useful life.

Because the XJTU-SY dataset does not provide directly measured defect-size labels, an equivalent reference health indicator H was constructed from the measured vibration signals. For each vibration record, six degradation-sensitive features, including RMS, peak amplitude, kurtosis, envelope RMS, high-frequency energy ratio, and BPFO envelope amplitude, were extracted. The features were positively standardized relative to the healthy baseline, and fused using correlation-based weights. The resulting composite degradation sequence was smoothed using a five-point moving average, denoted by $\tilde{D}_i$. The final equivalent health indicator was defined as

$$H_i = \frac{\sum_{k=2}^i \max(\tilde{D}_k - \tilde{D}_{k-1}, 0)}{\sum_{k=2}^M \max(\tilde{D}_k - \tilde{D}_{k-1}, 0)} \quad (28)$$

where M is the total number of vibration records. The constructed H is a relative equivalent degradation indicator rather than a directly measured physical defect size.

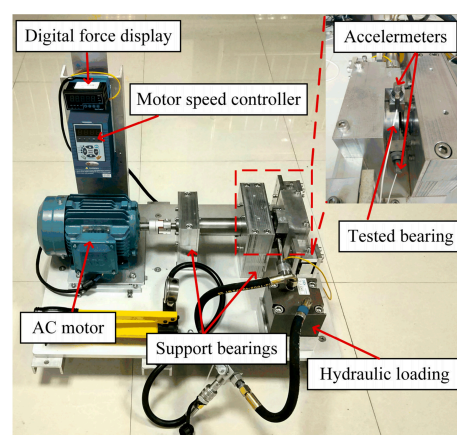


Figure 9. Experimental test rig for the XJTU-SY bearing dataset.

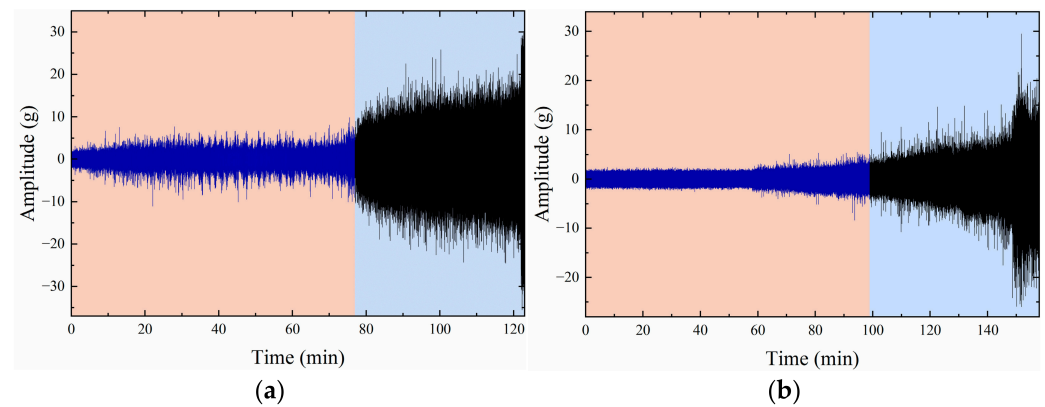


Figure 10. Selection of data samples: (a) Bearing1_1; (b) Bearing1_3.

The corresponding evolution results are shown in Figure 11. It can be seen that the value of H generally increases with operating time, indicating that the constructed health indicator can effectively characterize the gradual evolution of the bearing from the healthy state to the faulty state. During the initial stage, the value of H remains low and changes slowly, indicating that the fault characteristics are not yet obvious. During the intermediate stage, H exhibits a stepwise increase, suggesting that the local contact condition begins to change and the vibration impact components gradually become stronger. During the final stage, H rapidly approaches 1, indicating that the bearing has entered a pronounced degradation state.

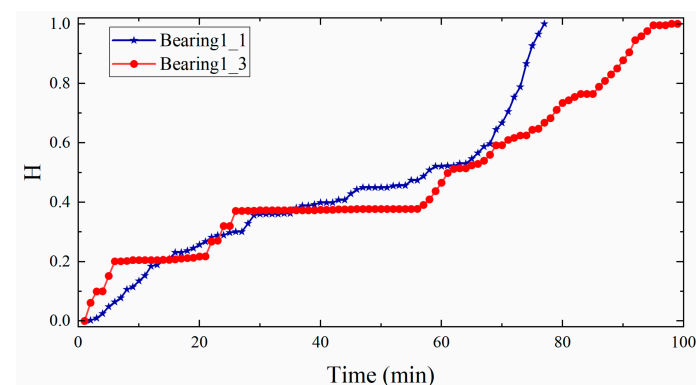


Figure 11. H-construction diagram.

Figure 12 compares the predicted fault-evolution variable with the constructed reference health indicator H for the independent Bearing1_1 test sequence. More specifically, during the initial and gradual degradation stages, the difference between the predicted and reference H values is generally below 0.01. During the transition stage of approximately 60–70 min, both curves increase from approximately H from 0.5 to 0.65, with only small local deviations. In the final rapid-degradation stage, the predicted H reproduces the sharp increase in the reference indicator and approaches 1.0 at the end of the selected sequence. A slightly larger deviation occurs locally during the rapid rising stage, but the maximum absolute error remains below approximately 0.005.

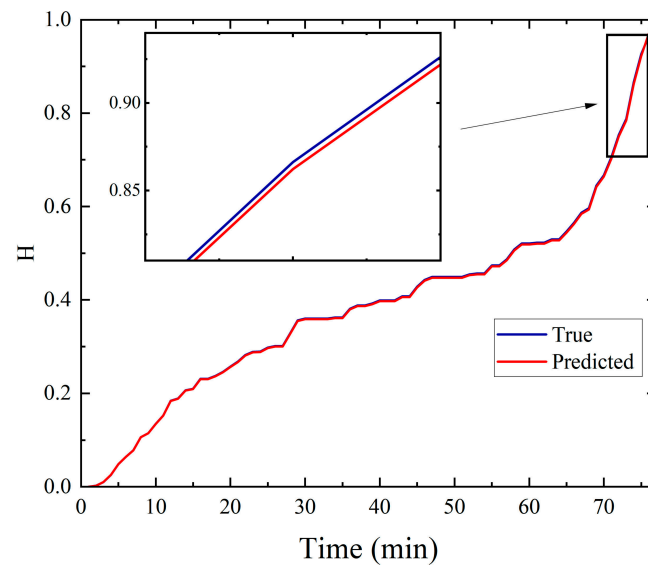


Figure 12. Comparison between the predicted and the constructed H .

Figure 13 further evaluates whether the predicted vibration response preserves the principal temporal and fault-related spectral characteristics of the measured signal. In the time domain, both signals exhibit an increasing vibration-amplitude range with operating time. The predicted signal reproduces the overall increase in vibration energy, although differences remain in several instantaneous impacts because random impact amplitudes and background disturbances are not fully represented by the model. In the envelope spectrum, the calculated outer race fault characteristic frequency f_o is 107.9 Hz. The f_o frequency of the measured signal is 108.6 Hz, while that of the predicted signal is 109.2 Hz. The difference between them is only 0.6 Hz. The measured signal contains stronger broadband components and more local spectral peaks, whereas the predicted response mainly preserves the dominant fault-related components. Therefore, the model reproduces the characteristic-frequency locations and their harmonic structure, while some discrepancies remain in spectral amplitude because of stochastic noise, structural transmission, and operating disturbances.

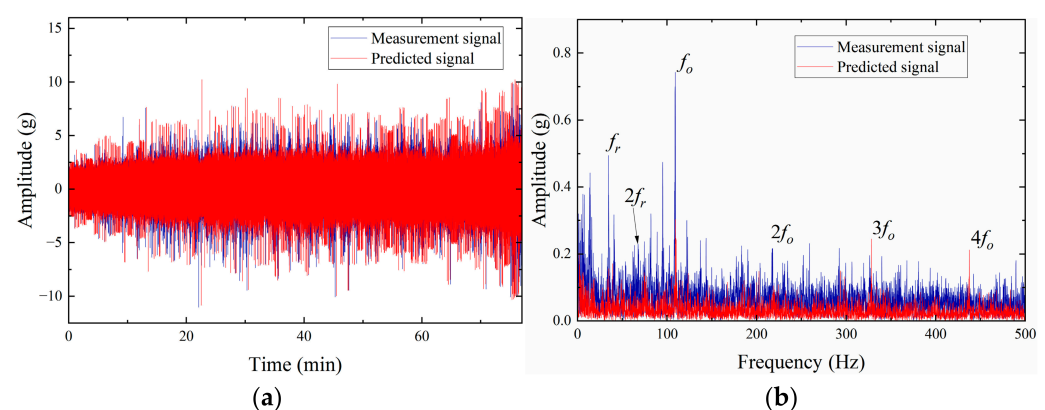


Figure 13. Comparison between the predicted and measured vibration signals: (a) time-domain waveforms; (b) spectrum envelope.

Figure 14 further evaluates whether the DC-GNN predicted vibration responses preserve degradation-sensitive statistical characteristics. The RMS values extracted from both the measured and predicted signals increase progressively with H , indicating a consistent evolution of vibration energy. Similarly, the peak amplitudes exhibit an overall increasing tendency, reflecting the enhancement of fault-induced impacts during degradation. Al-

though local discrepancies occur because of random impacts and stochastic disturbances in the measured signals, the predicted responses reproduce the principal degradation dependent evolution of both features.

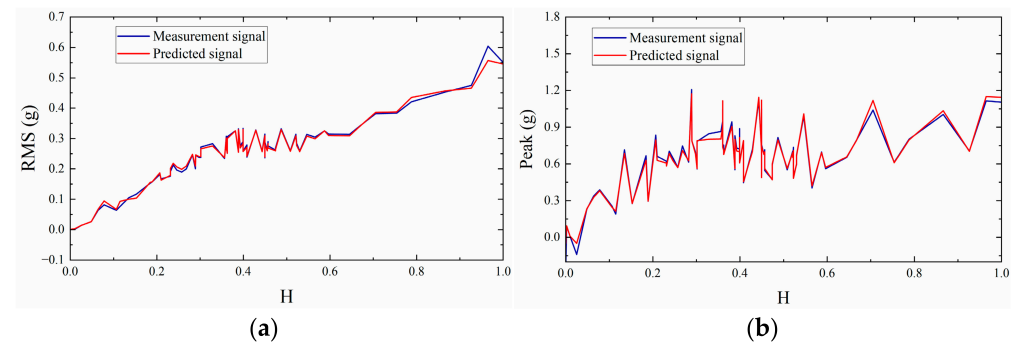


Figure 14. Comparison of the feature values extracted from the predicted and measured signals: (a) RMS; (b) peak.

To determine whether the above predictive performance is specifically attributable to the proposed DC-GNN architecture, rather than to a general data-driven prediction capability, additional comparisons were conducted with LSTM, Transformer, and GNN models under the same training and testing conditions. Figure 15 presents the quantitative comparison of their vibration-response prediction performance. The proposed DC-GNN exhibits the best and most balanced performance among the four models, with the Mean Absolute Error (MAE) of 0.155, Root Mean Square Error (RMSE) of 0.183, and Maximum Absolute Error (MaxAE) of 0.428, while $1 - R^2 = 0.424$ and $1 - \rho_s = 0.322$. In comparison, the GNN yields the MAE/RMSE/MaxAE of approximately 0.245/0.299/0.574, the LSTM yields 0.309/0.342/0.487, and the Transformer yields 0.195/0.234/0.460. Thus, relative to the GNN, LSTM, and Transformer, the DC-GNN reduces MAE by approximately 36.7%, 49.8%, and 20.5%, respectively, while simultaneously achieving the highest R^2 and Spearman correlation. These results indicate that, although the Transformer can reduce the average prediction error compared with conventional GNN and LSTM, its relatively low rank correlation suggests insufficient preservation of the degradation evolution trend. In contrast, the DC-GNN achieves consistently superior performance in both prediction accuracy and temporal trend consistency, demonstrating that incorporating bearing contact topology together with nonlinear dynamic constraints provides a more reliable representation of early fault evolution.

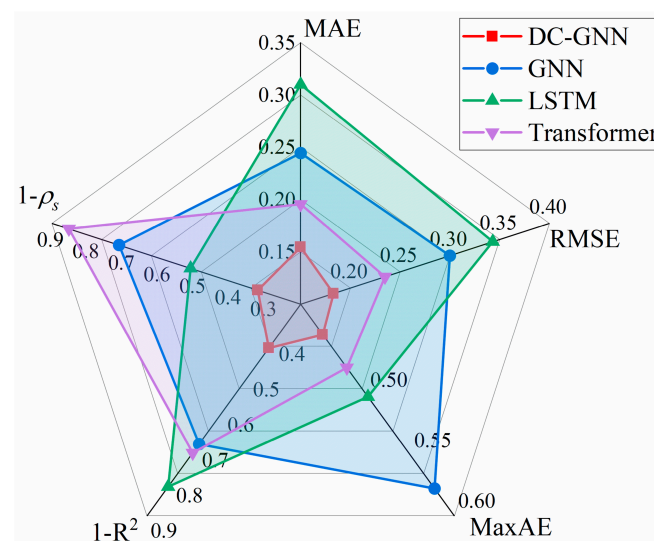


Figure 15. Method comparison.

5. Conclusions

To address the weak characteristics of incipient localized faults, the difficulty in characterizing the degradation process, and the limited physical consistency of purely data-driven models, the DC-GNN was proposed for incipient bearing fault evolution modeling. The contact dynamics model, fault evolution variable, and GNN were integrated to jointly model the bearing structural topology, nonlinear contact mechanisms, and measured degradation information. The main conclusions are summarized as follows:

- (1) By introducing the fault evolution variable, the localized fault was represented as contact stiffness degradation and an increase in the additional local displacement. This enabled the progressive evolution of contact deformation, contact force, and impact responses as the balls passed through the defect zone to be characterized. The simulation results show that increasing H led to stronger local contact force fluctuations and larger impact amplitudes in the time domain. The outer ring fault characteristic frequency and its harmonic components also became more pronounced in the envelope spectrum.
- (2) A graph-based representation was constructed for the ball bearing system. The inner ring, outer ring, and balls were defined as graph nodes, while the ball–raceway contact relationships were represented as graph edges. Contact deformation, contact stiffness, contact force, and the fault-zone indicator were incorporated as key graph features. This representation enabled the model to explicitly capture the contact coupling among the internal bearing components.
- (3) Validation using the XJTU-SY full-life dataset confirmed the performance of the proposed DC-GNN. The measured and predicted outer ring fault frequencies were 108.6 and 109.2 Hz, differing by 0.6 Hz. DC-GNN further reduced MAE by 36.7%, 49.8%, and 20.5% compared with GNN, LSTM, and Transformer, respectively, confirming superior prediction accuracy and degradation-trend consistency.

From an engineering perspective, excessive loads and impact loading should be avoided, while appropriate lubrication, radial clearance, and bearing alignment should be maintained to reduce local contact stress and uneven ball loading. During condition monitoring, a persistent increase in H , together with increasing RMS, peak amplitude, and BPFO-related components, should prompt inspection of the lubrication condition, bearing clearance, and raceway condition to prevent further development of incipient damage.

It should be noted that the present experimental validation was conducted using bearings of the same type under a fixed operating condition of 2100 r/min and 12 kN. Therefore, the generalization capability of the proposed method under variable-speed and variable-load conditions has not yet been fully investigated. Future work will incorporate rotational speed and load as additional operating variables, update the rotational frequency and BPFO-related features according to the measured speed, and further validate the proposed method under multiple operating conditions and long-term practical monitoring.

Author Contributions: E.L. oversaw the project as the team leader, provided strategic guidance on research direction, and supervised the progress of the study. Z.W. (Zinan Wang) developed the overall framework and wrote the manuscript. R.Z. conducted in-depth analysis with data investigation and optimized the methodology. S.X. contributed to data processing, indicator calculations, and visualization. Z.W. (Zhan Wang) proposed the research framework and technical roadmap. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Liaoning Provincial Science and Technology Plan Joint Project, grant number: 2025110045-JH2/1018; and Special Project for Science and Technology Talents of Shenyang City, grant number: RC240175.

Institutional Review Board Statement: Not applicable. This study did not involve humans or animals and therefore did not require ethical approval.

Informed Consent Statement: Not applicable. This study did not involve human participants.

Data Availability Statement: The XJTU-SY bearing dataset is publicly available at https://drive.google.com/open?id=1_ycmG46PARiykt82ShfnFfyQsaXv3_VK (accessed on 18 June 2026). Further in-quiries can be directed to the corresponding author.

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

Abbreviations

The following abbreviations are used in this manuscript:

DC-GNN	Dynamics-Constrained Graph Neural Network
GNN	Graph Neural Network
RMS	Root Mean Square

References

1. Hu, J.; Shi, J.; Huang, W.; Zhu, Z. Characteristic analysis of rolling bearing dynamic response under raceway multiple faults. *Measurement* **2025**, *256*, 118108. [\[CrossRef\]](#)
2. Zhao, J.; He, D.; Jin, Z.; Zhang, X.; Zhou, J. A new method for bearing remaining useful life prediction based on dynamic wavelet and physical information constraints. *Expert Syst. Appl.* **2026**, *296*, 129023. [\[CrossRef\]](#)
3. Liu, Y.; Yan, C.; Shi, J.; Kang, J.; Liu, B.; Wu, L. Nonlinear dynamic analysis of supporting bearing in rotor-bearing system considering vibration interaction. *Nonlinear Dynam* **2025**, *113*, 22435–22458. [\[CrossRef\]](#)
4. Jia, W.; Shi, H.; Dong, Z.; Zhang, X. A rolling bearing fault diagnosis framework under variable working conditions considers dynamic feature extraction. *Eng. Appl. Artif. Intel.* **2025**, *146*, 110255. [\[CrossRef\]](#)
5. Kucukoglu Dogan, B.; Dal, A.; Karacay, T. Modeling hybrid polymer ball bearing with elastoplastic contact model and its nonlinear dynamic response. *Arch. Appl. Mech.* **2025**, *95*, 66. [\[CrossRef\]](#)
6. Miao, H.; Ji, Y.; Ma, L.; Li, C.; Xu, M. Nonlinear dynamic characteristics analysis for support bearing of rotary worktable. *Mech. Based Des. Struct. Mach.* **2026**, *54*, 2544822. [\[CrossRef\]](#)
7. Yikun, E.; Feng, Z.; Liu, Z.; Chen, H.; Zhang, Y.; Gao, T.; Lin, R. Dynamic modeling and vibration analysis for cycloid gear bearing fault diagnosis of rotate vector reducers. *Mech. Syst. Signal Process.* **2026**, *244*, 113748. [\[CrossRef\]](#)
8. Wang, Z.; Zhao, R.; Chen, S.; Wang, Z.; Zhang, K.; Zhou, P.; Gao, L. Nonlinear vibration and slippage characteristics analysis of full-ceramic bearings with dynamic surface waviness. *Tribol. Int.* **2025**, *208*, 110664. [\[CrossRef\]](#)
9. Xu, Z.; Chen, Y. Dynamic modeling and analysis of local defect system of rolling bearing outer ring. *J. Vib. Eng. Technol.* **2025**, *14*, 24. [\[CrossRef\]](#)
10. Wang, Y.; Ran, Y.; Shi, Z.; Zhu, X.; Sagor, N.J. Dynamic modeling of ball bearing with irregular raceway spall profile and shaft speed fluctuation. *J. Vib. Eng. Technol.* **2026**, *14*, 97. [\[CrossRef\]](#)
11. Luo, Y.; Liu, N.; Xiao, S.; Wang, W. Dynamic behaviors of a rolling bearing-rotor system with bearing extended defect and shaft crack: Simulation and experimental investigation. *Mech. Syst. Signal Process.* **2026**, *247*, 113978. [\[CrossRef\]](#)
12. Eizmendi, M.; Heras, I.; Abasolo, M.; Aguirrebeitia, J. Four-point contact slewing bearing dynamics. Guidelines for FE modelling and mechanistic model correlation. *Finite Elem. Anal. Des.* **2025**, *247*, 104347. [\[CrossRef\]](#)
13. Sadaat, F.; Nutakor, C.; Sopanen, J. Different modeling methodologies for rotor-system on ball bearing cartridge with squeeze film damper. *J. Sound Vib.* **2026**, *625*, 119606. [\[CrossRef\]](#)
14. Han, Y.; Zhang, F.; Li, Z.; Wang, Q.; Li, C.; Lai, P.; Li, T.; Teng, F.; Jin, Z. MT-Convformer: A multitask bearing fault diagnosis method using a combination of cnn and transformer. *IEEE T Instrum. Meas.* **2025**, *74*, 3501816. [\[CrossRef\]](#)
15. Ahsan, M.; Rodriguez, J.; Abdelrahman, M. Bearing fault diagnosis in induction motors using low-cost triaxial ADXL355 accelerometer and a hybrid CWT-DCNN-LSTM model. *IEEE ACCESS* **2025**, *13*, 101037–101050. [\[CrossRef\]](#)
16. Orhan, A.; Yordanov, N.; Ertargin, M.; Zhilevski, M.; Mikhov, M. A comparative study of time-frequency representations for bearing and rotating fault diagnosis using vision transformer. *Machines* **2025**, *13*, 737. [\[CrossRef\]](#)
17. Alam, T.E.; Manjurul Ahsan, M.; Raman, S. Multimodal bearing fault classification under variable conditions: A 1D CNN with transfer learning. *Mach. Learn. Appl.-Neth.* **2025**, *21*, 100682. [\[CrossRef\]](#)
18. Gbashi, S.M.; Olatunji, O.O.; Adedeji, P.A.; Madushele, N. PSO-Optimised autoencoder for fault prediction in wind turbine planet carrier bearing. *Results Eng.* **2025**, *26*, 104844. [\[CrossRef\]](#)

19. Shi, H.; Li, Y.; Zhao, C.; Wang, J.; Liu, X. A method for discontinuous data repair and precise health monitoring of rolling bearings integrating particle filtering and digital twin. *Mech. Syst. Signal Process.* **2026**, *250*, 114136. [[CrossRef](#)]
20. Ko, T.; Park, J.; Jang, J.; Oh, K.-Y.; Lee, S.H. Adaptive bearing fault diagnosis using reference frequency and modified scalogram. *Struct. Health Monit.* **2024**, *23*, 2781–2794. [[CrossRef](#)]
21. Yu, Z.; Zhang, C.; Deng, C. An improved GNN using dynamic graph embedding mechanism: A novel end-to-end framework for rolling bearing fault diagnosis under variable working conditions. *Mech. Syst. Signal Process.* **2023**, *200*, 110534. [[CrossRef](#)]
22. Wang, Z.; Zhao, R.; Wang, Z.; Zhang, K.; Gao, L.; Liu, Z. Nonlinear vibration analysis of full-ceramic bearing-rotor systems with a ball defect considering thermal deformation. *J. Vib. Control* **2026**, 1–15. [[CrossRef](#)]
23. Gao, L.; Zhan, K.; Wang, Z. Nonlinear dynamic analysis of ceramic bearing-spindle system considering fit clearance dynamic variations. *Int. J. Nonlin Mech.* **2025**, *175*, 105104. [[CrossRef](#)]
24. Wang, B.; Lei, Y.; Li, N.; Li, N. A hybrid prognostics approach for estimating remaining useful life of rolling element bearings. *IEEE T Reliab.* **2018**, *69*, 401–412. [[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.