Digital twin-driven intelligent assessment of gear surface degradation

Ke Feng[1,2], J.C. Ji[3], Yongchao Zhang[1], Qing Ni[3], Zheng Liu[1,*], Michael Beer[4,5,6]

1. *School of Engineering, University of British Columbia, Kelowna, BC V1V 1V7, Canada*
2. *School of Mechanical and Manufacturing Engineering, University of New South Wales, Sydney, NSW 2052, Australia*
3. *School of Mechanical and Mechatronic Engineering, University of Technology Sydney, Ultimo, NSW 2007, Australia*
4. *Institute for Risk and Reliability, Leibniz Universität Hannover, Hannover, Germany*
5. *Institute for Risk and Uncertainty, University of Liverpool, Liverpool, United Kingdom*
6. *International Joint Research Center for Resilient Infrastructure & International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai, China*

**corresponding author: zheng.liu@ubc.ca*

Abstract

Gearbox has a compact structure, a stable transmission capability, and a high transmission efficiency. Thus, it is widely applied as a power transmission system in various applications, such as wind turbines, industrial machinery, aircraft, space vehicles, and land vehicles. The gearbox usually operates in harsh and non-stationary working environments, expediting the degradation process of the gear surface. The degradation process may lead to severe gear failures, such as tooth breakage and root crack, which could damage the gear transmission system. Therefore, it is essential to assess the progression of gear surface degradation in order to ensure a reliable operation. The digital twin is an emerging technology for machine health management. A high-fidelity digital twin model can help reflect the operation status of the gearbox and reveal the corresponding degradation mechanism, which could benefit the remaining useful life (RUL) prediction and the predictive maintenance-based decision-making framework. This paper develops a digital twin-driven intelligent health management method to monitor and assess the gear surface degradation progression. The developed method can effectively reveal the gear wear propagation characteristics and predict the RUL accurately. Furthermore, the knowledge learned from digital twin models can be well transferred to the

surface wear assessment of the physical gearbox in wide industrial applications, which is of great practical significance. Two endurance tests with different dominant degradation mechanisms were conducted to validate the effectiveness of the proposed methodology for gear wear assessment.

Keywords: gearbox, digital twin, surface degradation, health management, wear assessment

1 Introduction

Thanks to its compact structure, stable transmission performance, and high transmission efficiency, the gearbox has been widely utilized as a crucial transmission system in various industrial applications, such as wind turbines, vehicles, machinery, and aircraft [1-4]. However, the gearbox transmission system usually operates under harsh conditions such as inevitable fluctuating speeds and variable applied loads [5]. The adverse working environment expedites the degradation process of the gearbox transmission system. The gear degradation would lead to severe damage to gears, including gear tooth crack, gear surface spalling, and gear tooth breakage, all of which could destroy the transmission system. Thus, the health status of the gearbox has a significant impact on the safe operation of machinery and its corresponding production efficiency [6]. Therefore, it is crucial to assess the gear surface wear propagation process so that reliable predictive maintenance can be scheduled in advance to avoid the sudden shutdown of the gear transmission system.

Gear surface wear is an inevitable degradation behavior in the gearbox lifespan [7-11]. Abrasive wear and fatigue pitting are two typical gear wear mechanisms. The propagation of abrasive wear would reduce the tooth thickness of meshing gear pairs and increase the risk of gear tooth breakage, which may cause the sudden shutdown of the gear transmission system. Fatigue pitting usually exists in the transmission system with lubrication, and its propagation can significantly reduce the durability of the gear surface and affect transmission efficiency. In practice, there is a complex dynamic interaction between abrasive wear and fatigue pitting, and

this dynamic interaction brings a great challenge to the health management of the gearbox transmission system [9, 12, 13]. The development of gear wear monitoring techniques would bring significant benefits to industry practices.

Up to the present, the wear particle analysis technique has been the most widely utilized tool for gear wear analysis [14, 15]. The gearbox's degradation status can be identified by analyzing the particle size, concertation, and distribution. Nevertheless, in general, wear particle analysis is an offline technique. Also, it is laborious and costly to collect wear particles from the machinery under operation. In addition, some techniques use coordinate measurement to achieve gear wear monitoring. For example, a coordinate measurement machine was utilized in [16] to evaluate the actual gear tooth profile; these measurements were combined with the dynamic model and Archard wear model for gear wear monitoring. It should be mentioned that the gear tooth profile assessment was implemented once at the healthy condition in [16], and the coordinate measurement machine was not applied for real-time gear wear monitoring. The reason might be that this coordinate measurement technique requires an interruption of the normal gear transmission system. In addition, the Klingelnberg P40 (Hückeswagen, Germany) coordinate measuring machine and TalyScan 150 scanning instrument by Taylor Hobson (Leicester, UK) were applied in [17] for assessing the worn tooth surface. Unfortunately, this process also requires the dismantling of the gearbox, and it could not be used for online gear wear monitoring. A Talysurf Intra 50 profilometer was utilized in [18] to acquire the gear tooth profiles; however, due to the installation requirement of the profilometer, applying it for real-time gear wear monitoring is challenging. Therefore, it is worthwhile to develop online machine condition monitoring techniques which can efficiently and effectively reveal the degradation progression of the gearbox transmission system.

As a popular tool for real-time gear health management, vibration analysis has attracted considerable attention from the research community and the industry [19-22], and it has been

utilized to monitor gear wear propagation [23]. For instance, it was found that the gear wear propagation could bring increments to the gear meshing harmonics, and thus the magnitude of the gear meshing harmonics (higher-order) was applied in [24] to assess the propagation process of gear wear. Amplitudes of the quefrencies of cepstrum and gear meshing harmonics were used in [25] to assess gear wear progression. In industrial practice, the propagation process of gear wear is highly complicated. An intense competition can exist among multiple wear mechanisms (such as fatigue pitting and abrasive wear) during the wear progression [26-28]. As a result, the changes in gear mesh harmonics are not determined, although the average meshing gear tooth profile geometry stays deviating from the initial one during the whole wear process. More specifically, some specific gear meshing harmonics may have an increasing trend in a certain period, but a decreasing trend might appear in the subsequent period. Two new vibration indicators were developed in [29] to consider all the possible meshing harmonics of gear tooth that demonstrate noticeable changing behaviors. The effectiveness of these two indicators in gear wear monitoring was validated by two run-to-failure tests with different initial tooth surface morphologies.

Apart from the wear monitoring techniques developed on the basis of the gear meshing harmonics, some other analysis methodologies were also employed or developed for monitoring gear wear progression. For instance, the ability of RMS was investigated to evaluate the gear wear progression (mainly induced by gear fatigue pitting) in [30]. FM4 and NA4 were applied in [31] to monitor the gear system degradation behavior caused by wear. Interestingly, the results of studies [29, 32, 33] contradicted the ability of FM4 and NA4 to assess gear wear propagation. The underlying reason is that FM4 and NA4 were initially developed for the detection of local gear faults (such as gear root crack) [29], while gear wear is generally uniformly distributed on each gear tooth, which is different from the localized failures. Thus FM4 and NA4 might have a limited capability in monitoring gear wear propagation. Later,

modulation bispectrum of vibration and motor current signals was applied in [33, 34] to monitor gear wear propagation, and its effectiveness was validated using experiments. In addition, the correlation analysis was applied on the residual signal [32] to evaluate the gear wear propagation.

Based on the above discussions, it can be found that the gear surface wear monitoring techniques are yet to be well established and fully developed. The internal relationship between the underlying gear meshing physics and the unique gear wear features has not been well understood. Thus, it might be challenging to apply the current gear wear monitoring methodologies to track the surface degradation progression and fully reveal its complex degradation mechanisms.

Digital twin (DT) is an emerging technology to address the above-mentioned issues. DT is a virtual representation (mirror) of a physical structure along its lifecycle. Through real-time interaction between the virtual model and physical structure, the degradation status of the system and its RUL can be reflected and evaluated effectively. Thanks to its unique feature, DT has received substantial attention from the research community over the last decades. For example, a digital twin framework was developed in [35] to achieve structure damage detection, and the simulation signal was applied to demonstrate its performance. A digital twin-based toolbox was designed in [36] to identify high-value information from the data with high uncertainty. Also, a transfer learning-based digital twin was developed to detect the localized torsional friction in deviated wells [37]. Furthermore, with the help of the data-driven dynamic model updating and optimization-based operating condition estimation, a digital twin approach was developed in [38] for the on-load tap changers' health state estimation. However, due to the complex structures and harsh operation conditions, research on DT-based condition monitoring of mechanical transmission systems is rare. Moreover, the existing conceptual approaches [39-41] have limitations in indicating the specific contact status and providing

insights into the degradation stage of mechanical transmission systems, which are of great value to health management and RUL prediction. Therefore, it is necessary to develop DT-based methodologies for machinery health management.

However, there are some critical issues which impede the digital twin techniques from widely applied in gearbox health management, including gear wear monitoring. The first challenge is the establishment of a high-fidelity digital twin model. The gearbox transmission system usually consists of shafts, coupling, bearings, and gears. For each component, some properties, such as mass, inertia, stiffness, and damping coefficient, may vary under different operating conditions. Thus, it is challenging to tune these parameters manually to realize a high coherence between the simulated response and the realistic response from the physical structure. The second challenge is the practicability of the digital twin model in gear wear assessment. The gear wear propagation process is highly complex. The dynamic interactions between gear surface wear and gear dynamic response, lubrication conditions, and gear surface morphology are continuously changing in the gear lifespan. Accordingly, some inevitable feature differences exist in the responses of the digital twin model and gearbox physical structures during the gear wear progression. Also, the manufacturing, installation, and operation process would render the health status/performance of each gearbox different, even though they are designed and manufactured for the same. These issues would impede the digital twin techniques for wide applications in industrial practices.

To address the aforementioned issues, a digital twin-driven intelligent health management method is developed to assess gear wear progression critically in this paper. The highlights and contributions of this work are summarized as follows:

- High-fidelity digital twin models are established with the help of model updating. An intelligent approach is developed to build the digital model by interacting with the measurements from physical structures. This approach can eliminate the errors and

deficiencies induced by manually tuning the models, and it can ensure a high coherence between the simulated response and the realistic response from the physical structure.

- A transfer learning algorithm is developed and utilized to transfer the knowledge from the digital twin model and apply the learned knowledge to assess the surface wear of physical structure. This transfer learning algorithm can handle the feature differences that exist in the responses of the digital twin model and gearbox physical structures, increasing the practicability of the digital twin techniques for gear health management in wide industrial applications.

In summary, the developed digital twin-driven gear wear assessment method can effectively reveal the gear surface degradation status and predict the RUL accurately. Also, the knowledge learned from digital twin models can be well transferred for the surface wear assessment of the physical gearbox, which is of significant practical value. Two run-to-failure tests (with different dominant degradation mechanisms) are arranged to verify the effectiveness of the developed methodology for gear system health management.

The organization of this paper is as follows: Section 2 describes the digital twin-driven gear wear monitoring methodology, including the establishment of the high-fidelity digital twin model and the development of the transfer learning algorithms for surface degradation assessment of the physical gearbox. The surface degradation assessment results are presented and discussed in Section 3 with the help of the endurance tests. Conclusions and recommendations for further research are given in Section 4.

2 Methodology

In this section, the digital twin-driven intelligent gear surface degradation assessment method is developed. Also, the details of the high-fidelity digital twin model establishment and transfer learning algorithm development are presented.

2.1 Digital twin-driven gear wear propagation assessment methodology

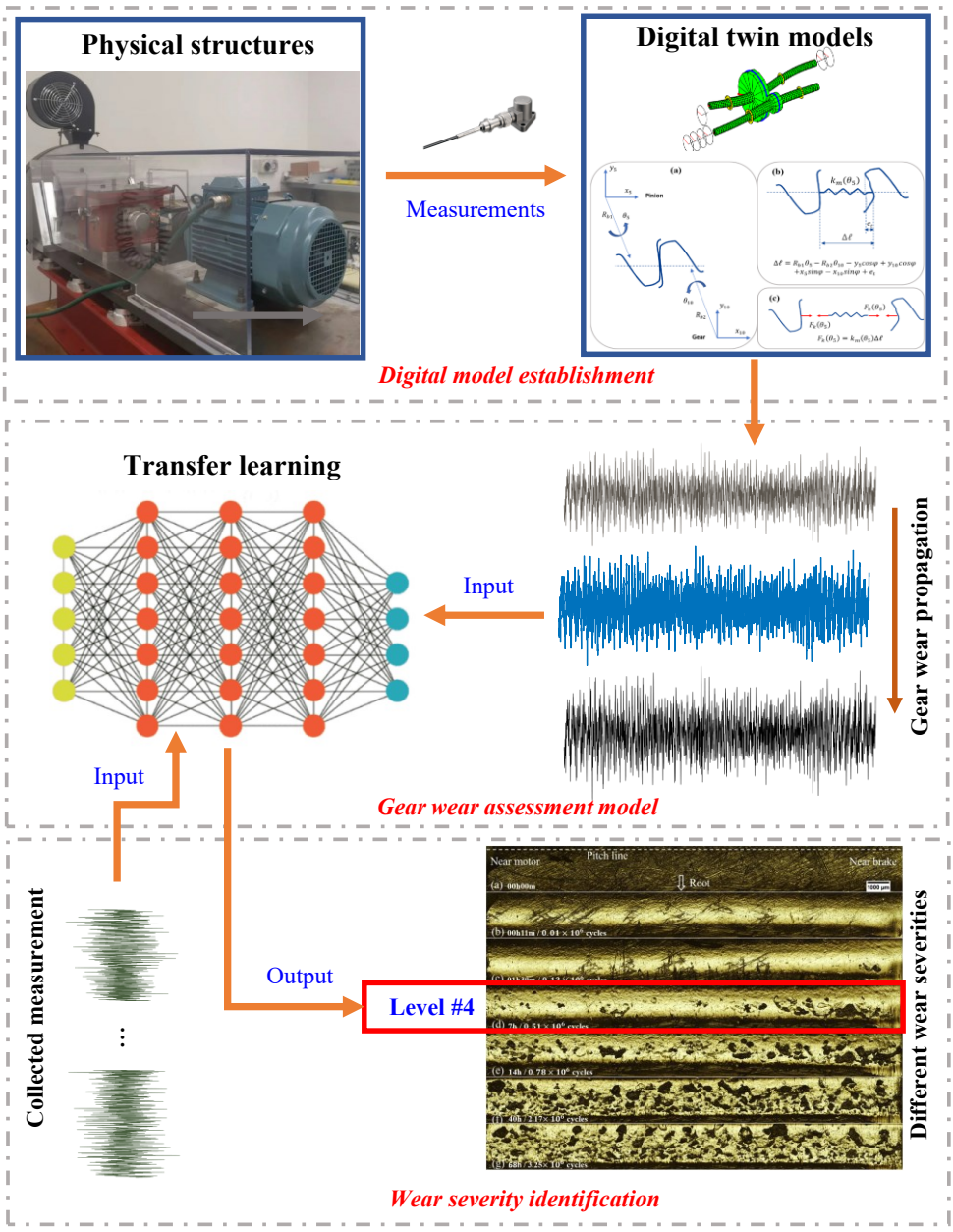

Figure 1 Basic procedure of the developed digital twin-based gear wear monitoring methodology

The basic procedure of the digital twin-driven gear wear monitoring scheme is illustrated in Figure 1. The digital twin-driven gear wear monitoring methodology mainly consists of two parts: 1) establishing a high-fidelity digital twin model with the assistance of the measurements from physical structures; and 2) developing a transfer learning algorithm for gear wear assessment. With the proposed digital twin-driven methodology, the gear wear severity can be accurately assessed non-destructively by inputting the measurement data collected during the

process of gear surface wear propagation. It can provide valuable information on the gear system degradation status so that proper predictive maintenance can be scheduled to guarantee the safe operation of the gearbox transmission system.

It should note that the high-fidelity digital twin models will be established based on the spur gearbox test rig at the University of New South Wales (UNSW), as shown in Figure 2. Also, the measurements from the physical test rig can be applied to bridge the digital twin models and physical structures.

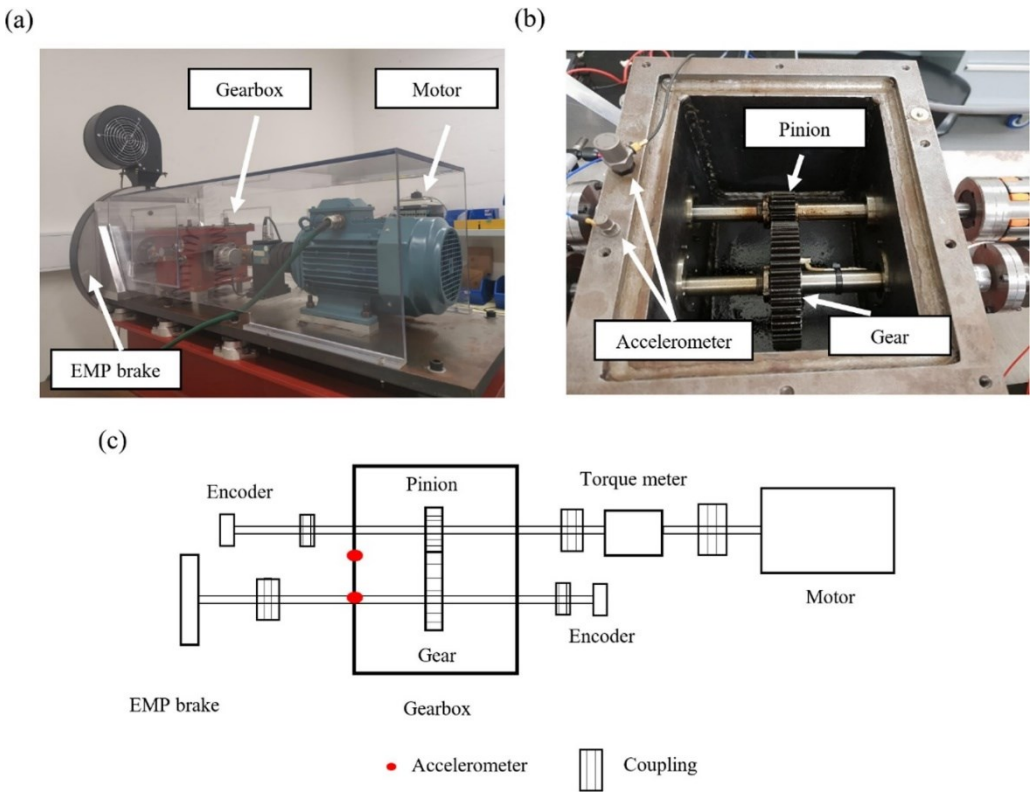

Figure 2. The spur gear test rig at UNSW. (a) Overall view of the test rig; (b) Gear meshing and accelerometers location; (c) The schematic diagram of the experimental test rig setup

In the spur gearbox transmission system, two modular gears with teeth numbers of 19 and 52 are installed. Mild steel (JIS S45C) is selected as the gear material to accelerate the wear propagation. An electric motor drives the input shaft, and a variable frequency drive (VFD) is applied to control its instantaneous speed. The applied load of the gear transmission system is controlled by an electromagnetic particle brake (EMP). Two encoders with 512 pulses per revolution are installed at the remaining free ends of the shafts. The lubrication is provided by

an oil bath, and the kinematic viscosity of the lubricating oil is 146 mm^2/s. Two accelerometers (B&K 4394 and B&K 4396) are placed on the gearbox casing, as shown in Figure 2(b). The following section will introduce the procedure of building the high-fidelity gear transmission digital twin model in detail.

2.2 A high-fidelity digital twin model for gear transmission system

For gear wear assessment, comprehensive dynamic digital twin models, including the dynamic model, the Archard gear wear model, and the fatigue pitting propagation model, are required to be established and calibrated intelligently. These high-fidelity digital twin models can represent the response of the physical structure and reveal the degradation mechanism. Also, the digital twin models can be applied to generate the system degradation data with low cost and high computation efficiency, which could benefit the degradation monitoring and RUL prediction of the gearbox transmission system. The intelligent construction processes of these high-fidelity digital twin models are introduced as follows.

2.2.1 Gear transmission dynamic model

A 21-degree-of-freedom (DOF) comprehensive dynamic model is built according to the spur gearbox test rig (Figure 2). As shown in Figure 3, the dynamic model consists of the gears, motor, coupling, shafts, and casing. The motion equation of the gearbox transmission system can be written as

$$\mathbf{M}\ddot{\mathbf{x}} + \mathbf{C}\dot{\mathbf{x}} + \mathbf{K}\mathbf{x} = \mathbf{f} \tag{1}$$

where $\mathbf{x}$ denotes the angular and translational displacements of different nodes of the gear transmission system, and it is in the plane perpendicular to the connecting shaft axes [42].

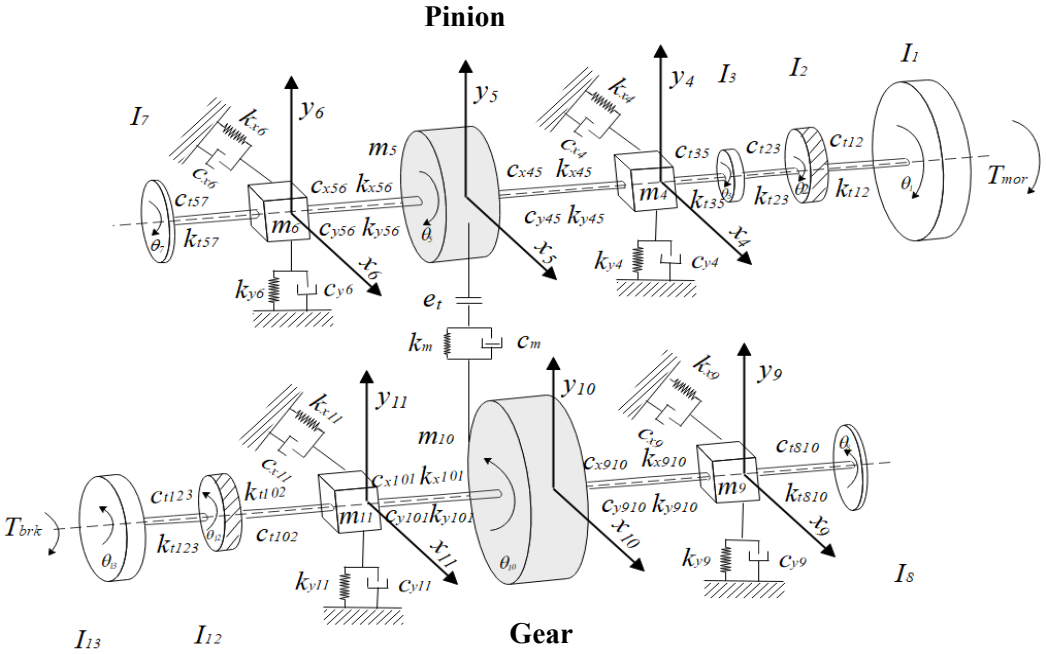

Figure 3 A 21 DOF gear dynamic model, in which θ_i and (x_i, y_i) refer to the torsional and translational displacements, respectively

The model parameters are generally determined by the analysts' prior experience or empirical formulas. However, the installation process and operational conditions could amend the component properties, resulting in the parameter values deviating from the nominal ones, especially for the damping coefficients. Also, manually fine-tuning the model parameters is time-consuming, and the accuracy cannot be guaranteed.

To ensure the responses of the gear transmission system can be accurately reflected by the established dynamic model, based on the digital twin framework, a novel model establishment approach is developed in this paper to calibrate the dynamic model intelligently. In the developed approach, the measured data from the gear system are applied to calibrate the dynamic model.

The basic procedure of the intelligent model updating approach is demonstrated in Figure 4. In the developed approach, multiple sources are considered, such as the signal characteristics in the time- and frequency- domains and the natural frequencies of the gear systems. In order to obtain the optimal values of parameters of the dynamic model to ensure the simulated responses

match well with the measured ones, the optimization algorithm will be employed to update the model parameters.

In the model calibrating process, multiple parameters need to be determined and updated, and also the simulations should match the physical measurements in multiple aspects. The mathematical expressions of the objective functions during the model updating process are expressed as:

$$\begin{cases} f_1\left(V_{paras_{optimal}}\right) = 1 - correlation(X^{MOD}, Y^{PHY}) \\ f_2\left(V_{paras_{optimal}}\right) = 1 - correlation(FFT^{MOD}, FFT^{PHY}) \\ f_3\left(V_{paras_{optimal}}\right) = \sum_{\iota=1}^{8} \left(f_{natural_\iota}^{MOD} - f_{natural_\iota}^{phy}\right)^2 \end{cases} \tag{2}$$

$$\left.\begin{array}{ll} \text{minimize: } \left[f_1\left(V_{paras_k}\right), f_2\left(V_{paras_k}\right), f_3\left(V_{paras_k}\right)\right] & \\ \text{subject to: } bd_i^L \le bd_i \le bd_i^U & i = 1,2,\cdots,n \end{array}\right\} \tag{3}$$

where $V_{paras_{optimal}}$ are the optimal parameter values after model calibrating/updating. The superscript MOD denotes the dynamic mode, and PHY indicates the physical gear system. Pearson's correlation is applied in this paper and is denoted as $correlation$. X and Y are the time waveforms from the dynamic model and physical gear system, respectively. FFT indicates the Fourier transform. $f_{natural}$ is the natural frequency. Note that there is no strict rule on the number of natural frequencies in Eq. (2). It can be adjusted based on the specific applications and requirements. In this paper, the first eight natural frequencies will cover the first five gear meshing harmonics (the input shaft speed is 20 Hz), which are usually used for gear condition monitoring. Therefore, the first eight natural frequencies are selected and used in Eq. (2).

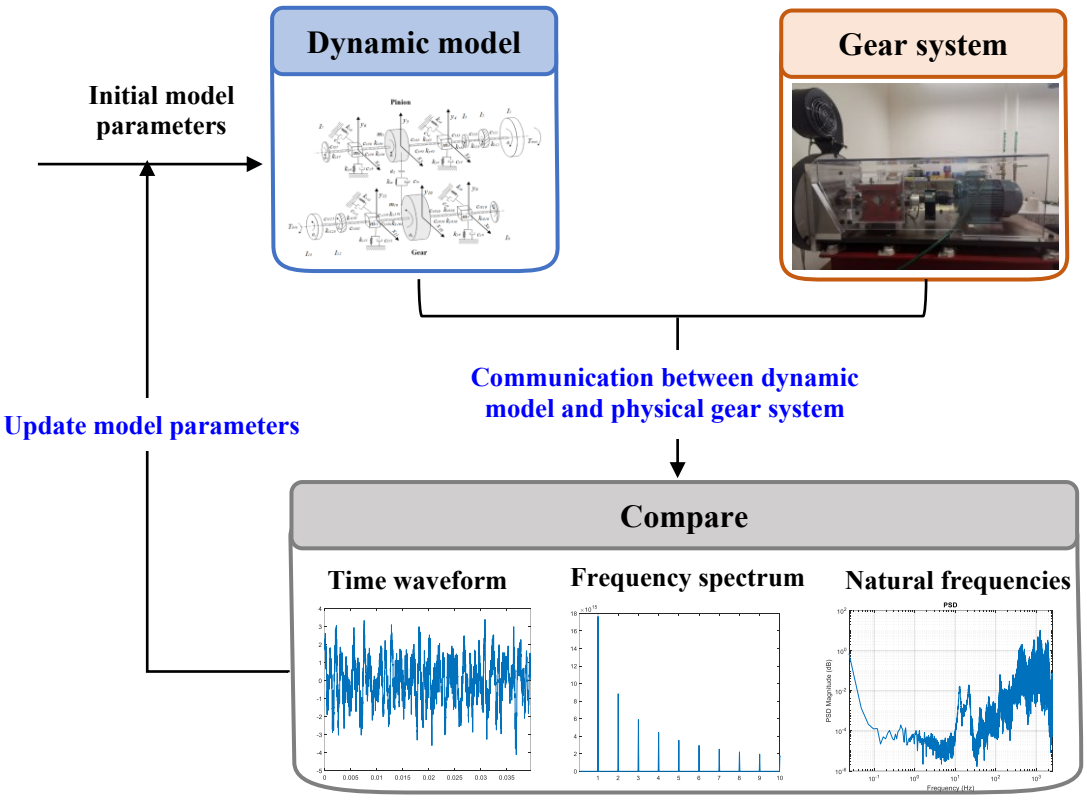

Figure 4 Communication between dynamic model and physical measurements

The multiobjective grasshopper optimization algorithm (MOGOA) has a high exploration capacity and breakneck convergence speed. The MOGOA is developed based on the grasshopper optimization algorithm (GOA) and is applied to find the optimal model parameters. As a metaheuristic bionic optimization algorithm, GOA simulates the grasshoppers' swarming behavior in nature to search for the optimal [43]. The grasshopper's life cycle includes two phases, namely, nymph and adulthood. For the nymph, the main swarming behavior characteristics are small steps with slow motion. While for adulthood, the main swarming behavior characteristics are long steps with abrupt movement. Thus, the bionic principle of GOA is to map the small steps of nymph to local exploitation with slow motion and the long steps of adulthood to global exploration with abrupt movement. The mathematical model of the swarming behavior of grasshoppers is

$$P_i = S_i + G_i + A_i \tag{4}$$

where P_i denotes the i^{th} grasshopper's position, S_i represents the social interaction between grasshoppers, G_i is the gravity force on the i^{th} grasshopper, and A_i indicates the wind advection.

The social interaction S_i between grasshoppers is

$$S_i = \sum_{j=1;j\neq i}^{K} s(|P_j - P_i|)\frac{P_j - P_i}{|P_j - P_i|} \tag{5}$$

K is the grasshopper number used for optimization, $|P_j - P_i|$ denotes the Euler distance between the i^{th} and j^{th} grasshoppers, $\frac{P_j - P_i}{|P_j - P_i|}$ is the unit vector from the i^{th} to j^{th} grasshoppers, $s(\cdot)$ represents the social force:

$$s(r) = fe^{\frac{-r}{l}} - e^{-r} \tag{6}$$

where f and l are the intensity of attraction and the length scale of attraction, respectively. If $s(r) > 0$, it indicates there is an attraction between grasshoppers. On the contrary, there is a repulsion between grasshoppers if $s(r) < 0$. And $s(r) = 0$ means no attraction and repulsion appear between grasshoppers, which is a comfort area. The value of $s(r)$ is determined by the combinations of f and l. Based on the suggestions in [43], the values of $f = 0.5$ and $l = 1.5$ are used in this paper.

However, employing the above-mentioned model could not make the solution converge to a specified point when arriving at the comfort area. An improved model is thus established to solve this problem:

$$P_i = c\left(\sum_{j=1;j\neq i}^{K} c\frac{u_b - l_b}{2}\sum_{j=1;j\neq i}^{K} s(|P_j - P_i|)\frac{P_j - P_i}{|P_j - P_i|}\right) + G_i + A_i \tag{7}$$

$$c = c_{max} - m\frac{c_{max} - c_{min}}{\mathcal{M}} \tag{8}$$

where u_b and l_b denote the upper and lower bounds for the optimization, m indicates the current iteration, $\mathcal{M}$ is the total iteration, and c is the decreasing coefficient (linearly decreases

from c_{max} to c_{min}). It should be mentioned that two $c's$ are used in Eq. (7), which are the external c and internal c. The two $c's$ aim to reduce the search coverage and social interaction over the increasing iteration, respectively.

By incorporating an archive and target selection technique into the GOA, the optimal solution can be estimated through the MOGOA. The approach for updating the target is the major difference between GOA and MOGOA. Unlike the GOA, which selects the target using the best solution acquired until now, MOGOA can achieve more than one solution. The following procedures are executed to select the optimal solution from the multiple solutions [44]: 1) calculate the Pareto optimal solutions and save these solutions in an archive; 2) integrate a roulette to select the optimal one, which can well balance all the objectives, as introduced in Eq. (9):

$$V_{paras_{optimal}} \rightarrow min\left(Pro_i = \frac{1}{\hbar_i}\right) \tag{9}$$

where $\hbar_i$ is the number of neighboring solutions of the i^{th} solution in the archive.

From Eq. (2) to Eq. (9), the optimal model parameters can be determined. In the process of determining the optimal parameters for the dynamic model, multiple measurements from various aspects (including time waveform, frequency spectrum, and natural frequencies) are utilized. This guarantees that the optimal parameters can still be obtained for the dynamic model by the multiobjective grasshopper optimization algorithm, even though some noise or errors exist in the time waveform, frequency spectrum, or the estimated natural frequencies. The optimization-assisted model calibration/updating approach can guarantee reliable outputs from the dynamic model, such as the vibration signal, transmission error, and dynamic contact force. Also, the dynamic model can duplicate the real-time dynamic behavior by communicating between the digital twin model and the physical measurements.

To conclude, this digital twin framework can help achieve intelligent model calibration and updating, which can uncover the dynamic characteristics of the gear system in real-time, benefiting the gear performance assessment and RUL prediction.

2.2.2 Gear transmission degradation models

In practice, the gearbox transmission usually operates under normal conditions. Proper routine maintenance or regular replacement is usually scheduled for safety operations before severe degradation occurs. Thus, the data labelled with various degradation severities are relatively rare compared with the measurements under healthy working conditions. The data with different degradation severities are invaluable for the RUL prediction of the gear transmission system and its health management. However, conducting endurance tests to generate the system degradation data is costly and time-consuming. To tackle the problem, degradation models are used to produce signals with well-defined degradation characteristics rather than waiting until such signals arise randomly.

In simulating the degradation behaviors, the Archard wear model is the most widely used degradation model. However, in the Archard wear model, the Hertzian contact-induced deformation is not fully considered, which could impair its performance in modeling abrasion characteristics. To address this issue, an improved Archard wear model was developed in our recent publication [45]:

$$h(\theta, t) = \int K_{wear} P(\theta, t) v(t) dt \tag{10}$$

where $h(\theta, t)$ represents the wear depth at rotation angle θ and time t, $P(\theta, t)$ denotes the Hertzian contact pressure distribution, $v(t)$ represents the sliding velocity of the engaging gear-pair, and K_{wear} is the wear coefficient. Unlike the existing research, the Archard wear model, after improvement, considers the influences coming from adjacent gear contacting

points that contribute to the wear depth accumulation. As a result, a smoother surface wear distribution curve can be achieved, and it is well approximated to worn gear tooth profiles that are observed in practice. The improved Archard wear model will be employed in this paper to simulate the degradation behavior induced by abrasive wear progression.

Fatigue pitting is another common surface degradation phenomenon; however, less research has been conducted on developing the model for simulating the surface degradation progression. To reveal the fatigue pitting propagation characteristics, a novel fatigue pitting model was recently developed in [46]. The details of the developed fatigue pitting model are as follows:

The accumulated fatigue pitting density D_{i+1} at $N + \Delta N$ running cycles is

$$D_{i+1} = D_i + \left(\exp\left(-\frac{K_{pitting} \times \tau_0(\theta)^c \times V(\theta) \times N^m}{z_0(\theta)^h}\right) - \exp\left(-\frac{K_{pitting} \times \tau_0(\theta)^c \times V(\theta) \times (N + \Delta N)^m}{z_0(\theta)^h}\right)\right) \quad (11)$$

where D_i is the fatigue pitting density corresponding to N running cycles, $K_{pitting}$ is pitting coefficient, and $\tau_0(\theta)$, $z_0(\theta)$ and $V(\theta)$ are the maximum shear stress, the depth of the maximum shear stress, and stressed volume, respectively (see Eqs. (12-14)):

$$\tau_0(\theta) = 0.3 \times P_{max}(\theta) \quad (12)$$

$$z_0(\theta) = 0.786 \times b(\theta) \quad (13)$$

$$V(\theta) = 2b(\theta) \times z_0(\theta) \times 2\pi \times R(\theta) \quad (14)$$

where P_{max} denotes the maximum Hertzian contact pressure corresponding to the rotation angle θ, R is the combined radius of curvature of the contact point, and b is Hertzian contact width.

Through integrating the calibrated dynamic model and the degradation models (abrasive wear model and fatigue pitting model), the signals with various degradation severities can be simulated, and a degradation archive can be obtained, which is meaningful to the gear system

performance assessment and its RUL prediction. The intergradation of the dynamic and degradation models is not the main focus of this paper; therefore, it will not be further introduced. More details on the model intergradation can refer to [46].

It should be mentioned that the model calibration process (as shown in Figure 4), based on the time waveforms, frequency spectra, and natural frequencies, is only applied to the dynamic model. The dynamic model calibration can ensure the dynamic model under healthy condition function well, and the realistic response can be generated from the dynamic model compared with the measurement from the physical structure. The gear digital twin model is achieved by integrating the well-calibrated dynamic model and degradation models. The introduction of degradation models will bring inevitable feature differences in responses between the digital twin model and physical structures, which are caused by the degradation rate change in engineering practice. The transfer learning algorithm will address this discrepancy instead of real-time updating based on the physical measurement, which is more practical in industrial practices. The proposed transfer learning algorithm will be introduced in the following section.

2.3 A transfer learning algorithm for gear wear severity assessment

2.3.1 Domain adaptation model development

In order to transfer the degradation knowledge of the simulation data from the digital twin model to the actual evaluation process, a domain adaptation model based on deep learning is built in this paper. In the domain adaptation model, labeled simulation data is called source domain data, and it is represented by $\mathcal{D}^s = \{\mathbf{x}_i^s, \mathbf{y}_i^s\}_{i=1}^{n_s}$, where $\mathbf{x}^s$ and $\mathbf{y}^s$ denote the source domain samples and corresponding labels, respectively. Moreover, actual unlabeled measured data is named the target domain data, and it is expressed as $\mathcal{D}^t = \mathbf{x}_i^{t\,n_t}{}_{i=1}$, where $\mathbf{x}^t$ represents target domain samples. The architecture of the domain adaptation model is shown in Figure 5, which consists of a feature extractor G and a fully connected classifier C. During the model

training process, the simulated source domain data and the measured target domain data are first fed into the feature extractor simultaneously to obtain high-level features. Then, the high-level features are processed through the classifier to obtain the predicted output of the model. The label data of the source domain is used to obtain the performance of extracting discrepancy features between different degradation conditions. In high-level feature layers, losses based on global domain alignment and class alignment are designed to achieve feature domain invariance between the simulated and measured data.

Within the deep learning architecture, the convolutional neural network (CNN) has many unique advantages [47, 48], including strong nonlinear feature extraction ability, easy construction, and easy training. Thus, a CNN-based feature extractor G is constructed in this study. In the feature extractor G, five convolution modules and a fully connected module are built for the feature extraction of source and target domain data. In each convolution module, there are four operations, namely convolution (Conv), batch normalization (BN) layer, rectified linear units activation function (ReLu) layer, and maximum pooling (Pooling).

The convolution module has fully connected operations, BN and ReLu. The first convolution module receives a one-dimensional raw vibration signal as input, and high-level features of the input data are obtained after five convolution modules and a fully connected module. In the high-level feature layer, a maximum mean discrepancy (MMD) loss and a class contrastive loss are built for domain-invariant learning of source and target domains. Finally, the high-level features of the source and target domains are input into the classifier C to obtain the predicted output.

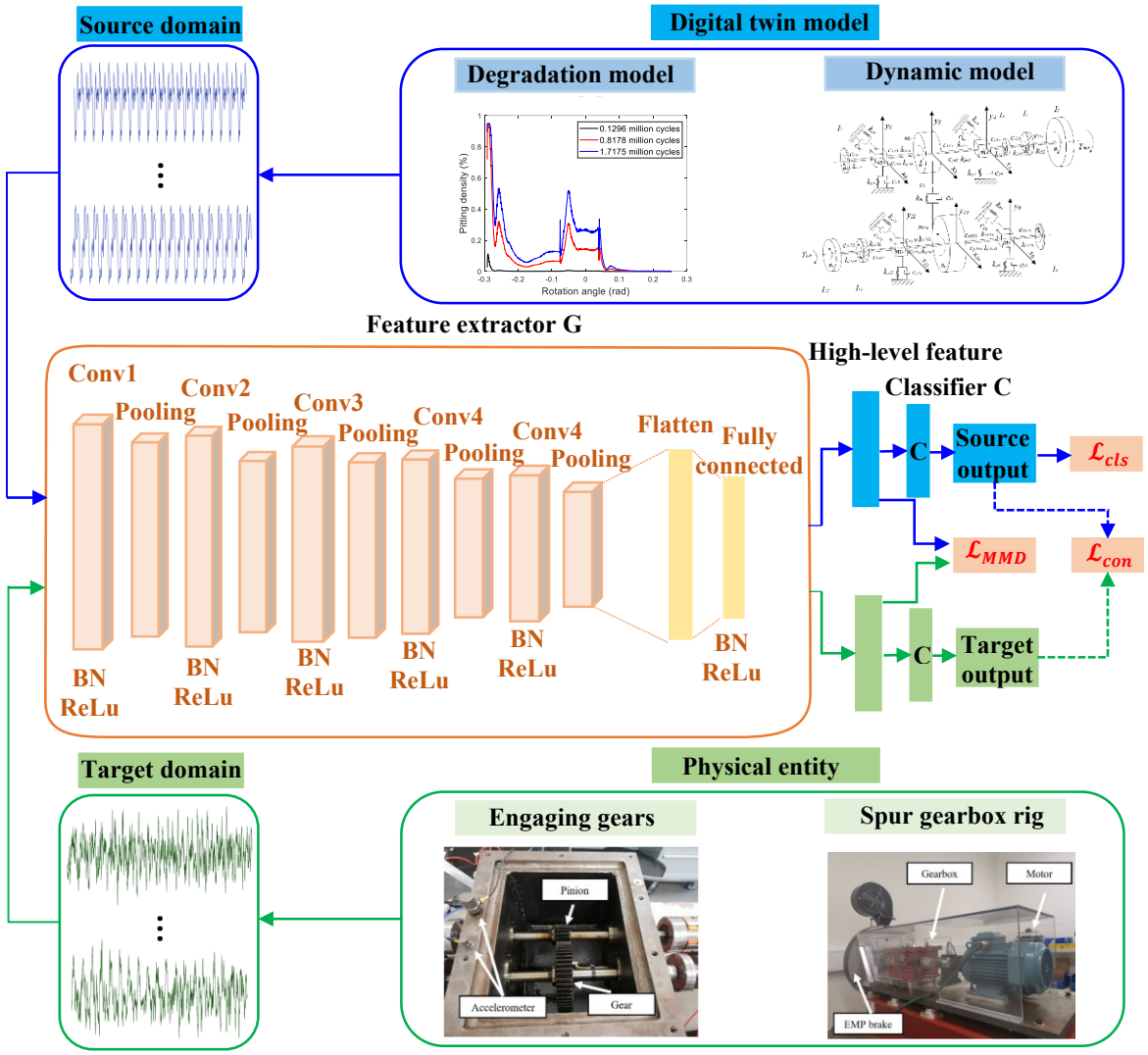

Figure 5 The architecture of the proposed domain adaptation model.

2.3.2 Domain adaptation model optimization

From Figure 5, it can be seen that three crucial parameters are introduced in the domain adaptation model to achieve accurate wear assessment results, which are $\mathscr{L}_{cls}$ (assessment loss), $\mathscr{L}_{MMD}$ (MMD loss), and $\mathscr{L}_{con}$ (class contrastive loss). In the following, these three parameters will be discussed in detail.

(1) Assessment loss

To extract discriminative features of the different degradation conditions, the cross-entropy loss is adopted as the loss function. The cross-entropy loss is defined as:

$$\mathscr{L}_{cls}(x^s, y^s) = -E_{(x^s,y^s)\in\mathcal{D}^s}\left[\sum_{k=1}^{K} \mathrm{I}_{[k=y^s]}\, log(\, C(G(x^s)))\right] \tag{15}$$

where K is the number of health conditions.

(2) MMD loss

Due to a large discrepancy in feature distribution between the simulated data and the measured data, direct use of the deep learning model trained on the simulated signal for the degradation condition assessment of the measured signal could result in poor performance. Therefore, reducing the domain shift in the feature layer is a necessary operation that can significantly improve the performance of deep learning models. In this work, the MMD loss is first introduced in the high-level feature layer to achieve global domain alignment. MMD can measure the distribution difference between two domains by comparing test statistics [49]. The MMD loss of source domain and target domain in the high-level feature layer can be defined as [50]:

$$\begin{aligned}\mathscr{L}_{MMD}(x^s, x^t) &= \left\| \frac{1}{n_1}\sum_{i=1}^{n_1} f(x_i^s) - \frac{1}{n_2}\sum_{j=1}^{n_2} f\left(x_j^t\right) \right\|_{\mathscr{H}}^2 \\ &= \frac{1}{n_1^2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_1} k\left(x_i^s, x_j^s\right) - \frac{2}{n_1 n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} k\left(x_i^s, x_j^t\right) + \frac{1}{n_2^2}\sum_{i=1}^{n_2}\sum_{j=1}^{n_2} k\left(x_i^t, x_j^t\right)\end{aligned} \tag{16}$$

where $\mathscr{H}$ represents the reproducing kernel Hilbert space (RKHS), $f: x^s, x^t \to \mathscr{H}$.

(3) Class contrastive loss

MMD loss only achieves global domain alignment, and there may be some misidentified conditions at the decision boundary of recognition due to the large domain shift between simulation data and test data. Thus, a class alignment strategy is established to further improve the assessment performance of the proposed model. In the class alignment strategy, a class contrastive loss based on the instance is introduced to measure the intra-class and the inter-class discrepancy between the source domain and target domain. The instance-based

contrastive loss can measure intra-class and inter-class discrepancy, which can be expressed as [51, 52]:

$$\ell_d\left(g_i^s, g_i^t, c_{ij}\right) = \begin{cases} \|g_i^s - g_i^t\|^2 & c_{ii} = 1 \\ max(0, m - \|g_i^s - g_i^t\|^2) & c_{ii} = 0 \end{cases} \tag{17}$$

$$\mathscr{L}_{con}(x^s, x^t) = \sum_{i=1}^{n} \ell_d(g_i^s, g_i^t, c_{ii}) \tag{18}$$

where g_i^s and g_i^t indicate the feature distributions of the *i*-th source domain sample and the *i*-th target domain sample, m is a pre-defined distance, $c_{ii} = c(y_i^s, \tilde{y}_i^t)$ can be defined as:

$$c(y_i^s, \tilde{y}_i^t) = \begin{cases} 1, & y_i = \tilde{y}_i^t \\ 0, & y_i \neq \tilde{y}_i^t \end{cases} \tag{19}$$

The proposed model is jointly trained with three losses, and the total loss can be expressed as:

$$\mathscr{L}_{all} = \mathscr{L}_{cls}(x^s, y^s) + \alpha \mathscr{L}_{MMD}(x^s, x^t) + \beta \mathscr{L}_{con}(x^s, x^t) \tag{20}$$

where α and β represent tradeoff parameters.

The $\mathscr{L}_{all}$ is minimized in the model training process to seek the optimal θ_G and θ_C, which can be expressed as:

$$\left(\hat{\theta}_G, \hat{\theta}_C\right) = \underset{\theta_G, \theta_C}{argmin} \mathscr{L}_{all}(x^s, y^s, x^t) \tag{21}$$

The following section will introduce the setting details of the domain adaption model and demonstrate its performance in assessing the surface degradation severities, using the endurance tests under different dominant wear mechanisms.

3 Validation and results

3.1 Experimental design and data collection

Two endurance tests were conducted to verify the effectiveness of the developed digital twin-driven intelligent gear surface degradation assessment methodology: (i) a long-running endurance test with lubrication, in which the fatigue pitting is the dominant wear mechanism;

and (ii) a relatively short endurance test without lubrication, in which the abrasive wear is the dominant wear mechanism. A surface moulding technique was used periodically to duplicate the gear tooth surfaces during the run-to-failure tests [53] (see Figure 6). The acquired mould of gear tooth was then analyzed by optical microscopy and laser scanning confocal microscopy to obtain its surface images, which was used for further relevant analysis, like the surface pitting density calculation. More details about the surface duplication process using mould can refer to [53]. The tribology information obtained from the moulds can be applied for quantitative methodology validation. In addition, to quantify the profile change of the gear tooth during the dry test, an adhesive paper was placed at the bottom of the gearbox casing. The adhesive paper can collect the wear particles produced by the meshing gear surfaces (see Figure 7). The weights of the collected gear wear particles were used to calculate the gear tooth average wear depths. And the wear depth can be applied as an index of the wear severity of the dry test.

The lubricated test was performed under the applied load of 20 Nm and the operating speed of 20 Hz. In comparison, the operating conditions were adjusted to the applied load of 5 Nm and the operating speed of 10 Hz for the dry test. The vibration signals (from B&K 4394 and B&K 4396) were recorded regularly at a sampling rate of 100 kHz for a period of 10 seconds. As shown in Figure 2, B&K 4396 accelerometer is close to the pinion (driving gear), and the pinion has much severer wear than the driven gear [53]. Therefore, the measurements from B&K 4396 accelerometer are utilized in this paper for the corresponding gear wear assessment.

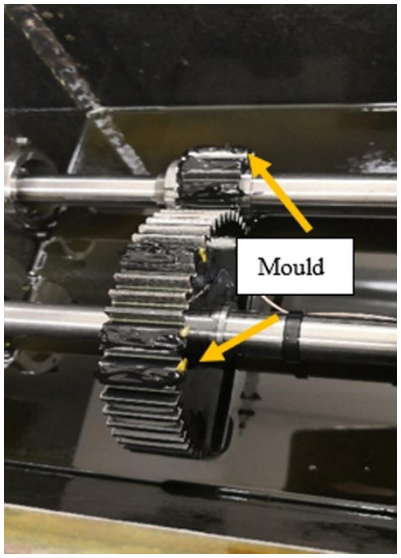

Figure 6 Gear pair and mould making (in lubricated test)

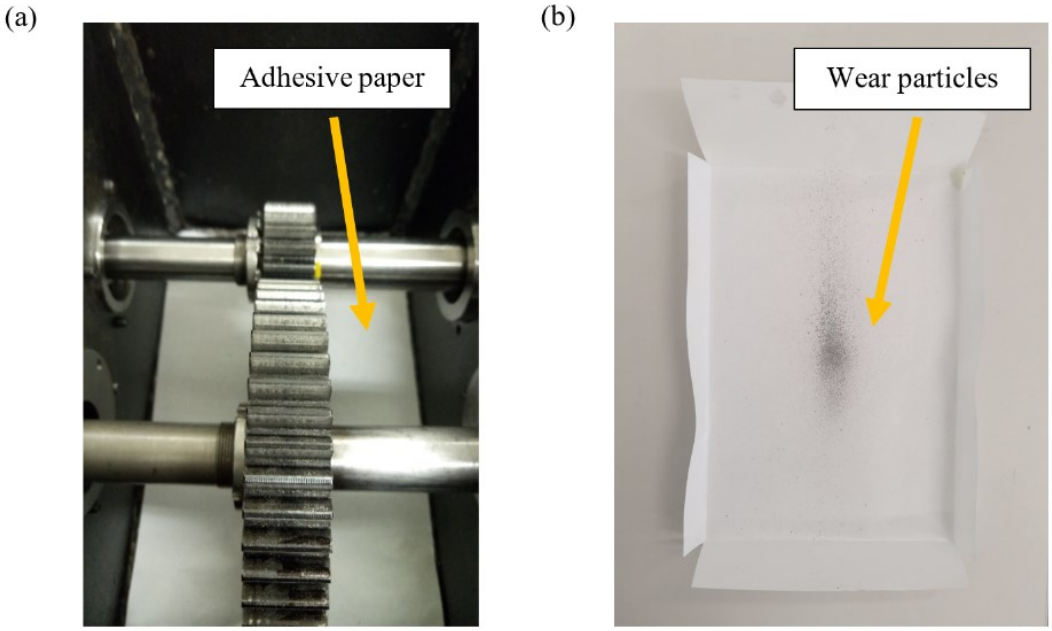

Figure 7 Gear meshing pair and the adhesive paper (used in dry test): (a) an adhesive paper which is placed at the gear casing bottom; (b) the wear particles collected by the adhesive paper

3.2 Results of gear transmission digital twin

The dynamic model is built using Matlab R2021a Simulink, and it takes 2 minutes to generate the corresponding responses on the PC with NVIDIA RTX 3070 GPU. As introduced in Section 2.2.2, the time waveforms, frequency spectrum, and natural frequencies under healthy conditions will be applied to update and calibrate the dynamic model. The time waveforms were recorded by the B&K 4394 and B&K 4396 accelerometers, and the frequency spectra can

be obtained using the Fourier transform. To obtain the natural frequencies of the gear system, a speed ramp test was arranged. This method involves collecting vibration signals over a period when the gearbox is ramping up toward full speed, as shown in Figure 8. This particular test uses the vibration of the shaft as a forcing function to provide energy input into the gear system. This process is used to excite resonances when the shaft vibration passes through the critical speeds.

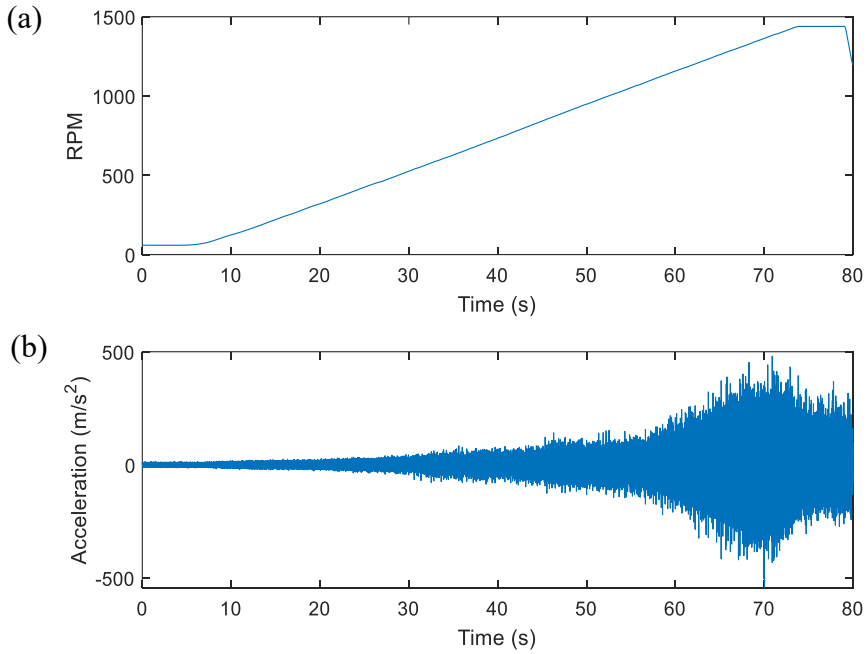

Figure 8 Measured vibration during the spur gearbox ramp test: (a) rotational speed; (b) measured vibrations

Power spectral density (PSD) analysis is a traditional frequency-domain analysis tool to identify the vibration modes/natural frequencies of mechanical systems. The underlying mechanism is that resonance is the amplification of a signal when its frequency is close to the natural frequency of a system. With the help of PSD analysis, the resonances in the spectrum are amplified, enabling the vibration modes of the systems to be easily detected. The PSD analysis of the spur gearbox is shown in Figure 9.

From Figure 9, the modes/natural frequencies of the whole spur gear system can be identified, which are summarized in Table 1. These natural frequencies can be used as references to

calibrate the developed dynamic model. It should be noted that in the developed dynamic model (as shown in Figure 3), the foundation of the spur gearbox is not simulated and included; therefore, to make the natural frequency from ramp tests being comparable with the developed dynamic model of the gear system, the foundation of the gearbox's natural frequencies should be excluded from the natural frequencies acquired from ramp tests. In this regard, an impact test was applied to determine the natural frequencies of the gearbox's foundation so that it can be subtracted from the whole spur gear system, see Figure 10. Note that a slight load is applied to ensure the gear pairs are in contact when conducting the hammer test on the foundation of the gearbox.

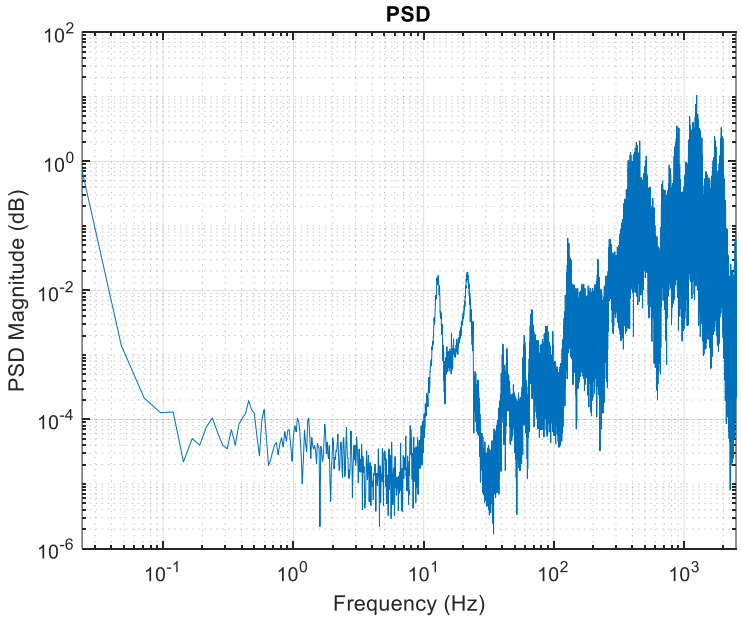

Figure 9 PSD analysis of the vibrations of the spur gear system

Table 1 Identified natural frequencies from the PSD analysis of experimental measurements

Natural frequencies (Hz)						
13	22	41	59	68	87	128
218	270	430	512	6883	767	880
1103	1167	1248	1367	1471	1714	1927

The quadrature picking method is applied here to determine the natural frequencies/modes and mode shapes of the gearbox's foundation [54]. As for the foundation of the gear system, an

accelerometer (B&K 4396) is fixed on the root of the gearbox casing to collect the responses excited by varying inputs generated by impact hammers, as shown in Figure 10. The modes of the foundation of the gearbox and its mode shapes can be calculated and extracted through Eq. (15)

$$H(\omega) = \frac{X(\omega)}{F(\omega)} \tag{15}$$

where $H(\omega)$ is the FRF, $X(\omega)$ is the output of the system, and $F(\omega)$ is the input of the system. More specifically, it is assumed that more noise exists in the harmer excitation (input) signal than in the receiving accelerometer (output) signal, since the harmer is moving while the accelerometer is fixed. The FRF can be calculated by

$$H(\omega) = G_{yy}/G_{yx} \tag{16}$$

where G_{yy} is the auto spectra of output (accelerometer) signal, and G_{yx} is the cross-spectrum between output signal and input (harmer) signal. Five averaging times for the hammer tests were executed. Note that there are some advanced signal processing techniques that can improve the FRF estimation results; for example, singular value decomposition (SVD) was applied in [55] to improve the FRF. However, the FRF estimation is not the main focus of this paper; therefore, the advanced signal process algorithms will not be investigated here to further improve FRF results.

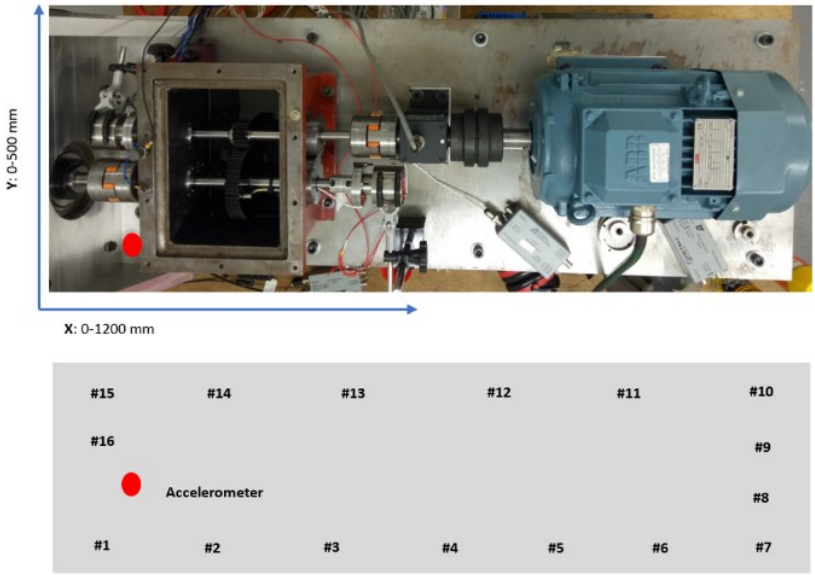

Figure 10 Harmer test on gearbox's foundation

Figure 11 gives the first four mode shapes for demonstration purposes. Based on Eq. (15) and Eq. (16), the FRF of the gearbox's foundation is shown in Figure 12, including the magnitude, phase, and coherence. On the modes of the gear system, there would be a peak in the magnitude spectrum, and the corresponding phase would shift 180 degrees. Coherence is a function versus frequency that indicates the relationship between the input and output of FRF. It can be an indicator of the quality of the FRF, which evaluates the consistency of the FRF from the repetitive measurements: (a) coherence's value is 1 at a particular frequency, indicating that the FRF amplitude and phase are very repeatable from one measurement to another measurement; (b) while coherence's value is 0 indicating the opposite – the measurements are not repeatable, which is a possible "warning flag" that there is an error in the measurement setup.

From the magnitude, phase, and coherence spectra in Figure 12, all the modes of the gearbox's foundation can be identified. By subtracting the obtained modes/natural frequencies from the whole gear system's modes (from the speed ramp test), the modes/natural frequencies of the gear system without foundation can be obtained, which are comparable with the developed dynamic model.

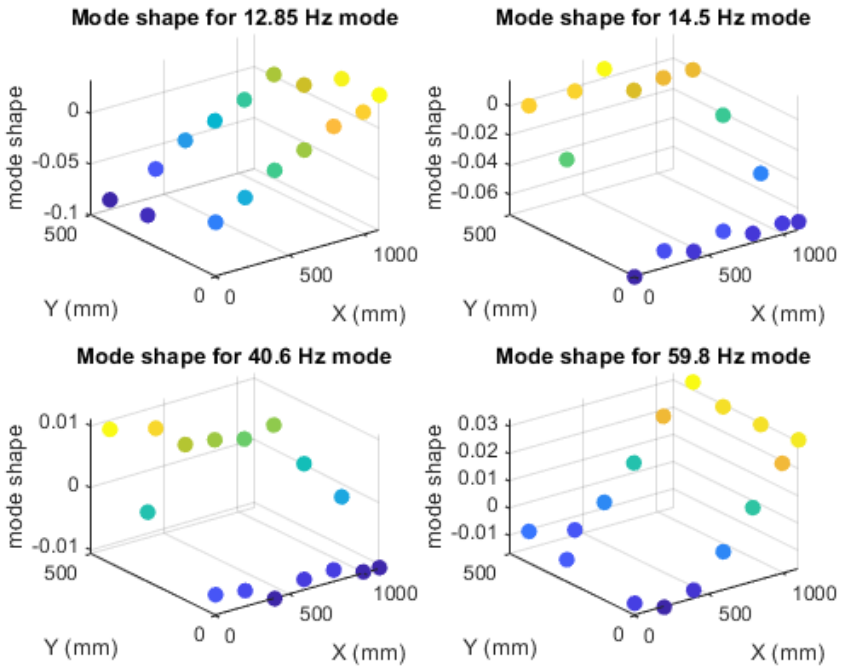

Figure 11 The first four mode shapes of the gearbox's foundation

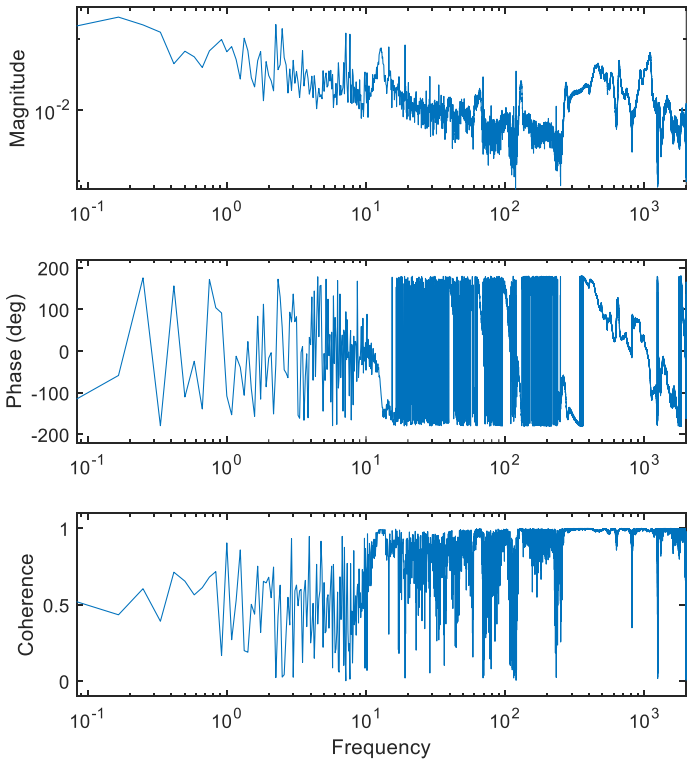

Figure 12 Frequency Response Function (FRF) of the gearbox's foundation

From Eqs. (2)-(9), the optimal parameters from the dynamic model can be determined. The updated or calibrated model can reliably represent the system's dynamic response. For example, the natural frequencies of the dynamic model and gear system are compared in Table 2. Table

2 shows noticeable differences in the natural frequencies between the dynamic model and gear systems. After updating, the first eight natural frequencies are in good agreement between the experiments and the dynamic model within the discrepancy of eight percent, which is fair and reasonable to approximate the system response. In addition, to prove that the calibrated dynamic model can well reflect the gear meshing behavior, the PSD comparison results of transmission error (TE) signals are applied here to validate the effectiveness of the model parameter intelligent updating approach, as shown in Figure 13. By comparing with the traditionally measured vibrations from accelerometers, TE is closer to the source (such as contact force) than the vibration response measurements and less affected by different transmission paths. Thus, the TE signal can more reliably represent the gear meshing characteristics. From Figure 13, it can be seen that the simulation from the calibrated dynamic model matches well with the experimental results. The modes of the gear system are well captured and represented using the dynamic model, indicating that the dynamic model, after intelligent updating, can well reveal the gear system's dynamic characteristics.

Table 2 Comparison of natural frequencies of the dynamic model and experimental data

Experiments (Hz)	Before intelligent updating		After intelligent updating	
	Dynamic model (Hz)	Difference (%)	Dynamic model (Hz)	Difference (%)
22	15	-31.8	23	5.1
87	62	-28.7	81	-6.7
218	156	-28.4	201	-7.8
430	320	-25.6	427	-0.8
512	598	16.8	529	3.5
767	856	11.6	740	3.5
1248	1543	23.6	1268	1.6
1927	2658	37.9	1885	2.2

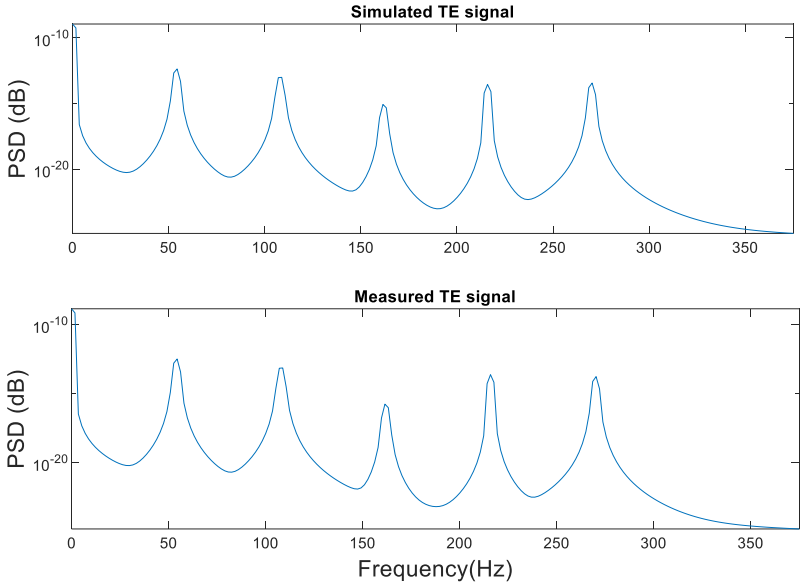

Figure 13 Comparisons of PSD spectra of transmission error signals from dynamic model and physical measurements (based on encoder signals)

3.3 Results of gear wear propagation assessment

3.3.1 Surface pitting severity assessment

To represent and reveal the system degradation status during the gear fatigue pitting progression, the dynamic model and the fatigue pitting model are applied to generate the system response, especially the vibration response. A total of six groups of vibrations are simulated, corresponding to different fatigue pitting severities during the whole lifespan of the gearbox (the maximal pitting density reaches 20%). It should be noted that the endurance tests presented in this paper are accelerated degradation tests, and the data acquisition interval is around 5 minutes, resulting in the amount of measured data is limited. Therefore, balancing the minimum change in each condition and the data length is required for the transfer learning algorithm. The number of health conditions/stages is set as six (around 3% pitting density change for each stage), based on our prior experience [45, 46] and suggestions from the ASM handbook [56]. The real pitting density corresponding to each measured signal from the physical gearbox can be known from the duplicated gear surface, as shown in Figure 6. The real pitting density information can be used here as the kind of "ground truth" to label the actual

degradation stages of the endurance tests. Also, the measured pitting density can be applied to validate the effectiveness of the developed method for gear surface degradation assessment. The simulated vibration signals are applied as the source domain data. The measured vibration signals are the target domain data, whose corresponding wear severity needs to be identified. Then the developed transfer learning algorithm will be implemented, as introduced in Section 2.3. In the proposed domain model, the kernel size of each convolution operation is 9, and the number of channels of the convolution operation is 4, 8, 16, 32, and 64, respectively. The number of units in the fully connected layer is 300. The parameter settings in the experiment are shown in Table 3. Each experiment is performed six times, and the average value is used as the final surface degradation assessment result. The developed domain model is implemented on the PC with NVIDIA RTX 3070 GPU, and it takes 0.2s to obtain the assessment result once the model is pre-trained.

Table 3 Parameter settings of the domain adaptation model

Parameter	Value	Parameter	Value
Source training samples	400	Batch size	20
Target training samples	300	Epochs	80
Target test samples	100	α	0.1
Initial learning rate	0.01	β	0.1

To more intuitively reflect the performance of the model, the confusion matrix visualization technique is used to display the surface wear assessment accuracy of all conditions. In the confusion matrix, the numbers on the diagonal represent the correct rate of prediction for each condition, and the data in other positions represent the error rates for the corresponding incorrect condition. The confusion matrix of the developed method is shown in Figure 14 to show the surface wear assessment accuracy. From Figure 14, it can be seen that the average surface wear assessment accuracy is approximately 94.80%, which proves its excellent performance in surface wear evaluations. Also, the t-SNE technology is employed to visualize the high-level features of the target test data, as shown in Figure 15. In Figure 15, each wear

state in the proposed method achieves good individual clustering, indicating the effectiveness of the developed surface wear assessment method.

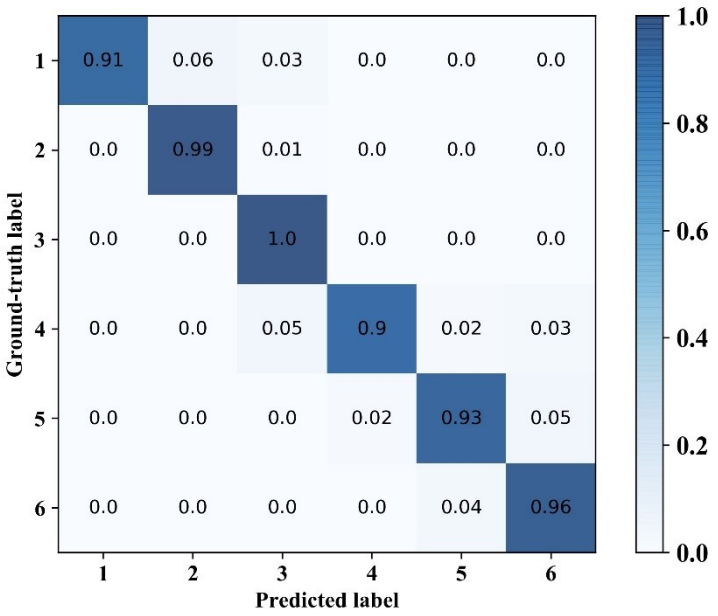

Figure 14 Confusion matrix of the developed method: lubricated test

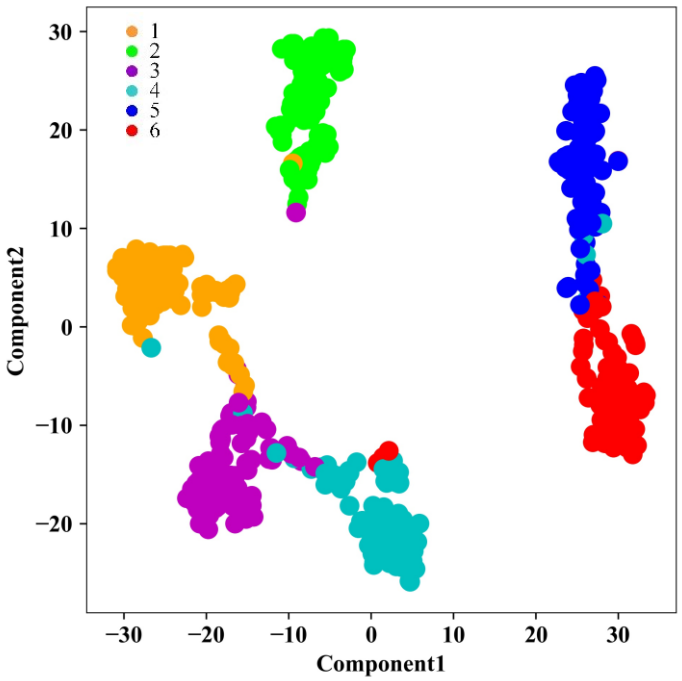

Figure 15 Visualization of the developed method via t-SNE: lubricated test

To demonstrate the superiority of the proposed domain-adaptive model for assessing the surface degradation severity over other conventional techniques, two comparison methods are designed under the same experimental scenario. The first comparison method is the

convolutional neural network (CNN), in which the model trained on the simulated data is directly used for testing the measured data. Another comparison method is MMD, in which only MMD loss is introduced for domain invariant learning.

To more comprehensively evaluate the performance of the methods, each method is executed six times, and their average evaluation accuracies are 80.45% and 90.26%, respectively. The confusion matrices of these two methods are shown in Figure 16. From the comparisons between Figure 14 and Figure 16, it can be seen that the proposed digital twin-driven surface degradation assessment method can obtain much more accurate degradation assessment results under all degradation conditions, and the assessment performance is superior to those of the none-domain adaptation method and the traditional domain adaptation method.

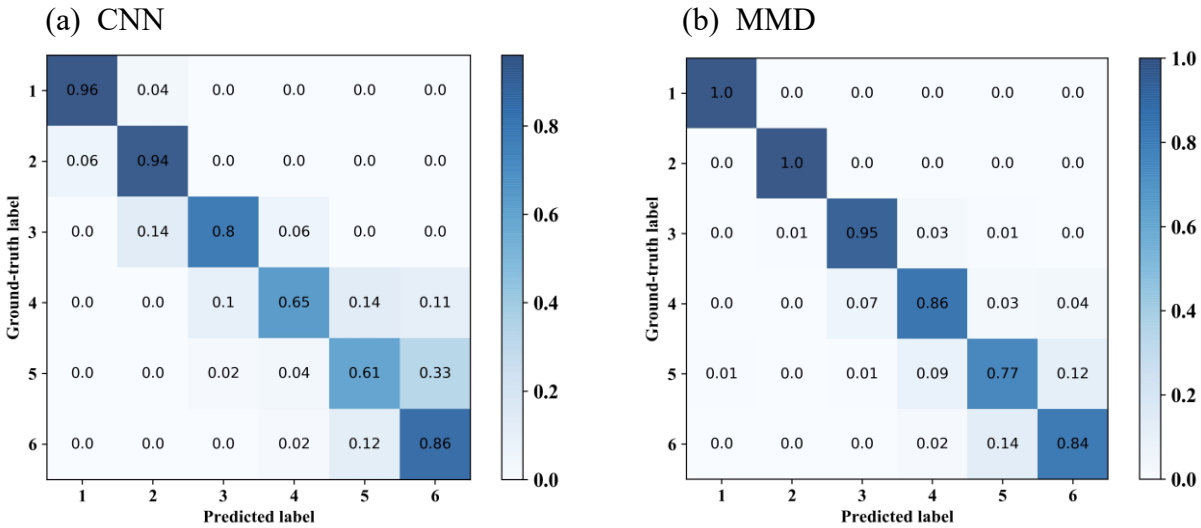

Figure 16 Lubricated test: the confusion matrices of the two methods: CNN and MMD

To further demonstrate the superiority of the proposed method, the t-SNE technology is employed to visualize the high-level features of the target test data in CNN and MMD. The visualization results of these two methods are shown in Figure 17. From Figure 17, it can be seen that for the CNN method without domain adaptation, there are more feature overlaps in most classes. The feature overlap indicates a noticeable discrepancy in feature distribution between the simulation data and the test data. Consequently, the model obtained from the simulation data being directly used for evaluating characteristics of measured data does not

work well. Compared with CNN, the introduction of the MMD method improves the evaluation accuracy. However, there are still more feature overlaps among conditions 4, 5, and 6, which affect the performance of the model. In contrast, as shown in Figure 15, each class in the proposed method achieves better individual clustering. The comparison of Figure 15 and Figure 17 illustrates that the proposed method can efficiently achieve both domain alignment and class alignment for accurate degradation condition assessment.

(a) CNN (b) MMD

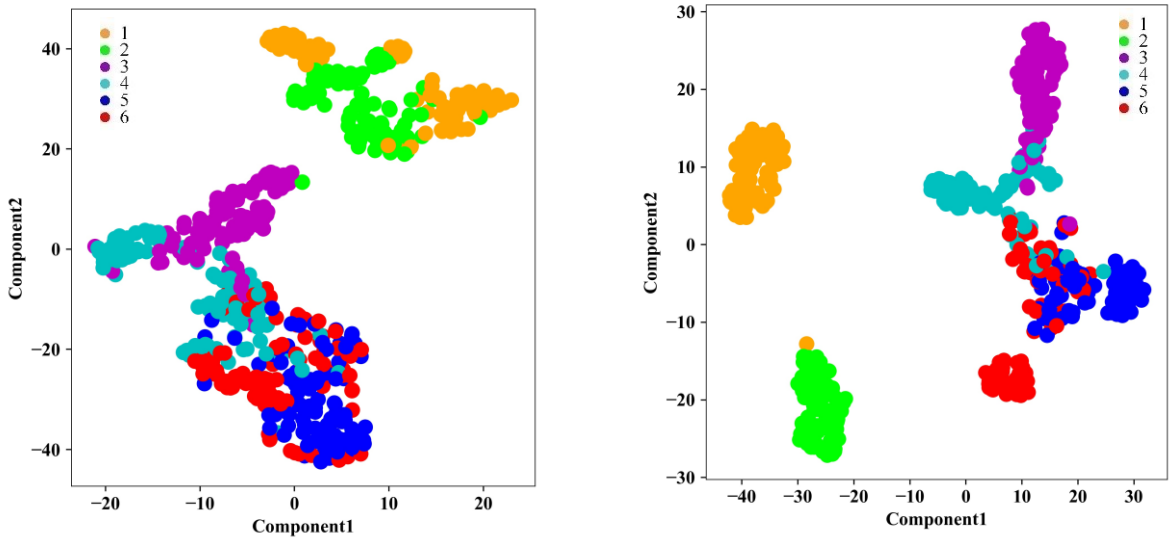

Figure 17 Lubricated test: visualization of the methods-CNN and MMD

3.3.2 Tooth profile change severity assessment

In the short dry test, the dominant gear wear mechanism is the abrasive gear wear, and thus the gear meshing tooth profile has changed significantly. Similar to the lubricated test, by balancing the minimum change contained in each stage and the data length required for the transfer learning algorithm, four degradation stages with different tooth profile changes are defined for the endurance test (around 60 μm tooth profile change for each stage). Note that the profile change of the gear tooth can be evaluated with the help of the collected wear particles (Figure 7). This information can serve as the kind of "ground truth" to verify the effectiveness of the developed method.

Similar procedures are implemented for the short dry tests, as introduced in Section 3.3.1. As for the developed method, the average gear surface degradation assessment accuracy is 92.71%, and the corresponding confusion matrix is shown in Figure 18. The visualization result of the developed method is shown in Figure 19.

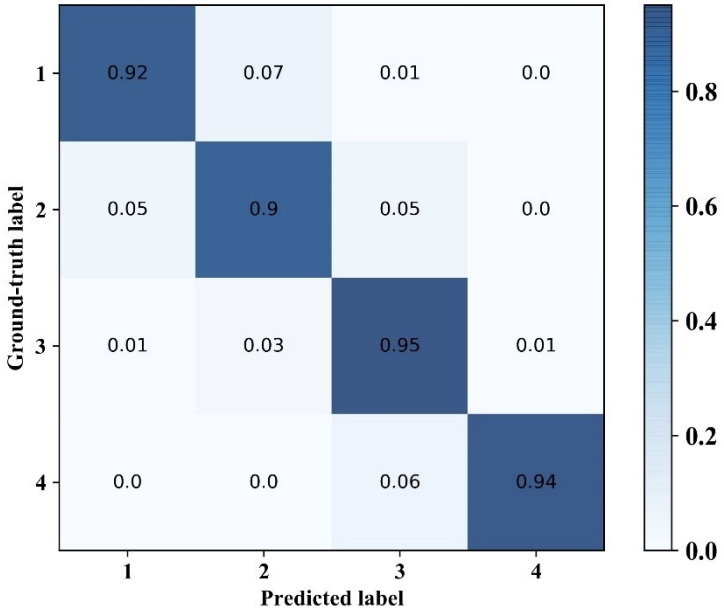

Figure 18 Confusion matrix of the developed method: dry test

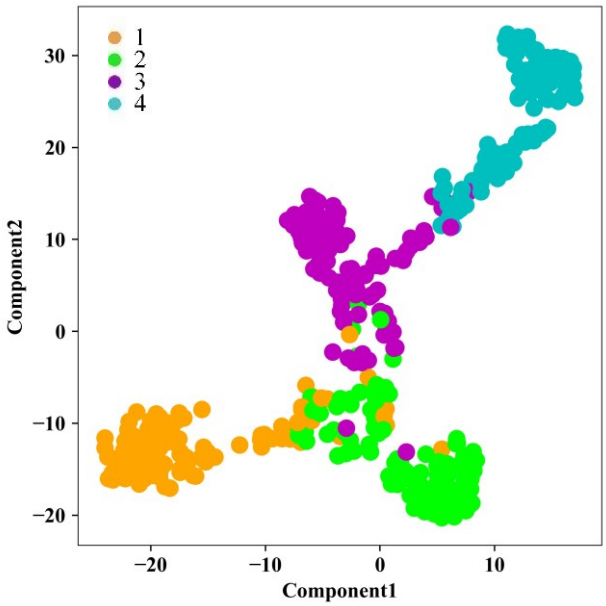

Figure 19 Visualization of the developed method via t-SNE: lubricated test

Similar to Section 3.3.1, in order to illustrate the superiority of the proposed method in the gear wear assessment, CNN and MMD are applied here to implement the relevant comparisons. The

evaluation accuracies of these two methods are 80.13% and 85.41%, respectively. The corresponding confusion matrices are shown in Figure 20, and the visualization results are shown in Figure 21. From the comparisons, it can be seen that the proposed method achieves the most accurate recognition under each condition. This proves that the proposed method has excellent generalization performance. Also, from the visualization results shown in Figure 19 and Figure 21, it can be observed that for stages 1, 2, and 3 with severe feature overlaps, the proposed method can significantly improve the feature distinguishability among them.

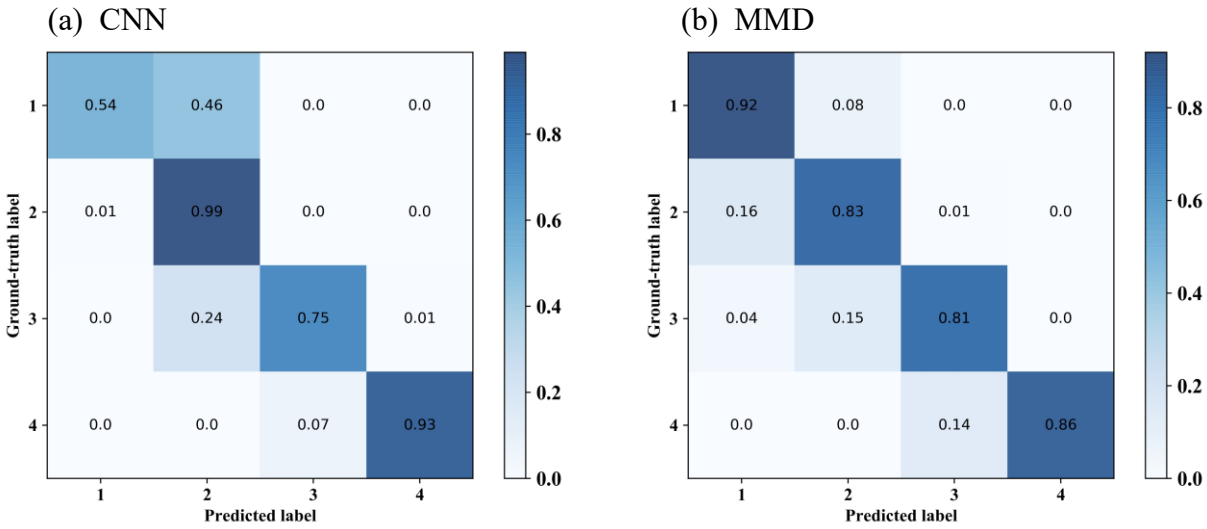

Figure 20 Dry test: the confusion matrices of the two methods-CNN and MMD

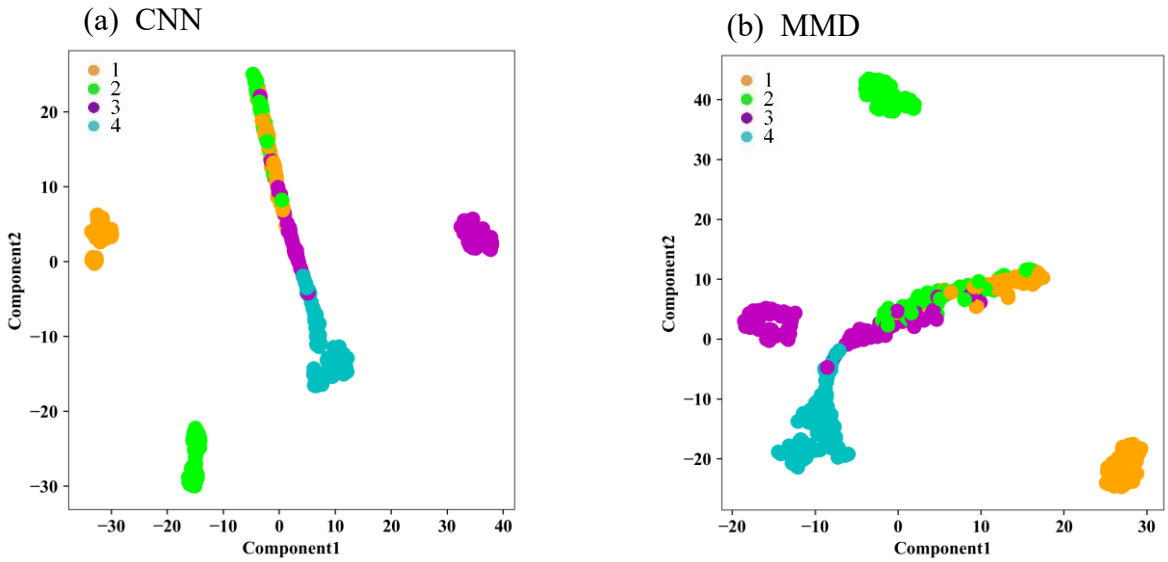

Figure 21 Lubricated test: visualization of the methods-CNN and MMD

4 Conclusion and future work

Gear surface wear is a common but inevitable phenomenon in the lifespan of the gearbox, and its propagation can induce severe failures, which could result in unexpected economic loss and even severe incidents. Real-time assessment of the system degradation status could provide many benefits to industrial applications. In this paper, a digital twin-driven methodology has been proposed for intelligent gear surface degradation assessment. With the help of the measurements from the physical system, the high-fidelity digital twin model was established automatically to reveal the dynamic response of the gear system during the gear wear progression. The developed transfer learning algorithm was adopted to transfer the knowledge learned from the digital twin models to the surface wear assessment of the physical gearbox. The developed methodology could assess the surface wear severity accurately with a non-destructive approach. Two kinds of endurance tests were performed to verify the effectiveness of the developed methodology and prove the practical values and significance of the developed methodology in industrial practices. Moreover, the developed methodology could improve the applicability of the digital twin techniques in gear wear monitoring, promoting the wide application of digital twin techniques in gear health management.

In future work, other failure wear mechanisms will be considered. The digital twin model will also be improved so that it can be timely updated during the progression process of gear wear, and the actual system degradation behavior can be revealed in real-time and with high accuracy. The FRF estimation is crucial for gearbox modeling; thus, advanced signal processing methods will be investigated to improve the FRF results. Also, the current digital twin model is complex, bringing challenges in model establishment and calibration. Thus, the model order reduction techniques will be studied to reduce the complexity of gear wear assessment using the digital twin solution and improve its practicability. Furthermore, some advanced deep learning

algorithms, such as long short-term memory network (LSTM) and Transformer, will be investigated to further improve the diagnosis accuracy of the gearbox transmission system.

Acknowledgment

This work was supported by the National Research Council of Canada through the Canada - Germany 3+2 Joint Project "Digital Twin Platform for Infrastructure Asset Lifecycle Management" (Agreement No. INT-016-1). The authors are grateful for the support. In addition, the authors would like to thank the assistance provided by the Tribology and Machine Condition Monitoring Lab from the University of New South Wales.

Appendix A

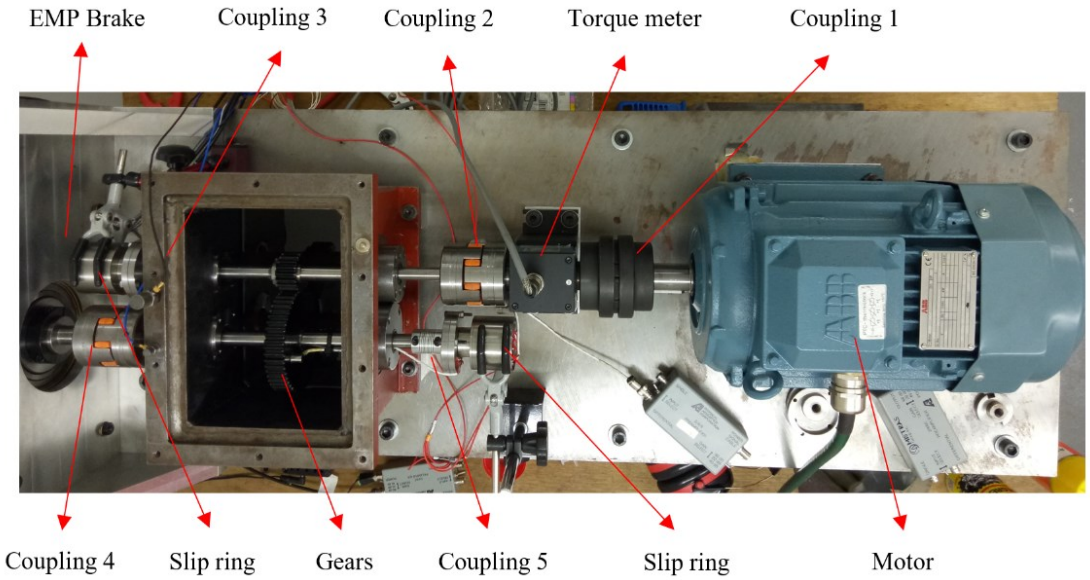

Figure A1. Spur gearbox test rig with each labeled modeling component [46]

Table A1. Mass and inertia of the dynamic model of spur gearbox system [46]

Inertia (kgm^2)		Mass (kg)	
I_1	Inertia of motor, inertia of motor shaft and $1/2$ inertia of coupling 1	m_4	Mass of pedestal and bearing
I_2	$1/2$ inertia of coupling 1, $1/2$ inertia of coupling 2 and inertia of torque meter shaft	m_5	Mass of pinion
I_3	$1/2$ inertia of coupling 2 and inertia of input shaft to bearing (not including bearing)	m_6	Mass of pedestal and bearing

I_5	Inertia of pinion, inertia of input shaft (whole shaft section through casing)	m_9	Mass of pedestal and bearing
I_7	Inertia of slip ring (rotor part), adapter, coupling 3 and shaft section outside casing (free end)	m_{10}	Mass of gear
I_8	Inertia of slip ring (rotor part), adapter, coupling 5 and shaft section outside casing (free end)	m_{11}	Mass of pedestal and bearing
I_{10}	Inertia of gear, inertia of whole output shaft section through casing		
I_{12}	$1/2$ inertia of coupling 4 and output shaft section to bearing (not including bearing)		
I_{13}	Inertia of brake, inertia of brake shaft and $1/2$ inertia of coupling 4		

Table A2. Stiffness and damping of the dynamic model of spur gearbox system [46]

Stiffness and damping			
k_{t12}, C_{t12}	Torsional stiffness and damping of coupling 1	k_{y56}, C_{y56}	Vertical stiffness and damping of shaft
k_{t23}, C_{t23}	Torsional stiffness and damping of coupling 2	k_{y101}, C_{y101}	Vertical stiffness and damping of shaft
k_{t35}, C_{t35}	Torsional stiffness and damping of shaft (from coupling 2 to pinion)	k_{y910}, C_{y910}	Vertical stiffness and damping of shaft
k_{t57}, C_{t57}	Torsional stiffness and damping of shaft (from pinion to slip ring)	k_{x45}, C_{x45}	Horizontal stiffness and damping of shaft
k_{t810}, C_{t810}	Torsional stiffness and damping of shaft (from gear to slip ring)	k_{x56}, C_{x56}	Horizontal stiffness and damping of shaft
k_{t102}, C_{t102}	Torsional stiffness and damping of shaft (from gear to coupling 4)	k_{x101}, C_{x101}	Horizontal stiffness and damping of shaft
k_{t123}, C_{t123}	Torsional stiffness and damping of coupling 4	k_{x910}, C_{x45}	Horizontal stiffness and damping of shaft
k_{y45}, C_{y45}	Vertical stiffness and damping of shaft	k_{y4}, k_{y6}, k_{y9}, k_{y11}	Vertical stiffness of pedestal/bearing

References

[1] X. Liang, M.J. Zuo, Z. Feng, Dynamic modeling of gearbox faults: A review, Mechanical Systems and Signal Processing, 98 (2018) 852-876.
[2] K. Feng, J.C. Ji, Q. Ni, M. Beer, A review of vibration-based gear wear monitoring and prediction techniques, Mechanical Systems and Signal Processing, 182 (2023) 109605.
[3] Y. Chen, M. Rao, K. Feng, M.J. Zuo, Physics-Informed LSTM hyperparameters selection for gearbox fault detection, Mechanical Systems and Signal Processing, 171 (2022) 108907.
[4] S. Schmidt, R. Zimroz, P.S. Heyns, Enhancing gearbox vibration signals under time-varying operating conditions by combining a whitening procedure and a synchronous processing method, Mechanical Systems and Signal Processing, 156 (2021) 107668.
[5] Y. Li, K. Feng, X. Liang, M.J. Zuo, A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved Vold-Kalman filter and multi-scale sample entropy, Journal of Sound and Vibration, 439 (2019) 271-286.
[6] K. Feng, J.C. Ji, Q. Ni, A novel adaptive bandwidth selection method for Vold–Kalman filtering and its application in wind turbine planetary gearbox diagnostics, Structural Health Monitoring, (2022) 14759217221099966.
[7] Y. Huangfu, X. Dong, K. Chen, G. Tu, X. Long, Z. Peng, A tribo-dynamic based pitting evolution model of planetary gear sets: A topographical updating approach, International Journal of Mechanical Sciences, 220 (2022) 107157.
[8] K. Feng, J.C. Ji, Y. Li, Q. Ni, H. Wu, J. Zheng, A novel cyclic-correntropy based indicator for gear wear monitoring, Tribology International, 171 (2022) 107528.
[9] S. Sendlbeck, A. Fimpel, B. Siewerin, M. Otto, K. Stahl, Condition monitoring of slow-speed gear wear using a transmission error-based approach with automated feature selection, International Journal of Prognostics and Health Management, 12 (2021).
[10] M. Amarnath, S. Chandramohan, S. Seetharaman, Experimental investigations of surface wear assessment of spur gear teeth, Journal of Vibration and Control, 18 (2011) 1009-1024.
[11] K. Feng, Q. Ni, M. Beer, H. Du, C. Li, A novel similarity-based status characterization methodology for gear surface wear propagation monitoring, Tribology International, 174 (2022) 107765.
[12] F. Karpat, S. Ekwaro-Osire, Influence of tip relief modification on the wear of spur gears with asymmetric teeth, Tribology Transactions, 51 (2008) 581-588.
[13] F. Karpat, A.E. Dirik, O.C. Kalay, C. Yüce, O. Doğan, B. Korcuklu, Fault diagnosis with deep learning for standard and asymmetric involute spur gears, ASME 2021 International Mechanical Engineering Congress and Exposition, 2021.
[14] S. Raadnui, Condition monitoring of worm gear wear and wear particle analysis of industrial worm gear sets, Wear, 476 (2021) 203687.
[15] G.P. Stachowiak, G.W. Stachowiak, P. Podsiadlo, Automated classification of wear particles based on their surface texture and shape features, Tribology International, 41 (2008) 34-43.
[16] P. Bajpai, A. Kahraman, N.E. Anderson, A surface wear prediction methodology for parallel-axis gear pairs, Journal of Tribology, 126 (2004) 597-605.
[17] J. Pisula, G. Budzik, P. Turek, M. Cieplak, An Analysis of Polymer Gear Wear in a Spur Gear Train Made Using FDM and FFF Methods Based on Tooth Surface Topography Assessment, Polymers, 13 (2021) 1649.
[18] J. Lin, C. Teng, E. Bergstedt, H. Li, Z. Shi, U. Olofsson, A quantitatively distributed wear-measurement method for spur gears during micro-pitting and pitting tests, Tribology International, 157 (2021) 106839.

[19] L.F. Villa, A. Reñones, J.R. Perán, L.J. de Miguel, Statistical fault diagnosis based on vibration analysis for gear test-bench under non-stationary conditions of speed and load, Mechanical Systems and Signal Processing, 29 (2012) 436-446.
[20] Q. Ni, J.C. Ji, K. Feng, B. Halkon, A fault information-guided variational mode decomposition (FIVMD) method for rolling element bearings diagnosis, Mechanical Systems and Signal Processing, 164 (2022) 108216.
[21] D. Wang, K.-L. Tsui, Two novel mixed effects models for prognostics of rolling element bearings, Mechanical Systems and Signal Processing, 99 (2018) 1-13.
[22] Q. Ni, J. Ji, K. Feng, Data-driven prognostic scheme for bearings based on a novel health indicator and gated recurrent unit network, IEEE Transactions on Industrial Informatics, (2022) 1-1.
[23] K. Feng, W.A. Smith, P. Borghesani, R.B. Randall, Z. Peng, Use of cyclostationary properties of vibration signals to identify gear wear mechanisms and track wear evolution, Mechanical Systems and Signal Processing, 150 (2021) 107258.
[24] R.B. Randall, A new method of modeling gear faults, Journal of Mechanical Design, 104 (1982) 259-267.
[25] S. Ziaran, R. Darula, Determination of the state of wear of high contact ratio gear sets by means of spectrum and cepstrum analysis, Journal of Vibration and Acoustics, 135 (2013).
[26] G.E. Morales-Espejel, P. Rycerz, A. Kadiric, Prediction of micropitting damage in gear teeth contacts considering the concurrent effects of surface fatigue and mild wear, Wear, 398-399 (2018) 99-115.
[27] H. Chang, P. Borghesani, Z. Peng, Investigation on the relationship between macropits and wear particles in a gear fatigue process, Wear, 484-485 (2021) 203724.
[28] M. Weibring, L. Gondecki, P. Tenberge, Simulation of fatigue failure on tooth flanks in consideration of pitting initiation and growth, Tribology International, 131 (2019) 299-307.
[29] C. Hu, W.A. Smith, R.B. Randall, Z. Peng, Development of a gear vibration indicator and its application in gear wear monitoring, Mechanical Systems and Signal Processing, 76-77 (2016) 319-336.
[30] C.K. Tan, P. Irving, D. Mba, A comparative experimental study on the diagnostic and prognostic capabilities of acoustics emission, vibration and spectrometric oil analysis for spur gears, Mechanical Systems and Signal Processing, 21 (2007) 208-233.
[31] P.J. Dempsey, Integrating oil debris and vibration measurements for intelligent machine health monitoring, Doctor of Philosophy Thesis, The University of Toledo, 2002.
[32] P. Kundu, A.K. Darpe, M.S. Kulkarni, A correlation coefficient based vibration indicator for detecting natural pitting progression in spur gears, Mechanical Systems and Signal Processing, 129 (2019) 741-763.
[33] R. Zhang, X. Gu, F. Gu, T. Wang, A.D. Ball, Gear wear process monitoring using a sideband estimator based on modulation signal bispectrum, Applied Sciences, 7 (2017) 274.
[34] R. Zhang, F. Gu, H. Mansaf, T. Wang, A.D. Ball, Gear wear monitoring by modulation signal bispectrum based on motor current signal analysis, Mechanical Systems and Signal Processing, 94 (2017) 202-213.
[35] T.G. Ritto, F.A. Rochinha, Digital twin, physics-based model, and machine learning applied to damage detection in structures, Mechanical Systems and Signal Processing, 155 (2021) 107614.
[36] J. Yang, R.S. Langley, L. Andrade, Digital twins for design in the presence of uncertainties, Mechanical Systems and Signal Processing, 179 (2022) 109338.
[37] T.G. Ritto, K. Worden, D.J. Wagg, F.A. Rochinha, P. Gardner, A transfer learning-based digital twin for detecting localised torsional friction in deviated wells, Mechanical Systems and Signal Processing, 173 (2022) 109000.

[38] W. Kim, S. Kim, J. Jeong, H. Kim, H. Lee, B.D. Youn, Digital twin approach for on-load tap changers using data-driven dynamic model updating and optimization-based operating condition estimation, Mechanical Systems and Signal Processing, 181 (2022) 109471.
[39] H. Jiang, S. Qin, J. Fu, J. Zhang, G. Ding, How to model and implement connections between physical and virtual models for digital twin application, Journal of Manufacturing Systems, 58 (2021) 36-51.
[40] Q. Qi, F. Tao, T. Hu, N. Anwer, A. Liu, Y. Wei, L. Wang, A.Y.C. Nee, Enabling technologies and tools for digital twin, Journal of Manufacturing Systems, 58 (2021) 3-21.
[41] J. Ma, H. Chen, Y. Zhang, H. Guo, Y. Ren, R. Mo, L. Liu, A digital twin-driven production management system for production workshop, The International Journal of Advanced Manufacturing Technology, 110 (2020) 1385-1397.
[42] K. Feng, P. Borghesani, W.A. Smith, R.B. Randall, Z.Y. Chin, J. Ren, Z. Peng, Vibration-based updating of wear prediction for spur gears, Wear, 426-427 (2019) 1410-1415.
[43] S. Saremi, S. Mirjalili, A. Lewis, Grasshopper Optimisation Algorithm: Theory and application, Advances in Engineering Software, 105 (2017) 30-47.
[44] S.Z. Mirjalili, S. Mirjalili, S. Saremi, H. Faris, I. Aljarah, Grasshopper optimization algorithm for multi-objective optimization problems, Applied Intelligence, 48 (2018) 805-820.
[45] K. Feng, W.A. Smith, Z. Peng, Use of an improved vibration-based updating methodology for gear wear prediction, Engineering Failure Analysis, 120 (2021) 105066.
[46] K. Feng, W.A. Smith, R.B. Randall, H. Wu, Z. Peng, Vibration-based monitoring and prediction of surface profile change and pitting density in a spur gear wear process, Mechanical Systems and Signal Processing, 165 (2022) 108319.
[47] H. Cao, H. Shao, X. Zhong, Q. Deng, X. Yang, J. Xuan, Unsupervised domain-share CNN for machine fault transfer diagnosis from steady speeds to time-varying speeds, Journal of Manufacturing Systems, 62 (2022) 186-198.
[48] Y. Zhang, Z. Ren, S. Zhou, K. Feng, K. Yu, Z. Liu, Supervised Contrastive Learning-Based Domain Adaptation Network for Intelligent Unsupervised Fault Diagnosis of Rolling Bearing, IEEE/ASME Transactions on Mechatronics, (2022) 1-10.
[49] K.M. Borgwardt, A. Gretton, M.J. Rasch, H.-P. Kriegel, B. Schölkopf, A.J. Smola, Integrating structured biological data by Kernel Maximum Mean Discrepancy, Bioinformatics, 22 (2006) e49-e57.
[50] M. Long, J. Wang, G. Ding, J. Sun, P.S. Yu, Transfer feature learning with joint distribution adaptation, Proceedings of the IEEE international conference on computer vision, 2013, pp. 2200-2207.
[51] C. Chen, Z. Chen, B. Jiang, X. Jin, Joint domain alignment and discriminative feature learning for unsupervised deep domain adaptation, Proceedings of the AAAI conference on artificial intelligence, 2019, pp. 3296-3303.
[52] Y. Zhang, K. Yu, Z. Ren, S. Zhou, Joint Domain Alignment and Class Alignment Method for Cross-Domain Fault Diagnosis of Rotating Machinery, IEEE Transactions on Instrumentation and Measurement, 70 (2021) 1-12.
[53] H. Chang, P. Borghesani, W.A. Smith, Z. Peng, Application of surface replication combined with image analysis to investigate wear evolution on gear teeth – A case study, Wear, 430-431 (2019) 355-368.
[54] Brüel and Kjær. Modal Analysis and Simulation: Structural Testing - Part II. Accessed on https://www.bksv.com/en/Knowledge-center/blog/articles/vibration/structural-testing-part-two. Retreved June 14, 2021.
[55] K.Y. Sanliturk, O. Cakar, Noise elimination from measured frequency response functions, Mechanical Systems and Signal Processing, 19 (2005) 615-631.
[56] ASM Handbook, Volume 19: Fatigue and Fracture, ASM International, 1996.