



OPEN

Research on fault diagnosis method of planetary gearbox based on dynamic simulation and deep transfer learning

Meng-Meng Song¹, Zi-Cheng Xiong^{1,2}, Jian-Hua Zhong^{2✉}, Shun-Gen Xiao^{1,2✉} & Yao-Hong Tang¹

To address the issue of not having enough labeled fault data for planetary gearboxes in actual production, this research develops a simulation data-driven deep transfer learning fault diagnosis method that applies fault diagnosis knowledge from a dynamic simulation model to an actual planetary gearbox. Massive amounts of different fault simulation data are collected by creating a dynamic simulation model of a planetary gearbox. A fresh deep transfer learning network model is built by fusing one-dimensional convolutional neural networks, attention mechanisms, and domain adaptation methods. The network model is used to learn domain invariant features from simulated data, thereby enabling fault diagnosis on real data. The fault diagnosis experiment is verified by using the Drivetrain Diagnostics Simulator test bench. The validity of the proposed means is evaluated by comparing the diagnostic accuracy of various means on various diagnostic tasks.

Planetary gearbox (PG) is a critical part of rotating machines, in virtue of its merit of great carrying capacity, small volume, and high driving efficiency, it has been more and more extensively applied in mechanical transmission systems of wind power, aviation, lifting and transportation, and other industries¹. However, the working environment of PG is usually very harsh, such as often changes in the working environment, heavy loads, bad weather, and many other factors, making it easy to fail. Once the PG fails, it may cause mechanical equipment downtime. Health monitoring and fault diagnosis of PG can keep the mechanical equipment running safely and stably, and reduce possible losses caused by faults.

The conventional fault diagnosis means comprises three steps: data collection, feature extraction, and fault classification². First, the data during the operation of the mechanical equipment is collected by sensors, and then the features are manually picked up from the collected data. Finally, the extracted features are used to train machine learning models such as SVM³, KNN⁴, RF⁵, and ANN⁶ for failure prediction. However, manual feature extraction has obvious disadvantages. It requires a lot of manpower and some professional basic knowledge.

Deep learning (DL) is a major breakthrough in artificial intelligence in recent decades. It can automatically pick out representative features from raw data, and accurately build nonlinear mapping relationships of different health status features, which greatly surmounts the deficiencies of traditional fault diagnosis means. DL models such as CNN⁷, AE⁸, DBN⁹, and RNN¹⁰ have been diffusely researched and utilized in the fault diagnosis field.

When making use of DL means for fault diagnosis, the following two assumptions usually need to be satisfied: (1) sufficient labeled data is available; (2) the feature distribution in the training data and the test data are identical. However, in actual production, majority of data are gathered when the machines is running healthy¹¹, so it is very hard to obtain comprehensive and extensive labeled failure data.

In view of the scarcity of fault data, dynamic simulation analysis may be a good solution. The dynamic simulation analysis and diagnostic means frequently begins with the creation of a dynamic model of machines to simulate dynamic response of machines under multiple conditions. Then, the dynamic response may be used to do diagnostics¹². Xiao et al.^{13–15} revealed the dynamic response of mechanical system under the influence of rub impact fault by establishing a dynamic model. Han et al.¹⁶ built a revised lumped parameter dynamic model, which revealed the fluctuation of the dynamic load distribution coefficient when gear crack faults were considered. Chen et al.¹⁷ developed an analytical model for meshing stiffness calculation using the potential energy

¹College of Information, Mechanical and Electrical Engineering, Ningde Normal University, Ningde, China. ²College of Mechanical Engineering and Automation, Fuzhou University, Fuzhou, China. ✉email: jHzhong@fzu.edu.cn; xiaoshungen022@163.com

principle and uniform bending Timoshenko beam theory, and demonstrated the effect law of flexible ring gear and tooth root cracking fault on dynamic response. Park et al.¹⁸ designed a lumped parameter model of planetary gear dynamics to extract transmission error signal, and applied the rotation error signal to fault diagnosis. Fan et al.¹⁹ built a dynamic model of the planetary gear train with planetary carrier crack fault, and demonstrated a correlation between vibration characteristics and planetary carrier crack conditions. Duan et al.²⁰ used a system-level rigid-flexible coupled model consisting of shells, gears, shafts and bearings to detect the crack evolution-induced vibration, and acquired relevant spectral characteristics, statistical indicators and instantaneous energy of the crack evolution-induced vibration.

By constructing a dynamic model of the mechanical equipment to be diagnosed and solving it numerically, the simulation data of various working conditions and common fault types can be obtained easily, which solves the problem of insufficient real faults. However, limited by the complexity of the simulation model and calculation, the simulation data obtained are often too ideal, lacking noise in the actual environment, which is significantly different from the real data. Therefore, the fault diagnosis model built by the simulation data set will not attain a better precision on the real data set. Transfer learning (TL) is a useful way to settle the matter of data discrepancy.

The purpose of TL is to adapt acquired from one or more tasks (source domain) to other associated but distinct tasks (target domain). It is inspired by the human ability to reuse knowledge from previous tasks to new ones without having to start from the beginning. He et al.²¹ proposed an enhanced depth transmission automatic encoder to address the issue of inadequate bearing vibration data with marks, which initialized the target model through the source model's trained parameter. Then only just a small subset of target training samples are used to fine-tune the target model to fit the features of the remaining target test samples. Li et al.²² put forward a joint attention feature transfer network model. The established feature transfer module transfers the generalized representation obtained from the class with more samples to the class with few samples, thus expanding its feature space and solving the matter of data imbalance. Zhu et al.²³ calculated the domain loss between the source domain and the target domain through the linear combination of multiple Gaussian kernels, which enhanced the adaptive ability of the diagnostic model. Wan et al.²⁴ put forward a TL means combining sensitive feature selection and sparse automatic encoder, which reduces the interference of insensitive information in the original signal and raises the diagnostic precision of the model under complex operating conditions. Han et al.²⁵ extended the marginal distribution adaptation to joint distribution adaptation, allowing the suggested network to employ the identification structure linked to the labeled data in the source domain to adapt to the conditional distribution of the unlabeled target data and ensure higher precise distribution matching. Zheng et al.²⁶ put forward a normalized recursive dynamic adaptive network, which can estimate the relative weight of marginal and conditional distributions dynamically and quantitatively. Deng et al.²⁷ put forward a sample weighted joint adversarial network that takes use of the classification information to improve the joint domain adaptability of adversarial learning. Han et al.²⁸ proposed a new framework to address the issue of transfer diagnosis for sparse target data. Individual domain adaptation is performed on the source domain data and target data under the same work conditions to reduce distribution differences and avoid the negative transfer.

Aiming at the scarcity of marked fault data in practical fault diagnosis, taking PG as the research object, a fault diagnosis means based on dynamic simulation and deep transfer learning (DTL) is put forward in this paper. By building the dynamic simulation model of the PG, sufficient fault simulation data can be gathered. On this basis, the time-domain graphs (TDG) and frequency-domain graphs (FDG) of the simulation data are evaluated to confirm the validity of established models. In addition, the influence law of various simulation step sizes on simulation data is explored. A fresh DTL network model is built. The network model is composed of three parts: one-dimensional convolution neural network, attention mechanism, and domain adaptive method. Through the network model, the fault diagnosis knowledge contained in a simulation data is used in a real PG's fault diagnosis, and the influence law of simulation data with distinct parameters on the diagnosis precision is explored.

The remaining of this article is arranged as below: “[Theoretical background](#)” outlines the relevant theories. “[Proposed method](#)” details the suggested means. “[Experiment and analysis](#)” studies a practical case. “[Conclusions](#)” summarizes this paper.

Theoretical background

Dynamics theory based on ADAMS. In ADAMS, the Cartesian coordinates of the rigid body's centroid and the Euler angles representing its orientation are used as generalized coordinates, hence, $q = [x, y, z, \psi, \theta, \varphi]^T$, Eq. (1) is obtained by the energy form of Lagrange equations of the first kind.

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_j} \right) - \frac{\partial T}{\partial q_j} = Q_j + \sum_{i=1}^n \lambda_i \frac{\partial \phi}{\partial q_j} \quad (1)$$

Here, T denotes the kinetic energy indicated in the generalized coordinate of the system, q_j denotes the generalized coordinate, Q_j represents the generalized force in the direction of q_j , and the last term of the equation represents the constraint reaction in the direction of q_j .

Introducing generalized momentum p_j

$$p_j = \frac{\partial T}{\partial \dot{q}_j} \quad (2)$$

The dynamic differential-algebraic equations of the system are established by integrating the constraint equations.

$$\left. \begin{aligned} \dot{p} - \frac{\partial T}{\partial q} + \Phi_q^T \lambda + H^T F \\ P = \frac{\partial T}{\partial \dot{q}} \\ u = \dot{q} \\ \Phi(q, t) = 0 \\ F = f(u, q, t) \end{aligned} \right\} \quad (3)$$

Here, Φ is the constraint function, Φ_q is the Jacobian matrix of the constraint equation, H is the coordinate transformation matrix of the external force, λ is the Lagrange multiplier, and F represents the external force on the system.

There are three different integration schemes in ADAMS to reduce the order of dynamic differential–algebraic equations.

I3 integral format:

$$\left. \begin{aligned} \dot{p} - \frac{\partial T}{\partial q} + \Phi_q^T \lambda + H^T F = 0 \\ P = \frac{\partial T}{\partial \dot{q}} \\ u = \dot{q} \\ \Phi(q, t) = 0 \\ F = f(u, q, t) \end{aligned} \right\} \quad (4)$$

SI2 integral format:

$$\left. \begin{aligned} \dot{p} - \frac{\partial T}{\partial q} + \Phi_q^T \lambda + H^T F = 0 \\ P = \frac{\partial T}{\partial \dot{q}} \\ u - \dot{q} + \Phi_q^T \mu = 0; (\mu = 0) \\ \Phi(q, t) = 0 \\ \dot{\Phi}(q, u, t) = 0 \\ F = f(u, q, t) \end{aligned} \right\} \quad (5)$$

SI1 integral format:

$$\left. \begin{aligned} \dot{p} - \frac{\partial T}{\partial q} + \Phi_q^T \dot{\eta} + H^T F = 0 \\ P = \frac{\partial T}{\partial \dot{q}} \\ u - \dot{q} + \Phi_q^T \dot{\xi} = 0 \\ \Phi(q, t) = 0 \\ \dot{\Phi}(q, u, t) = 0 \\ F = f(u, q, t) \end{aligned} \right\} \quad (6)$$

Transfer learning. First, various TL-related symbols are given to properly indicate the issues that need to be resolved. Given the source domain $D_{src} = \{(x_i, y_i)\}_{i=1}^{n_{src}}$ with n_{src} labeled samples and the target domain $D_{tar} = \{(x_i)\}_{i=1}^{n_{tar}}$ with n_{tar} unlabeled samples, where $X = \{(x_i)\}_{i=1}^n$ denotes the feature space and $Y = \{(y_i)\}_{i=1}^n$ denotes the label space. Let $P(x)$ and $Q(y|x)$ represent marginal and conditional probability distribution respectively. The goal of TL is to use D_{src} diagnostic knowledge with D_{tar} in the case of $X_{src} = X_{tar}$, $Y_{src} = Y_{tar}$, $P_{src}(x_{src}) \neq P_{tar}(x_{tar})$, $Q_{src}(y_{src}|x_{src}) \neq Q_{tar}(y_{tar}|x_{tar})$.

Joint distribution adaptation. The joint distribution adaptation (JDA) method was come up with Long et al.²⁹, and its goal is to find a transformation A to make the distance between $P_{src}(A^T x_{src})$ and $P_{tar}(A^T x_{tar})$ as close as possible, and at the same time, make the distance between the transformed conditional probability distribution $Q_{src}(y_{src}|A^T x_{src})$ and $Q_{tar}(y_{tar}|A^T x_{tar})$ also as close as possible.

The approach is broken down into two stages. First, the maximum mean discrepancy (MMD) would be utilized to judge the disparity between $P_{src}(x_{src})$ and $P_{tar}(x_{tar})$. The distance between the sample means of D_{src} sample and D_{tar} sample can be demonstrated that

$$\left\| \frac{1}{n_{src}} \sum_{i=1}^{n_{src}} A^T x_i - \frac{1}{n_{tar}} \sum_{j=n_{src}+1}^{n_{src}+n_{tar}} A^T x_j \right\|^2 = \text{tr}(A^T X M_o X^T A) \tag{7}$$

where $\text{tr}(\cdot)$ stands for a matrix's trace, X denotes the input data matrix, M_o represents the MMD matrix and Eq. (8) shows the computation equation.

$$(M_o)_{ij} = \begin{cases} \frac{1}{n_{src}n_{src}}, & x_i, x_j \in D_s \\ \frac{1}{n_{tar}n_{tar}}, & x_i, x_j \in D_t \\ \frac{-1}{n_{src}n_{tar}}, & \text{otherwise} \end{cases} \tag{8}$$

Then $Q_{src}(y_{src}|x_{src})$ and $Q_{tar}(y_{tar}|x_{tar})$ are adapted, and the distance between the conditional probability distributions can be written as

$$\left\| \frac{1}{n_{src}^{(c)}} \sum_{x_i \in D_{src}^{(c)}} A^T x_i - \frac{1}{n_{tar}^{(c)}} \sum_{x_j \in D_{tar}^{(c)}} A^T x_j \right\|^2 = \text{tr}(A^T X M_c X^T A) \tag{9}$$

where $D_{src}^{(c)}$ represents the source domain data with category c , $D_{tar}^{(c)}$ represents the target domain data with category c , M_c represents the MMD matrix involving class labels and its calculation formula is shown in Eq. (10).

$$(M_c)_{ij} = \begin{cases} \frac{1}{n_{src}^{(c)}n_{src}^{(c)}}, & x_i, x_j \in D_{src}^{(c)} \\ \frac{1}{n_{tar}^{(c)}n_{tar}^{(c)}}, & x_i, x_j \in D_{tar}^{(c)} \\ \frac{-1}{n_{src}^{(c)}n_{tar}^{(c)}}, & \begin{cases} x_i \in D_{src}^{(c)}, x_j \in D_{tar}^{(c)} \\ x_j \in D_{src}^{(c)}, x_i \in D_{tar}^{(c)} \end{cases} \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

Combining Eqs. (7) and (9), a general optimization objective is obtained

$$\min \sum_{c=0}^C \text{tr}(A^T X M_c X^T A) + \lambda \|A\|_F^2 \tag{11}$$

When Eq. (11) is the smallest, the JDA distance is the smallest, that is, the disparity between D_{src} sample and D_{tar} sample is the smallest.

Attention mechanism. Attention mechanism is derived from the studies of human vision. When mankind recognizes things, they often focus more on important information and ignore irrelevant information. Different features contribute differently to the final classification results throughout the training phase of a DL model. The attention method allows DL model to concentrate on features that have a significant impacts on classification precision and disregard those that do not matter, improving classification performance. CBAM put forward by woo et al.³⁰, is an attention mechanism module that integrates space and channel. The method first assigns various weights to various channels of the input feature through the channel attention module. The weight given by important channels is larger and that given by unimportant channels is smaller. Then, through the spatial attention module, different weights are given to different regions of the input features. The overall calculation process is shown in Eqs. (12) and (13).

$$C' = M_c(C) \otimes C \tag{12}$$

$$C'' = M_s(C') \otimes C' \tag{13}$$

Among them, C represents the input feature map, C' represents the feature map processed by the channel attention module, C'' represent the output feature map, M_c and M_s denote the channel and spatial attention modules, respectively, and \otimes represents the element by element multiplication.

Proposed method

In this paper, a fresh fault diagnosis means for PG is put forward based on dynamic simulation and DTL. This method mainly includes three parts: Building dynamic simulation models of PG, acquiring and analyzing the simulation data and real data, and fault diagnosis of PG based on TL. Figure 1 shows the general framework of the means.

Building dynamic simulation models of PG. In the light of the sizes of the actual PG related parts, use the 3D modeling software to build a 3D model, and import the 3D model into ADAMS to build a dynamic simulation model.

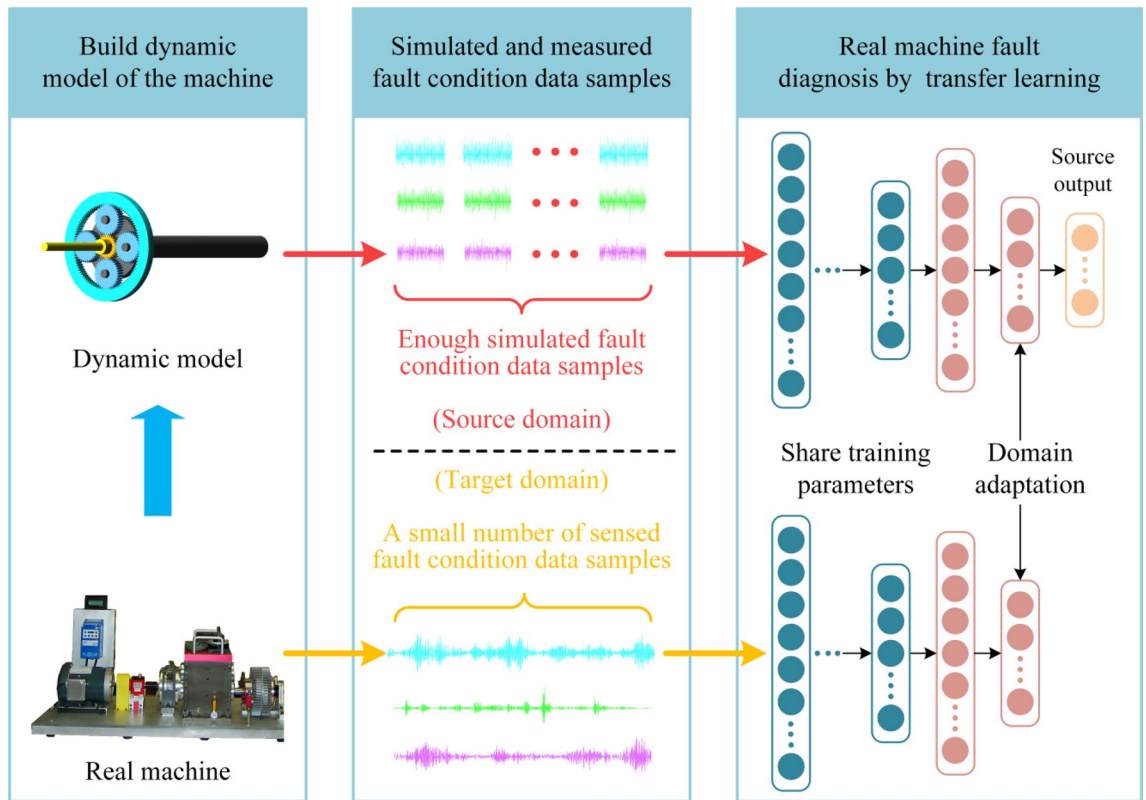


Figure 1. The framework of fault diagnosis means based on dynamic simulation and DTL.

Acquiring and analyzing the simulation data and real data. Through the dynamic simulation model, the dynamic response of the actual PG under various fault conditions is simulated, and sufficient simulation data are obtained. By analyzing the TDG and FDG of the fault simulation data, the simulation model’s validity is confirmed. By comparing the TDG and FDG of different simulation data, the influence law of simulation step size on simulation data is explored; Collect a small amount of real data from the actual PG.

Fault diagnosis of PG based on TL. The proposed DTL network model is comprised of three modules: feature extractor, health classifier, and domain adaptation. The model’s framework is shown in Fig. 2. The feature extractor is applied to extract high-level abstract features of data. Figure 3 shows the structure of the multi-scale feature extraction layer, which extracts features in parallel using three convolution kernel branches of various scales to obtain features of various scales. By stacking the features of these three scales, a rich multi-scale feature set is formed, and Table 1 displays the feature extractor’s network parameters. In the attention mechanism layer, CBAM is applied to enhance important features and suppress irrelevant features; The health status classifier is applied to distinguish health conditions of D_{src} features. The domain adaptation part adopts joint distribution adaptation (JDA) to execute domain adaptation on D_{src} features and D_{tar} features.

Two optimization objectives must be met during model training:

- (1) Minimize the classification loss L_c of the feature extractor on D_{src} data, L_c is calculated by the cross-entropy loss function. Following is the calculating formula:

$$L_c = - \sum_{i=1}^N y_{true}^i \log(y_{pred}^i) \tag{14}$$

where y_{true} represents the D_{src} data’s true label, y_{pred} represents the D_{src} data’s predicted label.

- (2) Minimize the domain loss L_d , which is calculated by JDA.

By combining the classification loss L_c and the domain loss L_d , the overall optimization target of the model can be obtained:

$$L = L_c + \lambda L_d \tag{15}$$

in which, λ Represents the penalty factor. By minimizing L , the features learned by the feature extractor can have small domain differences and become domain invariant features, and health condition classifier can accurately classify these features. Take the simulation samples as D_{src} and the unlabeled real samples as D_{tar} into the DTL

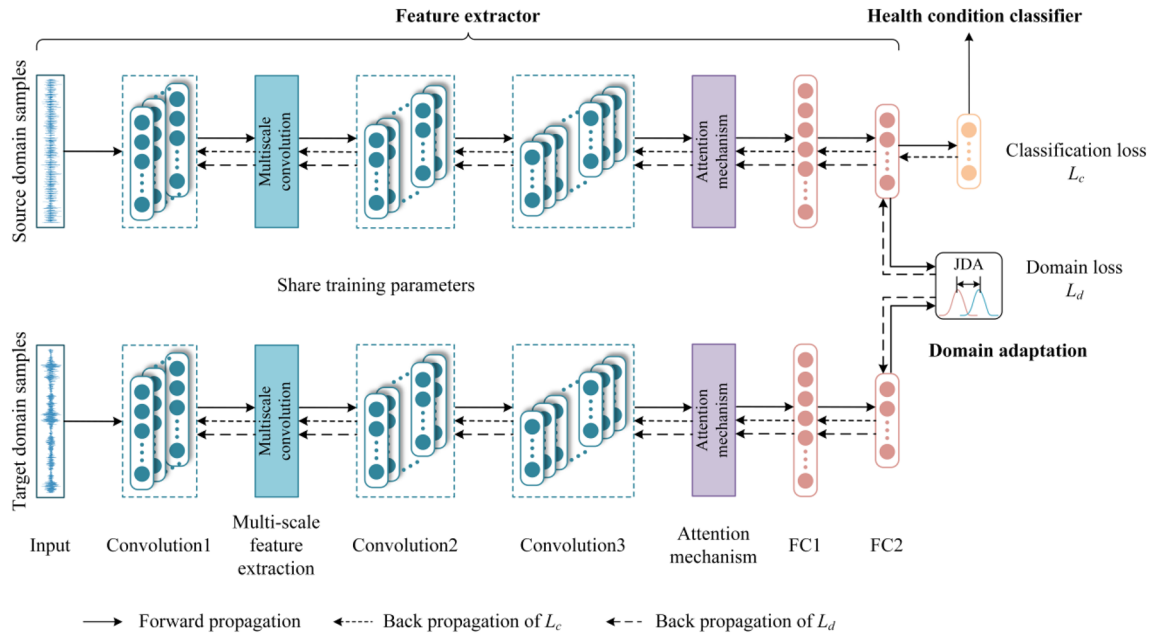


Figure 2. The framework of DTL network model.

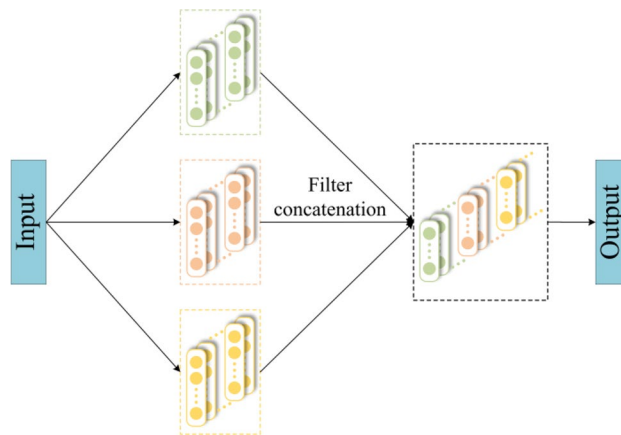


Figure 3. Multi-scale feature extraction layer.

Network structure	Number of convolution kernels	Convolution kernel size	Output size
Convolution layer1	16	64	16×128
Multiscale feature extraction layer	32	3	16×126
		5	
		7	
Convolution layer2	96	3	96×124
Convolution layer3	192	3	192×122
Full connection layer 1	128	–	128×1
Full connection layer 2	3	–	3×1

Table 1. Model network parameters.

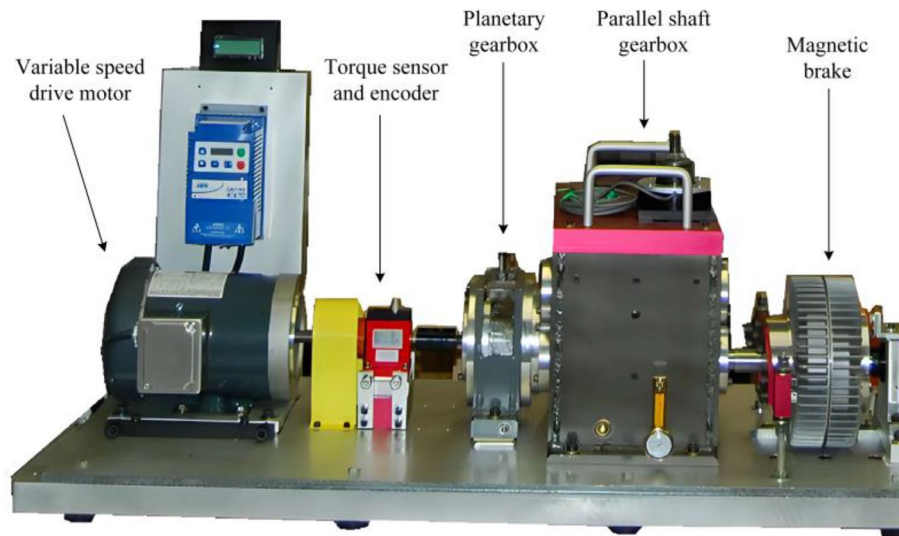


Figure 4. Physical drawing of DDS test bench.

Part	Quantity of teeth	Modulus	Pressure angle	Tooth width (mm)
Sun gear	28	1	20°	10
Planetary gear	36	1	20°	10
Ring gear	100	1	20°	10

Table 2. Parameters of the PG.

network model, implement fault diagnosis of actual PG, and explore influence laws of simulation data of different parameters on the diagnosis results.

Experiment and analysis

Establishment of the dynamic simulation model of PG. This paper takes the PG in Drivetrain Diagnostics Simulator (DDS) test bench designed by SpectraQuest company of America as the research target, Fig. 4 is the structural drawing of DDS laboratory bench, which is comprised of variable speed drive motor, torque sensor and encoder, PG, parallel shaft gearbox and magnetic brake. Table 2 demonstrates the fundamental parameters of the PG.

In the light of the parameters demonstrated in Table 2, the three-dimensional model of PG is built and assembled. In order to make the model more clearly reflect the dynamic characteristics, a number of parts that have seldom impact on the simulation results are simplified in the modeling process.

Save the 3D model as x_t format and import it into ADAMS. Considering that the rigid body model ignores the elastic deformation of the parts and cannot accurately transmit the vibration signal, the sun gear is constructed flexible. Import the 3D model of the sun gear into ANSYS, build and export a mode neutral file, replace the rigid body sun gear in ADAMS with this file, and obtain the rigid-flexible coupled model of the PG.

By replacing flexible body sun gears in the rigid-flexible coupled model, the PG simulation models of the sun gear under different health conditions are obtained. Figure 5 shows the physical drawing and flexible body model of the cracked and the missing tooth sun gear.

In line with the actual transmission relationship between the parts, add the constraints shown in Table 3. Since the gear pair can only transmit the speed between the gears and cannot reflect the change of the meshing force between the gears, contact forces are added between the sun gear and the planetary gears and between the planetary gears and the ring gear to replace gear pairs.

The PG model designed in this paper takes the input shaft as the input, the planetary carrier as the output, and the ring gear is fixed. Therefore, a 30 Hz rotation frequency is added to the input shaft as the speed drive, and the corresponding speed is 10,800°/s. In order to prevent the speed mutation of the input shaft, resulting in infinite acceleration, the step function is used to gradually increase the speed from 0 to 10,800°/s within 0.01 s. Figure 6 demonstrates the completed rigid-flexible coupling model of the PG.

Acquisition and analysis of simulation data. According to the calculation formula of the transmission ratio of the epicyclic gear train, the calculation formula of the planet carrier rotation frequency f_c , planetary gear train meshing frequency f_m , and sun gear local fault characteristic frequency f_g can be deduced

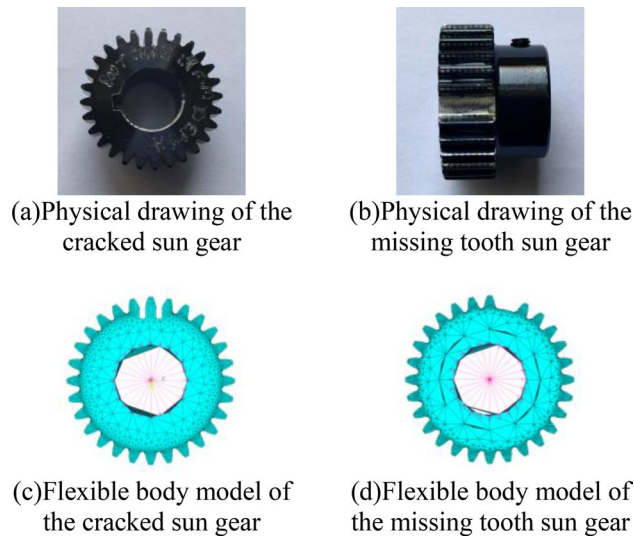


Figure 5. The physical drawing and flexible body model of the cracked and the missing tooth sun gear.

Part 1	Part 2	Constraint
Input shaft	Sun gear	fixed pair
Ring gear	Ground	fixed pair
Input shaft	Ground	Rotary pair
Planetary gear	Planet carrier	Rotary pair
Planet carrier	Ground	Rotary pair

Table 3. Constraints.

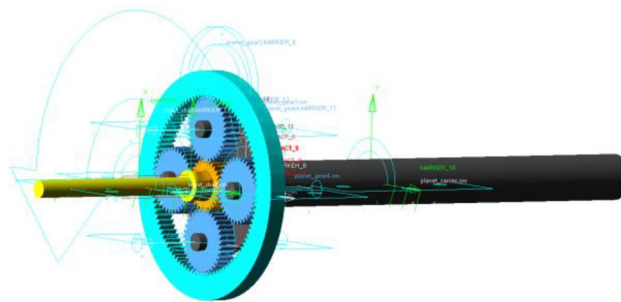


Figure 6. Rigid-flexible coupling model of the PG.

$$f_c = \frac{Z_s \times f_s}{Z_s + Z_r} \tag{16}$$

$$f_m = f_c \times Z_r \tag{17}$$

$$f_g = \frac{f_m \times N}{Z_s} \tag{18}$$

Here Z_s stands for the amount of teeth of the sun gear, Z_r stands for the amount of teeth of the ring gear, f_s denotes the rotation frequency of the sun gear, and N denotes the amount of planetary gears. Substitute $Z_s = 28$, $Z_r = 100$, $f_s = 30$ Hz, $N = 4$ into Eqs. (14)–(16) to obtain $f_c = 6.5625$ Hz, $f_m = 656.25$ Hz, $f_g = 93.75$ Hz.

When the sun gear’s faulty gear teeth mesh with the planetary gear teeth, the lubricating oil film between the sliding contact surfaces of the gear teeth will be ruptured, resulting in an impact phenomenon. The impact will bring about the amplitude modulation and frequency modulation effect on the meshing vibration, and the

Number	Simulation time (s)	Simulation step size (s)	Simulation data sampling frequency (kHz)
1	0.5	7.8125×10^{-5}	12.8
2	0.5	1.5625×10^{-5}	64
3	0.5	8.6806×10^{-6}	115.2

Table 4. Simulation parameters.

vibration signal at the meshing point can be described by the amplitude modulation and frequency modulation process. If the integer multiples of meshing frequency kf_m are used as the carrier frequency, the sidebands will appear at $kf_m + mf_g$, and the spacing between the sidebands is f_g ³¹.

In ADAMS, set the simulation time to 0.5 s, select WSTIFF as the integration solver, and I3 as the integration format. The simulation step size will affect the results of the simulation calculation. If the step length is set too big, the calculation accuracy will be low, and if the step length is set too little, the computing time will be long. Simulate on the basis of the three groups of parameters shown in Table 4 to explore the influence law of simulation step size on simulation data.

The models used in the simulation are all rigid-flexible coupled models with missing teeth sun gear. After the end of the simulation, angular acceleration on the planet carrier is exported as simulation data. Figure 7 demonstrates TDG and FDG of various simulation data.

As can be aware of the TDG in Fig. 7 that the smaller the simulation step size is, the more obvious the periodic shock phenomenon appears in the TDG, and the time interval between shocks is the same as the theoretical value $1/f_g = 0.01067$ s; As can be aware of the FDG of different simulation step size have obvious peaks at $f_m = 656.25$ Hz and its integer multiples. The smaller the simulation step size is, the more obvious sidebands appear around meshing frequency f_m and its integer multiple in the FDG, and the spacing between the sidebands are all $f_g = 93.75$ Hz.

Figure 8 demonstrates the TDG and FDG of the pure rigid body model. The sun gear in the model is missing tooth fault, and the simulation step size is 8.6806×10^{-6} s. As can be aware of Fig. 8, the TDG of the rigid body model has no periodic impact, and the FDG has obvious peaks at the meshing frequency and its frequency doubling, but there are no sidebands.

Based on the above comparative analysis, it can be found out that the impact characteristics in the TDG and the fault frequency characteristics in the FDG of the simulation data are consistent with the theory, which certifies the rationality of the dynamic simulation model. Furthermore, contrasted with the pure rigid body model, the rigid-flexible coupling model can better reflect the fault characteristics of the planetary gear train, and the smaller the simulation step size, the clearer the periodic impact in the TDG and the sidebands in the FDG.

Data description. The real data are the vibration signals acquired from PG of the DDS test bench. When collecting the signal, variable speed drive motor's rotation frequency is 30 Hz, and there are four different currents on the magnetic brake: 0 A, 0.4 A, 0.8 A, and 1.2 A (different loads can be applied to the output shaft by adjusting the magnetic brake's current). The sun gear contains three health conditions: sun gear cracks, sun gear missing teeth, sun gear normal, and sampling frequency is 12.8 kHz.

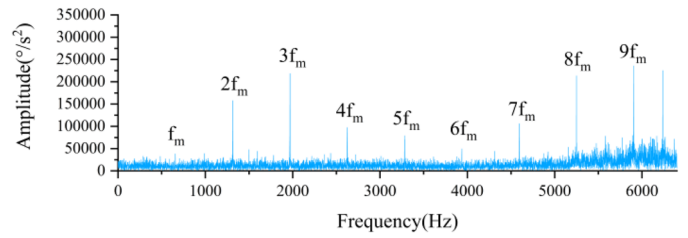
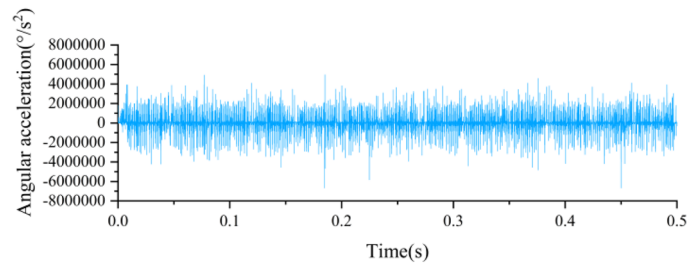
The input shaft rotation frequency of the simulation model is 30 Hz and there is no load. After the end of the simulation, angular acceleration on the planet carrier is obtained as simulation data.

The simulation data and real data are overlapping sampling to obtain simulation samples and real samples. 2048 points are sampled every 100 points, that is, 2048 data points per sample. Table 5 displays the number of samples.

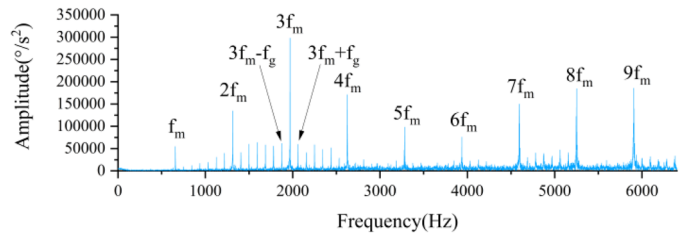
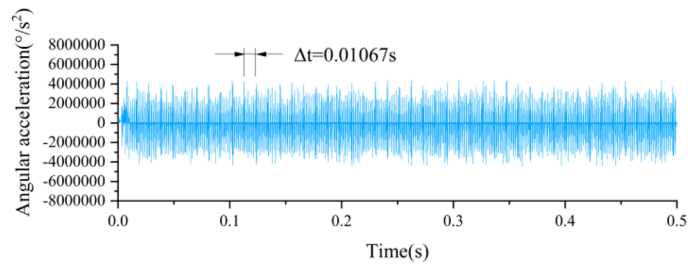
Analysis of influence law of simulation parameters. There are five different integration solvers in Adams: GSTIFF, WSTIFF, HHT, Newmark, HASTIFF, and three integration formats: I3, SI1, SI2. The characteristics of various solution means are demonstrated in Table 6. Take the simulation data of various solution means as D_{src} and the real data set B as D_{tar} for TL. The diagnosis precision is demonstrated in Fig. 9. As can be aware of Fig. 9, when the integration solver selects WSTIFF and the integration format selects I3, the diagnosis accuracy is the highest.

Classification results and comparison. The validity of the proposed means is proved by five TL tasks of $A \rightarrow B$, $A \rightarrow C$, $A \rightarrow D$, $A \rightarrow E$, and $B \rightarrow E$. In the first four TL tasks, the training dataset consists of 6000 labeled simulated samples and 3000 unlabeled real samples, and the leftover 600 real samples are used for testing. In TL task $B \rightarrow E$, The training dataset consisted of 3000 labeled B dataset samples and 3000 unlabeled E dataset samples, and the remaining 600 E dataset samples were used for testing. CNN, DeepCoral³², DDC³³, DANN³⁴ are selected for comparative experiments. In order to compare the accuracy of various means more reasonably, all the above means use the same CNN structure and parameters as the means proposed in this paper.

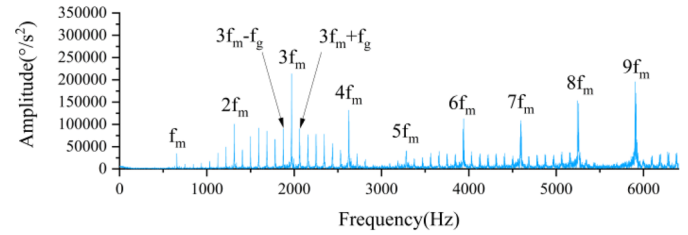
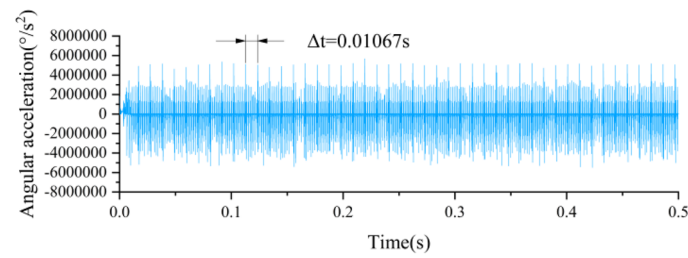
Figure 10 displays the result from several diagnostic means. In the four transfer learning tasks, the proposed means obtains the maximum accuracy, demonstrating its effectiveness; Additionally, the accuracy of transfer between real data under various working conditions reaches 100%, which is higher than the accuracy of the transfer from simulation data to real data. The reason is that there is a significant discrepancy between simulation data and real data, while the difference between real data under different working conditions is relatively small. This causes transfer between simulation data and real data to be more challenging and to have poorer accuracy. In order to intuitively show the difference between different data, the normal distribution histograms



(a)The simulation step size is $7.8125 \times 10^{-5}s$



(b)The simulation step size is $1.5625 \times 10^{-5}s$



(c)The simulation step size is $8.6806 \times 10^{-6}s$

Figure 7. TDG and FDG of different simulation step sizes of rigid-flexible coupling model.

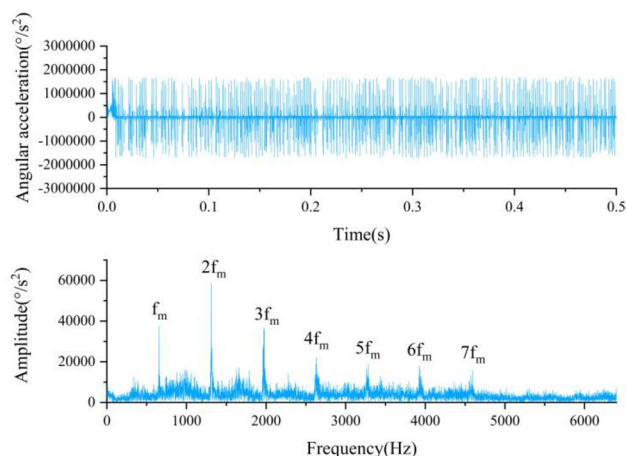


Figure 8. TDG and FDG of the pure rigid body model.

Name	Dataset	Health condition	Number of samples	Load
A	Simulation data	Crack	2000	0
		Missing teeth	2000	0
		Normal	2000	0
B	Real data	Crack	1200	0
		Missing teeth	1200	0
		Normal	1200	0
C	Real data	Crack	1200	0.4
		Missing teeth	1200	0.4
		Normal	1200	0.4
D	Real data	Crack	1200	0.8
		Missing teeth	1200	0.8
		Normal	1200	0.8
E	Real data	Crack	1200	1.2
		Missing teeth	1200	1.2
		Normal	1200	1.2

Table 5. Data set description.

Solver	Integral format	Characteristics of solution method	Calculation accuracy and simulation speed
GSTIFF	I3	Estimate-correction error is judged by displacement error	Highest calculation accuracy and fast simulation speed
WSTIFF	I3		
GSTIFF	SI2	Estimate-correction error is judged by speed error	High calculation accuracy, slowest simulation speed
WSTIFF	SI2		
HASTIFF	SI1/SI2		
HHT	I3	Estimate-correction error is judged by acceleration error	The lowest calculation accuracy and the fastest simulation speed
Newmark	I3		

Table 6. Characteristics of the different solution means in ADAMS.

of the simulated tooth missing tooth fault data with load of 0 A, the real tooth missing tooth fault data with load of 0 A and the real tooth missing tooth fault data with load of 1.2 A are drawn, respectively. It can also be seen from Fig. 11 that the disparities between the simulation data and the real data is far greater than the disparities between the real data under different working conditions.

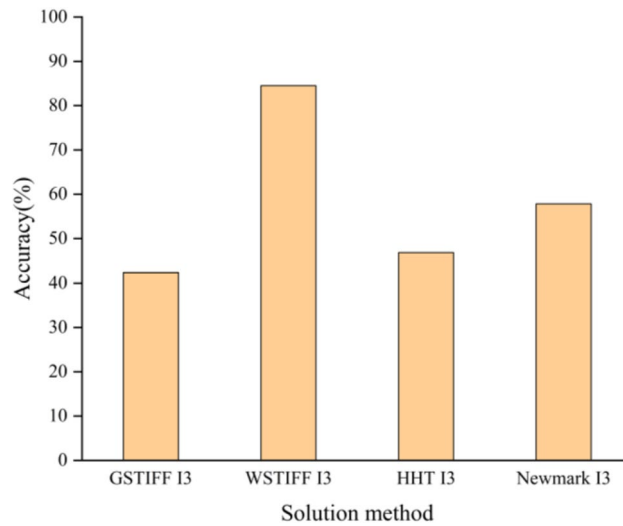


Figure 9. Diagnostic accuracy of simulation data for different solution methods.

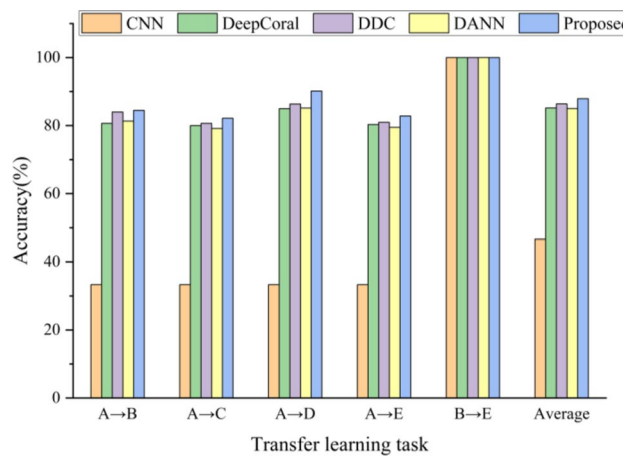
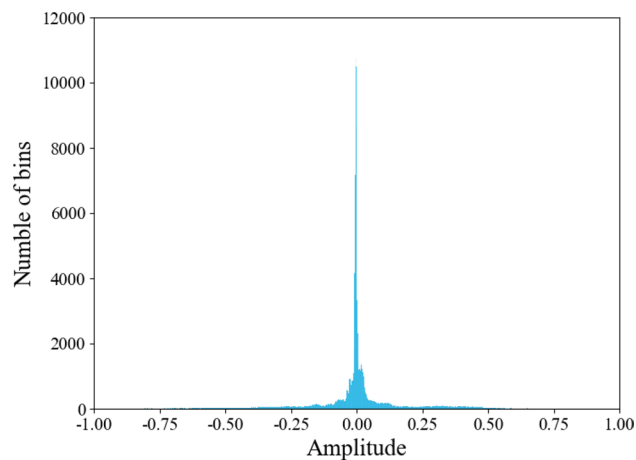


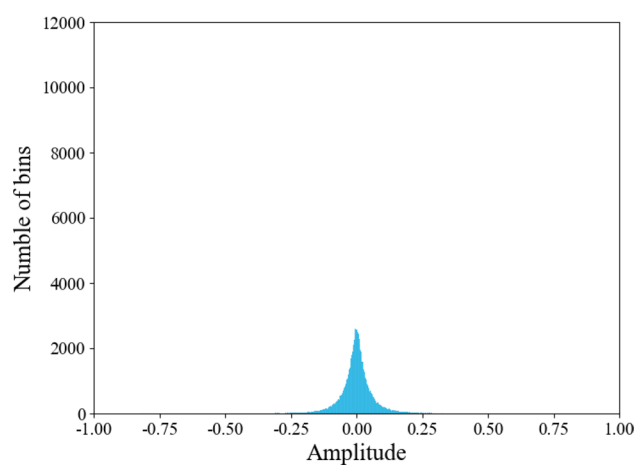
Figure 10. Diagnostic precision of various means on various transfer tasks.

Conclusions

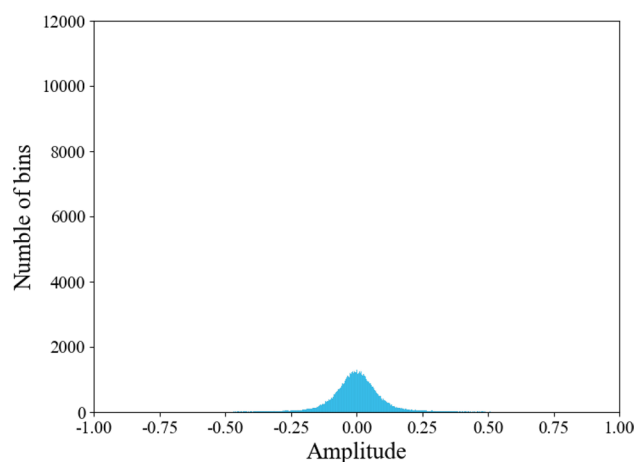
To achieve accurate fault diagnosis of PG when marked fault data is insufficient, a fault diagnosis means based on dynamic simulation and DTL is put forward in this paper. The dynamic simulation model of the PG was built by using ADAMS, and the simulation data of various health conditions were obtained. By analyzing the fault simulation data, it is found that the rigid-flexible coupling model can better mirror the fault characteristics of the planetary gear train, and the smaller the simulation step size is, the clearer the periodic impact in TDG and the sidebands in FDG are. By fusing one-dimensional convolutional neural network, attention mechanism, and domain adaptation method, a novel DTL network model is built, and the network model is used to apply diagnostic knowledge from simulation data to real PG’s fault diagnosis. The fault diagnosis experiment was verified by using DDS test bench. By contrasting the diagnosis results of the simulation data with various parameters, it was found that when the integration solver of the simulation data was WSTIFF, the integration format was I3, and the sampling frequency was the same as the real data, the diagnostic accuracy was the highest. The validity of the proposed means is verified by contrasting the diagnostic precision of various means on various transfer tasks.



(a) Simulate missing tooth fault data, the load is 0A



(b) Real missing tooth fault data, the load is 0A



(c) Real missing tooth fault data, the load is 1.2A

Figure 11. Normal distribution histograms.

Data availability

The datasets generated and/or analysed during the current study are not publicly available due the data also forms part of an ongoing study, but are available from the corresponding author on reasonable request.

Received: 6 June 2022; Accepted: 26 September 2022

Published online: 11 October 2022

References

1. Lei, Y., He, Z., Lin, J., Han, D. & Kong, D. Research advances of fault diagnosis technique for planetary gearboxes. *Chin. J. Mech. Eng.* **47**, 59–67. <https://doi.org/10.3901/JME.2011.19.059> (2011).
2. Glowacz, A. Fault diagnosis of single-phase induction motor based on acoustic signals. *Mech. Syst. Signal Process.* **117**, 65–80. <https://doi.org/10.1016/j.ymssp.2018.07.044> (2019).
3. Shi, Q. & Zhang, H. Fault diagnosis of an autonomous vehicle with an improved SVM algorithm subject to unbalanced datasets. *IEEE Trans. Ind. Electron.* **68**, 6248–6256. <https://doi.org/10.1109/tie.2020.2994868> (2021).
4. Li, D. *et al.* Continual learning classification method with the weighted k-nearest neighbor rule for time-varying data space based on the artificial immune system. *Knowl.-Based Syst.* **240**, 108145. <https://doi.org/10.1016/j.knosys.2022.108145> (2022).
5. Chen, Z. C. *et al.* Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents. *Energy Convers. Manage.* **178**, 250–264. <https://doi.org/10.1016/j.enconman.2018.10.040> (2018).
6. Liao, Y. X., Zhang, L. & Li, W. H. Regrouping particle swarm optimization based variable neural network for gearbox fault diagnosis. *J. Intell. Fuzzy Syst.* **34**, 3671–3680. <https://doi.org/10.3233/jifs-169542> (2018).
7. Han, T., Zhang, L., Yin, Z. & Tan, A. C. Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine. *Measurement* **177**, 109022. <https://doi.org/10.1016/j.measurement.2021.109022> (2021).
8. Choi, Y. & Yoon, S. Autoencoder-driven fault detection and diagnosis in building automation systems: Residual-based and latent space-based approaches. *Build. Environ.* <https://doi.org/10.1016/j.buildenv.2021.108066> (2021).
9. Gai, J., Zhong, K., Du, X., Yan, K. & Shen, J. Detection of gear fault severity based on parameter-optimized deep belief network using sparrow search algorithm. *Measurement* **185**, 110079. <https://doi.org/10.1016/j.measurement.2021.110079> (2021).
10. Liu, H., Zhou, J. Z., Zheng, Y., Jiang, W. & Zhang, Y. C. Fault diagnosis of rolling bearings with recurrent neural network based autoencoders. *ISA Trans.* **77**, 167–178. <https://doi.org/10.1016/j.isatra.2018.04.005> (2018).
11. Kwak, J., Lee, T. & Kim, C. O. An incremental clustering-based fault detection algorithm for class-imbalanced process data. *IEEE Trans. Semicond. Manuf.* **28**, 318–328. <https://doi.org/10.1109/TSM.2015.2445380> (2015).
12. Liang, X., Zuo, M. J. & Feng, Z. Dynamic modeling of gearbox faults: A review. *Mech. Syst. Signal Process.* **98**, 852–876. <https://doi.org/10.1016/j.ymssp.2017.05.024> (2018).
13. Xiao, S. *et al.* Nonlinear dynamics of coupling rub-impact of double translational joints with subsidence considering the flexibility of piston rod. *Nonlinear Dyn.* **100**, 1203–1229. <https://doi.org/10.1007/s11071-020-05566-x> (2020).
14. Xiao, S., Liu, S., Jiang, F., Song, M. & Cheng, S. Nonlinear dynamic response of reciprocating compressor system with rub-impact fault caused by subsidence. *J. Vib. Control* **25**, 1737–1751. <https://doi.org/10.1177/1077546319835281> (2019).
15. Xiao, S., Liu, S., Song, M., Ang, N. & Zhang, H. Coupling rub-impact dynamics of double translational joints with subsidence for time-varying load in a planar mechanical system. *Multibody Syst. Dyn.* **48**, 451–486. <https://doi.org/10.1007/s11044-019-09718-9> (2020).
16. Han, H. *et al.* Fault feature analysis of planetary gear set influenced by cracked gear tooth and pass effect of the planet gears. *Eng. Fail. Anal.* **121**, 105162. <https://doi.org/10.1016/j.engfailanal.2020.105162> (2021).
17. Chen, Z., Zhu, Z. & Shao, Y. Fault feature analysis of planetary gear system with tooth root crack and flexible ring gear rim. *Eng. Fail. Anal.* **49**, 92–103. <https://doi.org/10.1016/j.engfailanal.2014.12.014> (2015).
18. Park, J. *et al.* Model-based fault diagnosis of a planetary gear: A novel approach using transmission error. *IEEE Trans. Reliab.* **65**, 1830–1841. <https://doi.org/10.1109/tr.2016.2590997> (2016).
19. Fan, L., Wang, S., Wang, X., Han, F. & Lyu, H. Nonlinear dynamic modeling of a helicopter planetary gear train for carrier plate crack fault diagnosis. *Chin. J. Aeronaut.* **29**, 675–687. <https://doi.org/10.1016/j.cja.2016.04.008> (2016).
20. Duan, T. *et al.* Detecting the 3D spatial varying crack evolution-induced vibration of gearbox through a system level rigid-flexible coupling model. *Mech. Mach. Theory* **174**, 104892. <https://doi.org/10.1016/j.mechmachtheory.2022.104892> (2022).
21. Zhiyi, H., Haidong, S., Lin, J., Junsheng, C. & Yu, Y. Transfer fault diagnosis of bearing installed in different machines using enhanced deep auto-encoder. *Measurement* **152**, 107393. <https://doi.org/10.1016/j.measurement.2019.107393> (2020).
22. Li, B., Tang, B., Deng, L. & Wei, J. Joint attention feature transfer network for gearbox fault diagnosis with imbalanced data. *Mech. Syst. Signal Process.* **176**, 109146. <https://doi.org/10.1016/j.ymssp.2022.109146> (2022).
23. Zhu, J., Chen, N. & Shen, C. A new deep transfer learning method for bearing fault diagnosis under different working conditions. *IEEE Sens. J.* **20**, 8394–8402. <https://doi.org/10.1109/JSEN.2019.2936932> (2019).
24. Wan, Z., Yang, R. & Huang, M. Deep transfer learning-based fault diagnosis for gearbox under complex working conditions. *Shock Vib.* <https://doi.org/10.1155/2020/8884179> (2020).
25. Han, T., Liu, C., Yang, W. & Jiang, D. Deep transfer network with joint distribution adaptation: A new intelligent fault diagnosis framework for industry application. *ISA Trans.* **97**, 269–281. <https://doi.org/10.1016/j.isatra.2019.08.012> (2020).
26. Zheng, C., Wang, X., Hao, Y., Wang, K. & Xiong, X. Normalized recurrent dynamic adaption network: A new framework with dynamic alignment for intelligent fault diagnosis. *IEEE Access* **8**, 80243–80255. <https://doi.org/10.1109/ACCESS.2020.2990572> (2020).
27. Deng, M., Deng, A., Shi, Y., Liu, Y. & Xu, M. Intelligent fault diagnosis based on sample weighted joint adversarial network. *Neurocomputing* **488**, 168–182. <https://doi.org/10.1016/j.neucom.2022.03.005> (2022).
28. Han, T., Liu, C., Wu, R. & Jiang, D. Deep transfer learning with limited data for machinery fault diagnosis. *Appl. Soft Comput.* **103**, 107150. <https://doi.org/10.1016/j.asoc.2021.107150> (2021).
29. Long, M., Wang, J., Ding, G., Sun, J. & Yu, P. S. Transfer feature learning with joint distribution adaptation. In *Proceedings of the IEEE International Conference on Computer Vision*. 2200–2207. <https://doi.org/10.1109/ICCV.2013.274> (2013).
30. Woo, S., Park, J., Lee, J.-Y. & Kweon, I. S. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*. 3–19. https://doi.org/10.1007/978-3-030-01234-2_1 (2018).
31. Feng, Z., Zhao, L. & Chu, F. Vibration spectral characteristics of localized gear fault of planetary gearboxes. *Proc. Chin. Soc. Electr. Eng.* **33**, 119–127 (2013).
32. Sun, B. & Saenko, K. Deep coral: Correlation alignment for deep domain adaptation. In *European Conference on Computer Vision* 443–450 (Springer, 2016). https://doi.org/10.1007/978-3-319-49409-8_35.
33. Tzeng, E., Hoffman, J., Zhang, N., Saenko, K. & Darrell, T. Deep domain confusion: Maximizing for domain invariance. *arXiv:1412.3474* (arXiv preprint). <https://doi.org/10.48550/arXiv.1412.3474> (2014).
34. Ganin, Y. & Lempitsky, V. Unsupervised domain adaptation by backpropagation. In *International Conference on Machine Learning*. 1180–1189 (PMLR). doi: <https://doi.org/10.48550/arXiv.1409.7495> (2015).

Acknowledgements

This paper was supported by the following research projects: the Natural Science Foundation of Fujian Province (Grants no. 2022J011223 and no. 2020J01432), the Key Technology Innovation Project of Fujian Province (Grant no. 2022G02030), Youth and Middle-aged Science and Technology Project of Ningde Normal University (no. 2022ZQ102), Innovation Team of Ningde Normal University (no. 2020T02), and Engineering Research Center of Mindong Aquatic Product Deep-Processing, Fujian Province University. These supports are gratefully acknowledged.

Author contributions

J.-H.Z. and S.-G.X. conceived and designed the study. M.-M.S. and Y.-H.T. contributed to data acquisition and analysis. Z.-C.X. were responsible for the model building. M.-M.S. and Z.-C.X. contributed to writing of original manuscript. J.-H.Z. and S.-G.X. were responsible for revising and reviewing. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.-H.Z. or S.-G.X.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022