

Article

A Study of a Domain-Adaptive LSTM-DNN-Based Method for Remaining Useful Life Prediction of Planetary Gearbox

Zixuan Liu, Chaobin Tan *, Yuxin Liu, Hao Li, Beining Cui and Xuanzhe Zhang

North China Institute of Aerospace Engineering, Langfang 065000, China; zxbz999@163.com (Z.L.); lyx2267013175@163.com (Y.L.); lihao5563@163.com (H.L.); beiningcui@163.com (B.C.); zhe202416@163.com (X.Z.)
* Correspondence: tzb@nciae.edu.cn

Abstract: Remaining Useful Life (RUL) prediction is an important component of failure prediction and health management (PHM). Current life prediction studies require large amounts of tagged training data assuming that the training data and the test data follow a similar distribution. However, the RUL-prediction data of the planetary gearbox, which works in different conditions, will lead to statistical differences in the data distribution. In addition, the RUL-prediction accuracy will be affected seriously. In this paper, a planetary transmission test system was built, and the domain adaptive model was used to Implement the transfer learning (TL) between the planetary transmission system in different working conditions. LSTM-DNN network was used in the data feature extraction and regression analysis. Finally, a domain-adaptive LSTM-DNN-based method for remaining useful life prediction of Planetary Transmission was proposed. The experimental results show that not only the impact of different operating conditions on statistical data was reduced effectively, but also the efficiency and accuracy of RUL prediction improved.

Keywords: remaining useful life prediction; LSTM-DNN network; domain adaptation; planetary transmission



Citation: Liu, Z.; Tan, C.; Liu, Y.; Li, H.; Cui, B.; Zhang, X. A Study of a Domain-Adaptive LSTM-DNN-Based Method for Remaining Useful Life Prediction of Planetary Gearbox. *Processes* **2023**, *11*, 2002. <https://doi.org/10.3390/pr11072002>

Academic Editors: Sunghan Kim, Hongyun So and Jong-Won Park

Received: 25 April 2023

Revised: 2 June 2023

Accepted: 20 June 2023

Published: 3 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The planetary gearbox is widely used in various fields as an important core component of rotating machinery [1]. It often works in complex and bad environmental conditions. Many operation accidents are caused by malfunctions or failures of the planetary gearbox. The accurate remaining useful life (RUL) prediction of planetary transmissions has great significance for preventing major industrial accidents and economic losses [2].

The advent of failure prediction and health management (PHM) offers a solution for RUL prediction in a gearbox. PHM generally consists of five parts: data acquisition and processing, feature extraction, fault diagnosis, health state assessment, and life prediction. Health status assessment and life prediction are the main directions of PHM research. Health status assessment is conducted using constructing Health Index (HI)-curves for machinery and equipment, which usually have a monotonic upward or downward trend. A threshold value was set to assess the health status of the equipment. The historical data and the online condition data were used to predict the remaining life and health status of mechanical equipment in the future.

The current research on RUL prediction methods for mechanical equipment is based on physical models (Model-based) and data-driven models (Data-driven). The current RUL prediction methods mostly use data-driven model methods because the physical model requires a large amount of expert knowledge and experience. Chen et al. [3] proposed a multiscale long-term cyclic convolutional network deep learning framework (MSWR LRCN) with residual shrinkage construction units to address the low accuracy of traditional RUL prediction. This method can effectively improve the accuracy of RUL prediction. Fu et al. [4] proposed a combination of a convolutional neural network (CNN) and a short-term memory network (LSTM) to predict the RUL of a gearbox. He et al. [5] proposed

an improved temporal convolutional network (TCN) method for predicting the RUL of gearboxes. This method can learn more complete historical information and has higher prediction accuracy than other methods.

However, the above studies are all based on life prediction under the same operating conditions, with the pre-requisite assumption that the data in the training and test sets are required to belong to the same data distribution. In fact, the operating environment of mechanical equipment such as gearboxes is complex, the operating conditions are diverse, and the data distribution may vary greatly under different operating conditions, which will lead to low prediction accuracy using traditional machine learning algorithms.

The emergence of transfer learning (TL) methods avoids these problems. TL is a machine learning method that uses existing models and knowledge to solve unknown but related fields. Transfer learning can be divided into model-based, parameter-based, and feature-based methods. Among them, feature-based transfer learning has been a hot topic of research [6]. The traditional feature-based transfer learning, such as: In 2009, Professor Yang Qiang's [7] team proposed Transfer Component Analysis (TCA) for edge distribution adaptation, which mainly uses Maximum Mean Discrepancy (MMD) to calculate the difference between the mean values of the source and target domains after mapping. The main purpose of TCA is to calculate the difference between the mean values of the source and target domains after mapping using Maximum Mean Discrepancy (MMD). Subsequently, many scholars have extended the research on the basis of TCA [8–10]. In 2013, Ming-Sheng Long [11] et al. proposed the classical Joint Distribution Adaptation (JDA) method, which uses the MMD distance to calculate the difference between the edge distribution and the conditional distribution. In 2018, JinDong Wang [12] et al. argued that the edge distribution and the conditional distribution are not equally important and proposed Dynamic Distribution Adaptation (BDA). In recent years, deep learning technology has developed rapidly, and many research scholars have proposed deep transfer learning on top of traditional transfer learning methods [13–15]. Compared with traditional non-deep transfer learning methods, deep transfer learning has more adaptive feature extraction capability, as well as meeting the end-to-end training requirements in practical applications. The advantages of deep transfer learning have received attention from scholars both at home and abroad and have been successfully used in the fields of object recognition [16] and PHM research [17].

Deep transfer learning has achieved good application in fault diagnosis research [18]. However, there is relatively little research on RUL prediction. Furthermore, most prediction studies focus on bearings, with few of the literature related to the RUL prediction of gearboxes. For example, Wang et al. [19] predicted the RUL of bearings by combining transfer learning with the bidirectional long short-term memory (BiLSTM) model with an attention mechanism. Zhu et al. [20] proposed a bearing RUL prediction method combining transfer learning and LSTM. Compared with other methods, the LSTM method achieved better prediction results. LSTM and its variants have received attention from researchers in the prediction of bearings and have achieved good prediction accuracy. However, the predicted objects are mostly bearings and rarely involve objects such as planetary gearboxes. However, planetary gearboxes are also key components of rotating machinery, so accurately predicting the RUL of planetary gearboxes is of great significance for preventing major industrial accidents and economic losses.

In view of this, this article proposes a long, short-term memory network deep neural networks (LSTM-DNN) life prediction model based on domain adaptation. The proposed model can predict the lifespan between different operating conditions and has good generalization ability.

In this paper, labeled source domain data and unlabeled target domain data are used for training, which can be seen as a feature migration-based learning process, while various types of losses are considered to learn domain-invariant features. In order to verify the method proposed in this paper, a planetary gearbox test rig was built, and the experimental

results show the effectiveness of the method proposed in this paper. The main contributions of this paper are as follows:

1. A domain-based adaptive LSTM-DNN prediction algorithm is proposed, which utilizes the respective strengths of LSTM and DNN and combines transfer learning algorithms to reduce data distribution differences.
2. Experiments on planetary gearboxes were built, and the proposed network model was compared with other network models to show that the prediction accuracy is better by considering the regression loss and metric loss based on the dataset and the effectiveness and superiority of the proposed domain adaptive LSTM-DNN RUL prediction method for life prediction under different operating conditions were verified using the experiments.
3. The network model proposed in this paper is more accurate than the lifetime prediction method using a single LSTM, as evaluated with Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), and the results show that the method outperforms its single algorithm model.

The main arrangement of the remaining sections of the paper: Section 2 focuses on the theoretical analysis of the LSTM, LSTM-DNN, and domain adaptation algorithms; Section 3 focuses on the experimental details under different working conditions; Section 4 focuses on the experimental validation and analysis of the proposed algorithms, and finally Section 5 concludes the whole paper.

2. Theoretical Analysis

2.1. Definition of the Problem

This article uses X_s to represent the training task and X_t to represent the target task. Traditional RUL prediction assumes that the source and target domains follow the same distribution. However, due to different working conditions, there are distribution differences among datasets of different bearings. Therefore, when only training source domain data under one operating condition, the distribution difference between the source domain and the target domain will reduce the accuracy of the RUL prediction model. The TL method makes up for this defect. Given the source domain $D_s = \{(x_i, y_i)\}_{i=1}^{N_s}$ ($x_i \in X_s, y_i \in Y_s$) and target domain $D_t = \{x_j\}_{j=1}^{N_t}$ ($x_j \in X_t$). Among them, x_i is the i -th sample in the source domain and x_j is the j -th sample in the target domain. N_s and N_t are the total number of samples in the source and target domains, respectively. y_i is the label of the i -th sample in the source domain and Y_s is the label set of different samples in the source domain. The source domain D_s and target domain D_t have the same feature space $X_s = X_t$ and different probability distributions $P(X_s, Y_s) \neq P(X_t, Y_t)$.

2.2. LSTM Neural Networks

LSTM is the prediction of time series, for which the use of time series, in reality, is very common, such as weather prediction, health data analysis, traffic condition prediction, etc., all require the creation of time series models, which are composed of a sequence of data in time, space or other determined order [21]. Because there is continuity in time, temporal data is dynamic, and in particular certain statistics of a time series (e.g., mean, variance, etc.) are subject to dynamic temporal changes. To address this problem, traditional methods are usually modeled based on the Markov assumption that each observation on a time series depends only on the observation at its previous moment. Based on this assumption, Hidden Markov Models, Dynamic Bayesian Network Models, Kalman Filter Models, and other statistical models, such as autoregressive moving averages, have been found to be more effective in time series forecasting. With the rise of deep learning in recent years, approaches based on Recurrent Neural Networks (Recurrent Neural Networks, RNN) have achieved better results than these previous approaches. In contrast, RNNs rely on powerful neural networks that can automatically discover and model higher-order non-linear relationships in time series and enable prediction, making the RNN family of

methods very effective in solving short time series modelling. However, RNNs are prone to gradient disappearance and explosion, and cannot solve the long-time prediction problem. Therefore, the emergence of LSTM has solved the problem of RNNs, and LSTM has been better applied by constructing storage units to store long-term memory information, especially in lifetime prediction. The LSTM controls the state of the memory cell by linking the three cells of the forgetting gate, the input gate, and the output gate by means of point multiplication. The forget gate f_t is used to control whether the information in the memory cell is saved or discarded. The function of the input gate i_t is used to estimate whether to allow the input information to enter the current memory cell state. The outgoing gates O_t serve much the same purpose as the input gates and are used to determine whether the current signal will be output to the next layer. The structure of the specific LSTM network is shown in Figure 1.

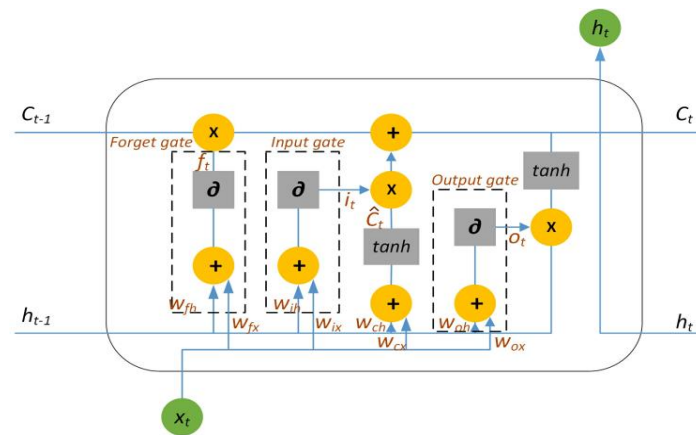


Figure 1. The structure of an LSTM memory cell.

Forgetting Gate: This forgetting gate will read the previous output h_{t-1} and the current input x_t to create a *sigmoid* non-linear mapping and then output a vector f_t

$$f_t = \text{sigmoid}(W_{fh}h_{t-1} + W_{fx}x_t + b_f) \quad (1)$$

Input gate: It mainly consists of two parts, first using h_{t-1} and x_t to determine which information to update through an operation called input gate. Then, new candidate cell information is obtained through a *tanh* layer using h_{t-1} and x_t .

$$i_t = \text{sigmoid}(W_{ih}h_{t-1} + W_{ix}x_t + b_i) \quad (2)$$

$$\widehat{C}_t = \text{tanh}(W_{ch}h_{t-1} + W_{cx}x_t + b_c) \quad (3)$$

Update cell status: Update old cell information C_{t-1} to new cell information C_t . The updated rule is to select a part of the forgotten old cell information through the forgetting gate and obtain new cell information C_t by selecting a part of the candidate cell information \widehat{C}_t to be added by the input.

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \widehat{C}_t \quad (4)$$

Output gate: After updating the cell state, it is necessary to determine which state features of the output cell based on the input h_{t-1} and x_t . Here, the input needs to pass through a *sigmoid* layer to obtain the judgment conditions. Then, the cell state is passed through the *tanh* layer to obtain a vector with values between -1 and 1 . The vector is multiplied by the judgment condition obtained by the output gate to obtain the final output of the LSTM unit.

$$O_t = \text{sigmoid}(W_{oh}h_{t-1} + W_{ox}x_t + b_o) \quad (5)$$

$$h_t = O_t \cdot \text{tanh}(C_t) \quad (6)$$

The above equation illustrates the principle of calculating the forgetting gate, the input gate, and the output gate, where c is the memory unit, C_t is the storage cell, and h_t is the hidden state. W is the weight matrix of the three gate cells and b is the threshold value. Sigmoid and tanh are the activation functions, and “ \bullet ” stands for dot product.

2.3. LSTM-DNN Prediction Model

Numerous studies have shown that shallow neural network architectures cannot accurately model the non-linear and wider range of time series and that deep neural network architectures are more generalizable than shallow architectures. Therefore, this paper proposes a prediction model based on domain adaptive LSTM-DNN. The proposed model can accurately predict its remaining life under different working conditions and has good generalization ability.

This article is mainly divided into three parts: feature extraction, domain adaptation, and RUL prediction, as shown in Figure 2. The feature extraction part mainly extracts traditional time-domain and frequency-domain features and then inputs them into the LSTM layer. The domain adaptation part mainly uses the MMD algorithm to calculate the difference in feature distribution between different domains and improves the accuracy of RUL prediction by reducing the difference between the source and target domains. The RUL prediction part inputs the features of the target domain into the trained source domain neural network model and outputs the final prediction results.

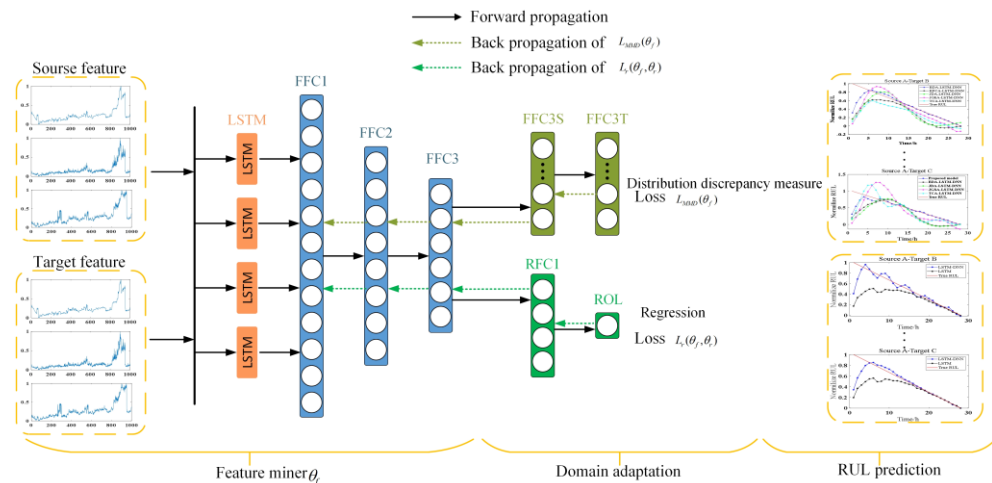


Figure 2. Prediction Flow Chart.

2.4. Domain Adaptive

The domain adaptation model proposed in this paper considers two main types of loss in combination: regression loss based on the dataset and metric loss.

(1) Maximum Mean Difference (MMD)

Maximum Mean Discrepancy (MMD) is a metric learning method in TL that calculates the distribution difference between two domains, which can reduce the distribution difference [21]. For two sets of random variables with N_s and N_t elements, respectively, the MMD distances of the two random variables are:

$$MMD^2(X, Y) = \left\| \frac{1}{N_s} \sum_{i=1}^{N_s} \phi(x_i) - \frac{1}{N_t} \sum_{j=1}^{N_t} \phi(y_j) \right\|_H^2 \quad (7)$$

N_s : Number of source domain features; the number of features in the N_t target domain. H represents the Reproducing Kernel Hilbert Space (RKHS), a kernel learning method,

and $\phi(\cdot)$ is the mapping, which is used to map the original variables into the regenerating kernel Hilbert space. MMD losses were calculated as follows [22]:

$$L_{MMD} = \frac{1}{N_s^2} \sum_{i=1}^{N_s} \sum_{j=1}^{N_s} k(x_{si}, x_{sj}) + \frac{1}{N_t^2} \sum_{i=1}^{N_t} \sum_{j=1}^{N_t} k(x_{ti}, x_{tj}) - \frac{2}{N_s N_t} \sum_{i=1}^{N_s} \sum_{j=1}^{N_t} k(x_{si}, x_{tj}) \quad (8)$$

where $k(\cdot, \cdot)$ denotes the kernel function.

(2) Training optimization

During training, mean square error (MSE) is selected as the loss function, and the formula is as follows [23]:

$$L_{regression} = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (9)$$

where m : batch size of s training set, y_i : real label; \hat{y}_i : Predictive labels.

The final optimized loss objective function is summarized as:

$$L_{total} = L_{regression} + L_{MMD} \quad (10)$$

(3) Assessment indicators

Average absolute error (MAE) and root mean square error (RMSE) are selected as the evaluation indicators of RUL prediction, and the formula is as follows:

$$MAE = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (11)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (12)$$

3. Experiments

In order to study the degradation process and remaining service life of planetary gearboxes under different operating conditions and to validate the remaining life prediction method proposed in this paper, the group built a planetary gearbox experimental bench and carried out 1003 h of planetary gearbox degradation experiments to collect the performance degradation data of planetary gearboxes from intact to failure.

The planetary gearbox test stand consists of a base, motor, magnetic powder brake, and planetary gearbox. This is shown in Figure 3. (1) The base is a steel table frame with recesses for fixing the other components of the test stand; (2) An electromagnetic speed-regulating motor (model YCT180-4A) with a rated power of 4 kW is used to provide power; (3) A speed-torque sensor (model JN338) is fitted between the output of the motor and the input of the planetary gearbox to provide the speed and torque signals of the input shaft of the planetary gearbox; (4) A magnetic powder brake (model FZJ-5) is connected to the output of the planetary gearbox to provide the load for the experimental process. The main components of the test stand are mounted coaxially on the base and are connected with couplings. The planetary gearbox used for this experiment is a single-pole planetary gearbox, type NGW11, which consists of a gear ring, a sun wheel, and three planetary wheels connected by a planetary carrier.

The star gearbox is fitted with four vibration sensors in different orientations, as shown in Figure 3. 1#, 2#, 3#, and 4# represent the positions of the first, second, third, and fourth sensors, respectively. The vibration signals are acquired at a sampling frequency (F_s) of 20 kHz and a sampling time of 12 s at 15 min intervals. The main parameters of speed and load current during the experiment were: speed of 1000 rpm and load current of 1 A. The vibration signals were collected for four operating conditions. Table 1 shows the operating

conditions of the experimental planetary gearbox, the speed was kept at 1000 rpm, and the loads were 0.8 A, 1.0 A, 1.2 A, and 1.4 A.

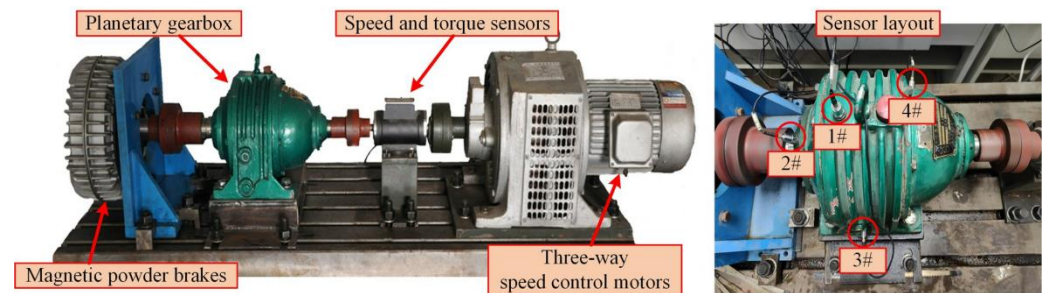


Figure 3. Planetary gearbox experiment rig.

Table 1. The description of the run-to-fail datasets.

Working Condition	Speed	Load
A	1000 rpm	1.0 A
B	1000 rpm	1.4 A
C	1000 rpm	1.2 A
D	1000 rpm	0.8 A

4. RUL Forecast

In order to achieve the prediction of RUL, an LSTM-DNN network needs to be built and the corresponding hyperparameters set. For different migration tasks, the optimal parameters of the model may vary. The parameters of the network model in this paper are mainly based on several experiments. The main parameters are: the network learning rate is set to 0.01, the number of training iterations is set to 2000, the batch size m is 128, and the weights are optimized for Stochastic Gradient Descent (SGD).

4.1. Feature Extraction and Analysis

The raw signal extracted by the sensors does not reflect the degradation trend of the planetary gearbox well, while using raw data for network training increases the network training cost and affects the final output. This paper collects vibration signal datasets from normal operation to fault failure under four operating conditions to extract useful features related to the remaining service life of the planetary gearbox; 12 degradation feature parameters are extracted in this paper. For the time domain feature parameters, information from the entire frequency band was chosen without specifying a particular band so that information from other bands would not be lost and would facilitate subsequent predictions. Table 2 shows the performance degradation feature parameters. The 12 degradation characteristics were used to form a multisensor data set as the source and target domain for the study of the remaining service life of the planetary gearbox. The performance degradation characteristics are shown in Figure 4, from which it can be seen that most of the degradation characteristics have good monotonicity and can reflect the performance degradation characteristics of the planetary gearbox.

Table 2. Classification of feature parameters.

	Characteristic Parameters
Time domain characteristic parameters	Mean Square Value, RMS, Average amplitude, Square root amplitude, Variance, Standard Variance,
Frequency domain characteristic parameters	Waveform Index, Residual clearance factor FC, RMSF, STDF

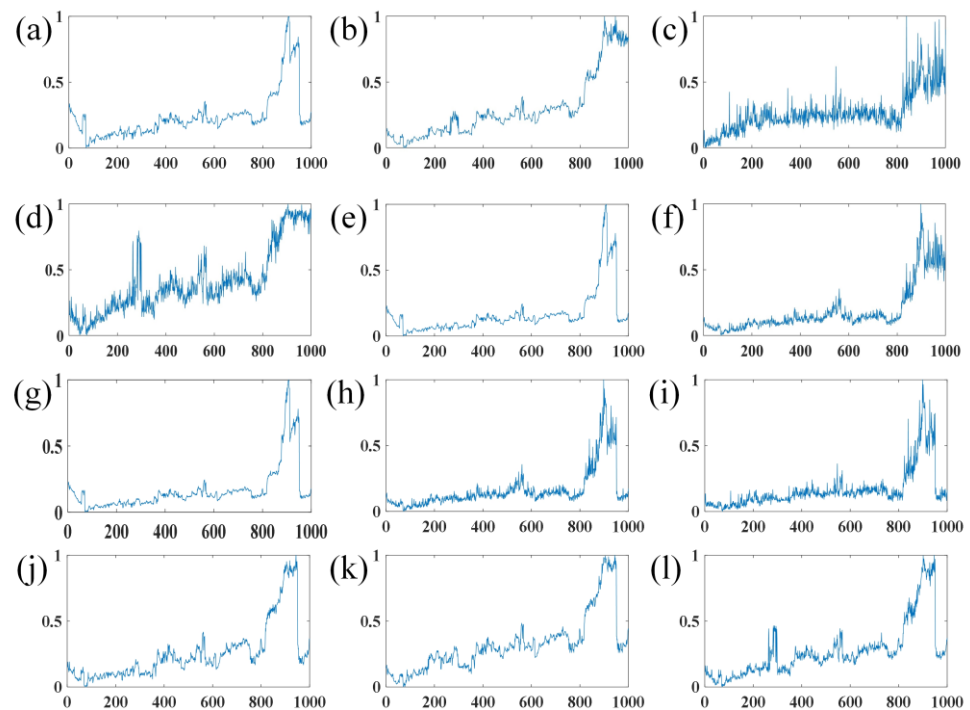


Figure 4. The particle extracted features for: (a) mean square, (b) Root Mean Square, (c) Mean amplitude; (d) Square root amplitude, (e) Variance, (f) Standard Deviation, (g) Energy, (h) Waveform index, (i) Clearance factor, (j) Frequency Center, (k) Root Mean Square frequency, (l) Frequency Standard Deviation.

4.2. Forecasting Models

The life prediction of planetary gearboxes under different operating conditions uses the data set in Table 3 to extract 12 characteristic degradation parameters for each operating condition. To obtain better experimental results, the extracted experimental data needs to be normalized by the normalization formula:

$$MF_{norm}(x) = \frac{MF(x) - MF_{min}}{MF_{max} - MF_{min}} \quad (13)$$

Table 3. Layer details of the model in this article.

	Layer Name	Details
Feature extractor	LSTM layer	Units = 96×3 , dropout = 0.8
Regression	Fully connected layer(FFC1)	Layer_size = 128
	Fully connected layer(FFC2)	Layer_size = 64
	Fully connected layer(FFC3)	Layer_size = 32
	Fully connected layer(RFC1)	Layer_size = 1
	Output layer(ROL)	Layer_size = 1

Among them, $MF(X)$: the feature set extracted for the x th time; MF_{min} : the minimum value; MF_{max} : the maximum value.

Table 3 provides the specific layer details of the proposed method in this article. In this paper, the dropout layer is added after each full connection layer, and its value is set to 0.5. ReLu is selected as the activation function of the prediction model.

4.3. Comparison with Related Transfer Learning Models

This article conducted four sets of experiments and collected four sets of operating conditions data from Table 3. This article uses the last 110 samples from each group of

operating conditions for RUL prediction. The experiment has been described earlier, with each collection lasting for 12 s and every 15 min. Therefore, the remaining 110 sample data are 28 h of data samples. This article first compares the proposed method with traditional migration methods (TCA (31), JDA (11), BDA (12), and JGSA (32)). Input the features constructed by TCA, JDA, BDA, and JGSA into the LSTM-DNN network for RUL prediction. For all LSTM-DNN models, the radial basis function is selected as the function, MSE is selected as the loss function, Adam (Adaptive Moment Estimation) is optimized as the weight value, and the learning rate is 0.001. In this article, the proposed domain adaptive LSTM-DNN model uses 96×3 units of LSTM layer and 128, 64, and 32 units of fully connected layer for feature modules. The regression module uses 16 units of a fully connected layer and one output layer. The model uses the Gaussian kernel function, optimizes the weight to SGD, and has a learning rate of 0.01. The method proposed in this article mainly utilizes labeled source domain data to train the LSTM-DNN model. Then the loss function selects MSE loss and MMD loss as the common loss function to adjust the specific parameters of the LSTM-DNN model. Finally, input the target domain dataset into the trained LSTM-DNN model and output the prediction results.

The proposed model is compared with other transfer learning methods, as shown in Figure 5. This article selects 28 time datasets from four working conditions for life prediction. It can be seen from the figure that the model proposed in this paper is superior to other transfer learning models and can adapt to different working conditions well. The transfer learning conducted under any working condition can well match the actual model. Regardless of the network model, the prediction effect between operating conditions B-C and C-B is very good because the load of operating conditions B is 1.4 A and the load of operating conditions C is 1.2 A, and the difference in operating conditions is very small. The distribution of the collected datasets is similar. The traditional transfer learning method cannot obtain a good prediction effect when there is a large difference in working conditions, such as B-D, C-D, D-B, D-C, and other cross-working conditions. Especially, TCA-LSTM-DNN exhibits negative transfer under B-D and C-D conditions. A possible reason is that the TCA method is simple, similar to PCA, which involves placing two large matrices inside, mapping them to higher dimensions, finding the minimum distance, and outputting two small matrices. Therefore, it cannot adapt well to each operating condition data. JDA, BDA, and JGSA have made a series of improvements to the classic TCA. However, when the load changes significantly, RUL prediction cannot be performed well. Although the method proposed in this article is relatively simple, it mainly consists of a feature layer and a regression layer. However, the focus of this article is to add an MMD adaptation layer after the feature layer to calculate the distance between the source and target domains. Furthermore, the focus is to add the MMD adaptation layer to the loss layer of the network for training, which can make the data distribution in the source and target domains closer. In addition, this article adds a DNN layer after LSTM to improve the fitting ability of the network, thereby improving the accuracy and stability of the network. Therefore, the method proposed in this article can adapt to RUL prediction between any cross-working conditions.

In order to accurately conclude that the proposed domain adaptive LSTM-DNN model outperforms other transfer learning models, Table 4 lists the prediction performance evaluation indicators of different transfer learning models, and the table shows that the proposed model has the smallest MAE and RMSE, which are 0.069 and 0.144, respectively, smaller than other migration models. Figure 6 shows the comparison histogram between the proposed model and other transfer learning models under different working conditions. From the figure, it can be seen that the values of MAE and RMSE of the proposed model are smaller than those of other models, which verifies that the proposed model has a better prediction effect and better generalization under different working conditions.

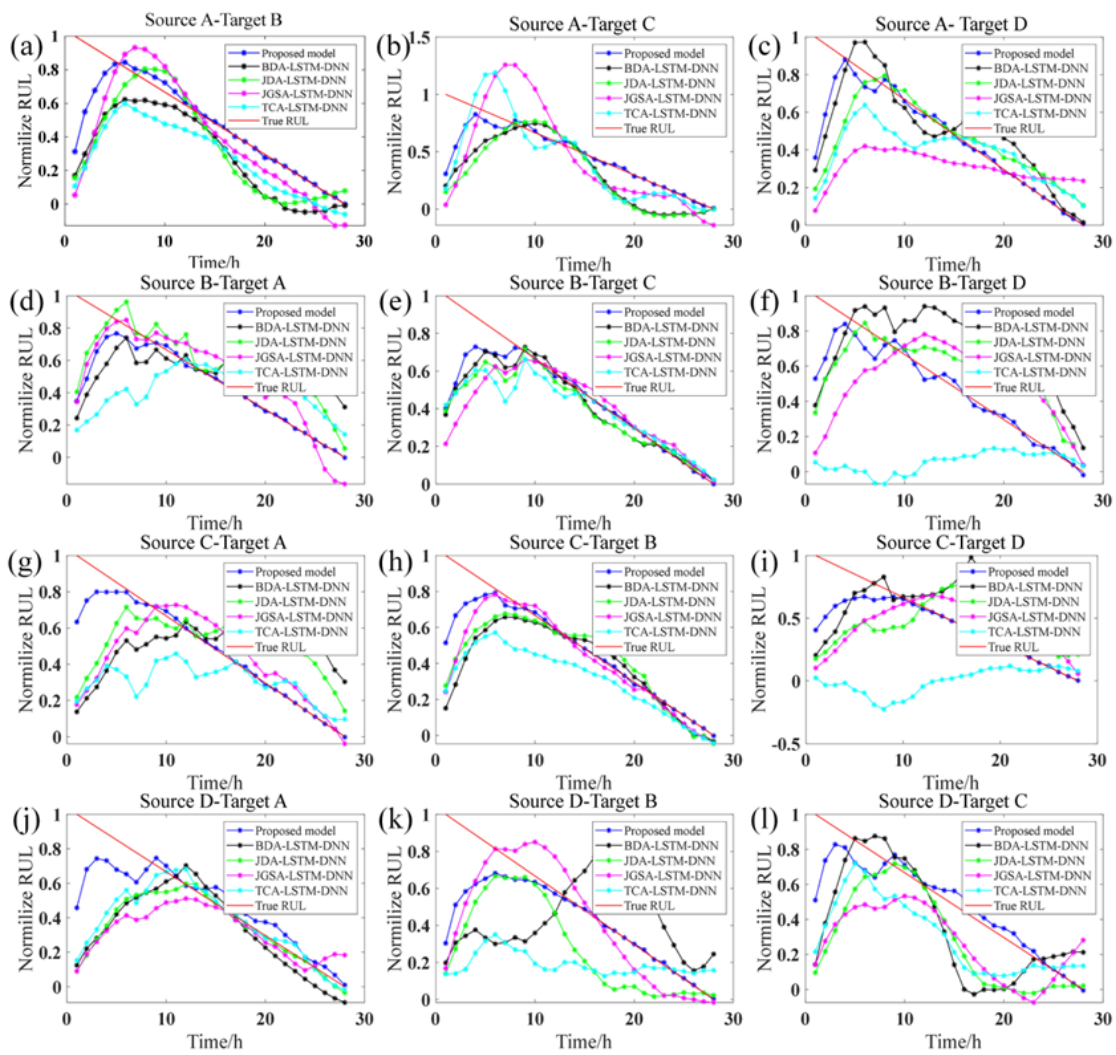


Figure 5. Comparative results of different transfer learning models. (a) RUL prediction for Source A—Target B, (b) RUL prediction for Source A—Target C, (c) RUL prediction for Source A—Target D; (d) RUL prediction for Source B—Target A, (e) RUL prediction for Source B—Target C, (f) RUL prediction for Source B—Target D, (g) RUL prediction for Source C—Target A, (h) RUL prediction for Source C—Target B, (i) RUL prediction for Source C—Target D, (j) RUL prediction for Source D—Target A, (k) RUL prediction for Source D—Target B, (l) RUL prediction for Source D—Target C.

Table 4. Predictive performance evaluation indicators of different transfer learning.

Source Domain	Target Domain	BDA-LSTM-DNN		TCA-LSTM-DNN		JDA-LSTM-DNN		JGSA-LSTM-DNN		Proposed Model	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
A	B	0.205	0.276	0.176	0.272	0.225	0.298	0.199	0.282	0.066	0.162
A	C	0.200	0.272	0.249	0.333	0.168	0.245	0.222	0.306	0.070	0.162
A	D	0.141	0.203	0.262	0.347	0.194	0.279	0.138	0.241	0.058	0.145
B	A	0.280	0.345	0.137	0.184	0.259	0.336	0.181	0.228	0.072	0.166
B	C	0.104	0.180	0.132	0.243	0.124	0.201	0.116	0.193	0.070	0.155
B	D	0.316	0.348	0.269	0.332	0.452	0.560	0.184	0.233	0.062	0.118
C	A	0.335	0.397	0.178	0.274	0.238	0.332	0.237	0.300	0.038	0.087
C	B	0.136	0.248	0.093	0.196	0.191	0.257	0.129	0.211	0.050	0.118
C	D	0.332	0.397	0.270	0.345	0.512	0.626	0.323	0.367	0.083	0.170
D	A	0.197	0.303	0.214	0.329	0.151	0.275	0.165	0.302	0.097	0.145
D	B	0.317	0.368	0.158	0.239	0.340	0.415	0.226	0.298	0.090	0.185
D	C	0.210	0.288	0.270	0.324	0.223	0.275	0.226	0.314	0.071	0.119
Total		0.231	0.302	0.201	0.285	0.256	0.342	0.196	0.273	0.069	0.144

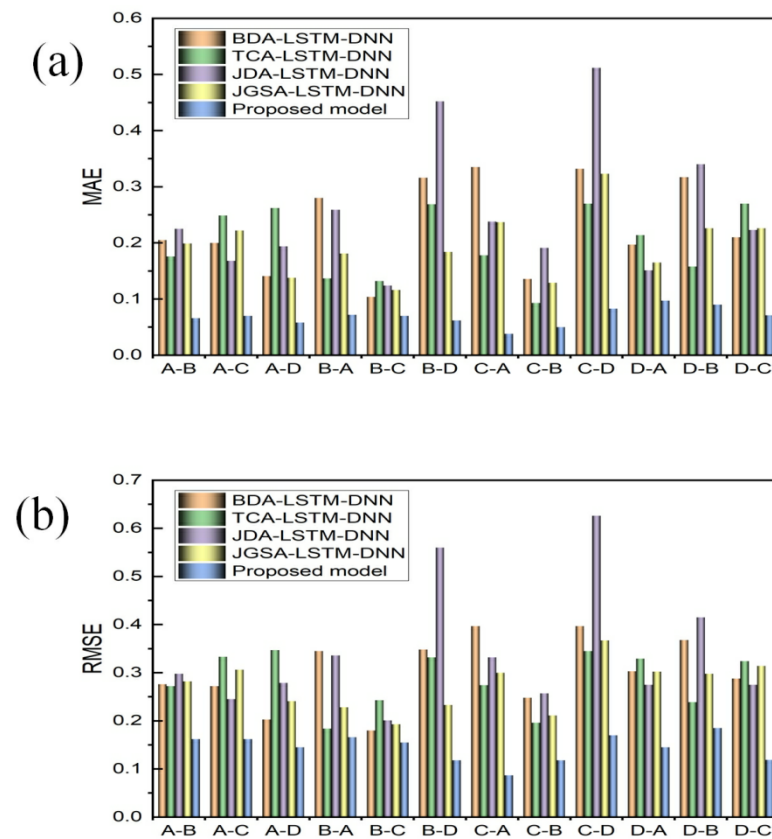


Figure 6. Error comparison histogram of different transfer learning models: MAE, RMSE. (a) MAE pair ratio of different Transfer learning algorithms; (b) RMSE pair ratio of different Transfer learning algorithms.

4.4. Comparison with LSTM Model

In order to better verify the effectiveness of the proposed model, this section compares the lifetime prediction results of the proposed network model with the domain adaptive LSTM model, as shown in Figure 7, which shows that the proposed model matches the actual results better than the LSTM model. From Figure 7, it can be concluded that using only the LSTM model has lower prediction accuracy. Because in LSTM networks, adding a DNN layer can increase the depth of the network, thereby improving its expressive and fitting abilities. Specifically, each time step of the LSTM network outputs a state vector, which typically contains the feature information of the input sequence for that time step. However, these vectors are usually high-dimensional and may not necessarily be the optimal feature representations. Therefore, adding a DNN layer to further process these vectors, mapping them to lower dimensional spaces, and extracting more discriminative features.

In addition, adding a DNN network after LSTM can also increase the non-linearity of the network, thereby improving its fitting ability. In deep learning, non-linear transformations are very important as they allow neural networks to approximate any complex function. Therefore, by adding a DNN network, the depth and non-linearity of the network can be increased, thereby improving the performance of the LSTM network.

The performance indicators of the proposed network model and the domain adaptive LSTM network model are shown in Table 5. It can be concluded from the table that the MAE and RMSE of the proposed network model are smaller than those of the domain adaptive LSTM network model for life prediction under different operating conditions. The comparative histograms of MAE and RMSE of the network model proposed in this paper and the domain adaptive LSTM network model are shown in Figure 8, through which it can be seen that the values of MAE and RMSE with the addition of the DNN structure are smaller than those without the DNN structure, verifying the superiority of

the model proposed in this paper and reflecting the fact that in general, shallow neural network structures cannot accurately model non-linear. It also reflects that, in general, shallow neural network structures cannot accurately model non-linear and wider range of time series and that deep neural network architectures have more generalization and prediction capabilities compared to shallow architectures.

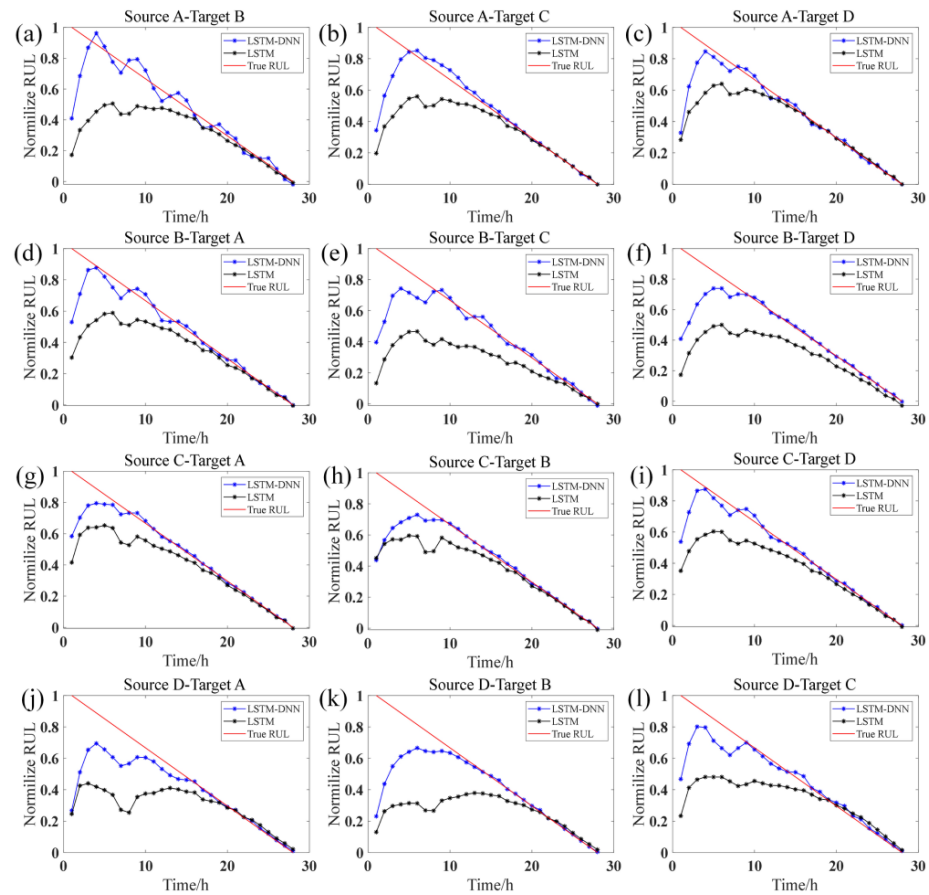


Figure 7. Comparison results between LSTM-DNN and LSTM models. Comparative results of different transfer learning models. (a) RUL prediction for Source A—Target B, (b) RUL prediction for Source A—Target C, (c) RUL prediction for Source A—Target D; (d) RUL prediction for Source B—Target A, (e) RUL prediction for Source B—Target C, (f) RUL prediction for Source B—Target D, (g) RUL prediction for Source C—Target A, (h) RUL prediction for Source C—Target B, (i) RUL prediction for Source C—Target D, (j) RUL prediction for Source D—Target A, (k) RUL prediction for Source D—Target B, (l) RUL prediction for Source D—Target C.

Table 5. RUL predicted performance indicators compared with LSTM.

Source Domain	Target Domain	LSTM		Proposed Model	
		MAE	RMSE	MAE	RMSE
A	B	0.175	0.275	0.065	0.130
A	C	0.147	0.251	0.065	0.154
A	D	0.110	0.206	0.057	0.147
B	A	0.142	0.223	0.047	0.105
B	C	0.228	0.311	0.076	0.156
B	D	0.201	0.284	0.070	0.158
C	A	0.111	0.174	0.042	0.100
C	B	0.122	0.191	0.068	0.150
C	D	0.134	0.206	0.040	0.101
D	A	0.211	0.303	0.106	0.193
D	B	0.242	0.351	0.098	0.206
D	C	0.179	0.263	0.066	0.128

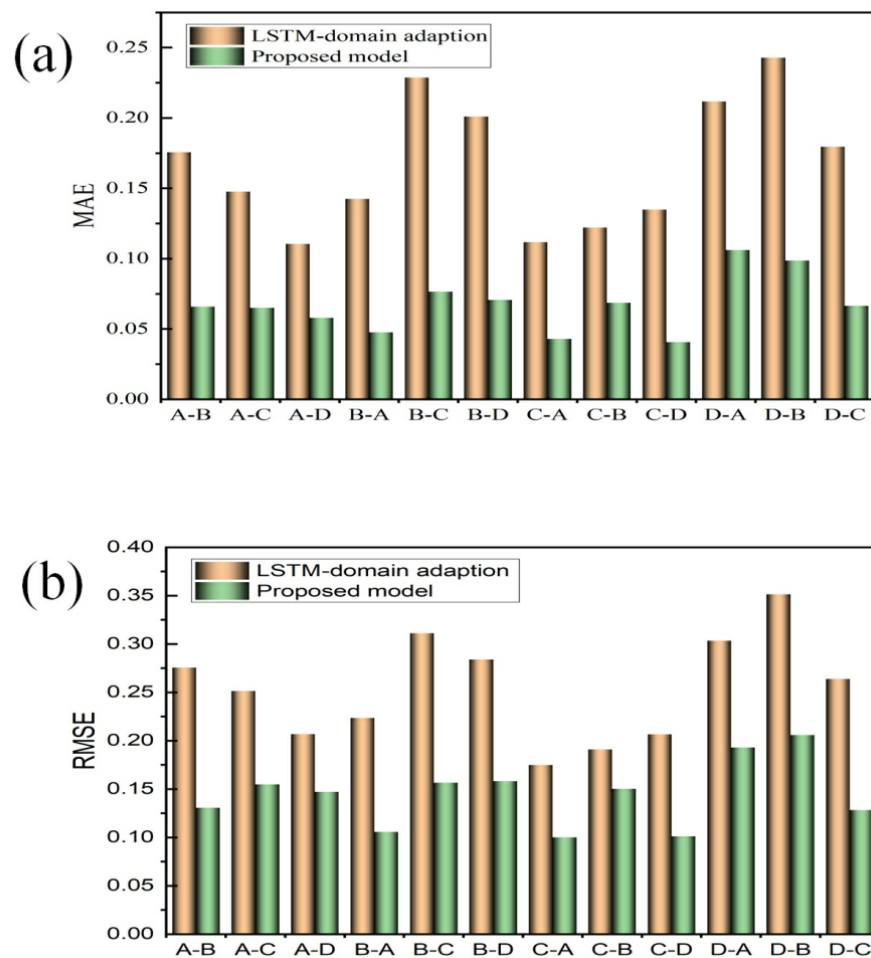


Figure 8. Comparison of MAE and RMSE between LSTM-DNN and LSTM. (a) Compare the MAE evaluation indicators between the LSTM domain adaptive method and the method proposed in this article; (b) Compare the RMSE evaluation indicators between the LSTM domain adaptive method and the method proposed in this article.

5. Conclusions

In this paper, a life prediction method for planetary gearboxes under different operating conditions is proposed. The domain-adaptive LSTM-DNN network proposed in this paper is validated against four other different prediction models on an experimental data set in our laboratory, and the effectiveness and superiority of the model proposed in this paper are derived. The main conclusions of this paper are as follows:

1. Compared with traditional transfer learning models, the method not only accurately predicts the remaining lifespan but also has better feature extraction capability and enhanced adaptivity of data distribution compared to other adaptive methods.
2. The model proposed in this paper can effectively extract degradation features from condition monitoring data under various operating conditions. Through domain adaption, the generalization capability of the data-driven RUL prediction model can be effectively improved, and to a certain extent, it can adapt to the RUL prediction tasks under different operating conditions, making up for the limitations of the traditional data-driven model.
3. The model has made predictions under different working conditions and obtained better prediction results, which facilitate engineers to make effective maintenance plans in advance and shorten maintenance intervals, which can further save maintenance costs, thus realizing planetary gearbox life prediction and health management.

In the future, we will make further improvements in the following areas. (1) Loss functions such as classification and CORAL are not considered in this paper, and multiple loss functions will be introduced in this network model for prediction in the next step. (2) Further refinement of the proposed models and algorithms to obtain more practical applications.

Author Contributions: Conceptualization, Z.L., C.T. and H.L.; Methodology, C.T., B.C. and X.Z.; Software, Z.L., H.L. and X.Z.; Validation, Z.L. and C.T.; Formal analysis, X.Z.; Investigation, Z.L., Y.L. and H.L.; Resources, C.T.; Data curation, Z.L. and B.C.; Writing—original draft, Z.L.; Writing—review & editing, C.T.; Visualization, Z.L. and Y.L.; Supervision, C.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data involved in this article have been presented in the article.

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

References

1. Wu, Z.; Bai, H.; Yan, H.; Zhan, X.; Guo, C.; Jia, X. Intelligent Fault Diagnosis Method for Gearboxes Based on Deep Transfer Learning. *Processes* **2023**, *11*, 68. [\[CrossRef\]](#)
2. Lu, Y.-W.; Hsu, C.-Y.; Huang, K.-C. An Autoencoder Gated Recurrent Unit for Remaining Useful Life Prediction. *Processes* **2020**, *8*, 1155. [\[CrossRef\]](#)
3. Chen, Y.; Zhang, D.; Zhang, W.A. MSWR-LRCN: A new deep learning approach to remaining useful life estimation of bearings. *Control Eng. Pract.* **2022**, *118*, 104969. [\[CrossRef\]](#)
4. Fu, J.; Chu, J.C.; Guo, P.; Chen, Z. Condition Monitoring of Wind Turbine Gearbox Bearing Based on Deep Learning Model. *IEEE Access* **2019**, *7*, 57078–57087. [\[CrossRef\]](#)
5. He, K.; Su, Z.; Tian, X.; Yu, H.; Luo, M. RUL prediction of wind turbine gearbox bearings based on self-calibration temporal convolutional network. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3501912. [\[CrossRef\]](#)
6. Mao, W.; He, J.; Zuo, M.J. Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 1594–1608. [\[CrossRef\]](#)
7. Pan, S.J.; Tsang, I.W.; Kwok, J.T.; Yang, Q. Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* **2010**, *22*, 199–210. [\[CrossRef\]](#)
8. Dorri, F.; Ghodsi, A. Adapting Component Analysis. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining, Brussels, Belgium, 10–13 December 2012; pp. 846–851.
9. Duan, L.; Tsang, I.W.; Xu, D. Domain transfer multiple kernel learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 465–479. [\[CrossRef\]](#)
10. Hsiao, P.H.; Chang, F.J.; Lin, Y.Y. Learning discriminatively reconstructed source data for object recognition with few examples. *IEEE Trans. Image Process.* **2016**, *25*, 3518–3532. [\[CrossRef\]](#)
11. 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, Sydney, NSW, Australia, 1–8 December 2013; pp. 2200–2207.
12. Wang, J.; Chen, Y.; Feng, W.; Yu, H.; Huang, M.; Yang, Q. Transfer learning with dynamic distribution adaptation. *ACM Trans. Intell. Syst. Technol. (TIST)* **2020**, *11*, 1–25. [\[CrossRef\]](#)
13. Ding, Y.; Ding, P.; Zhao, X.; Cao, Y.; Jia, M. Transfer learning for remaining useful life prediction across operating conditions based on multisource domain adaptation. *IEEE/ASME Trans. Mechatron.* **2022**, *27*, 4143–4152. [\[CrossRef\]](#)
14. Cheng, H.; Kong, X.; Wang, Q.; Ma, H.; Yang, S.; Chen, G. Deep transfer learning based on dynamic domain adaptation for remaining useful life prediction under different working conditions. *J. Intell. Manuf.* **2023**, *34*, 587–613. [\[CrossRef\]](#)
15. Mao, W.; Liu, J.; Chen, J.; Liang, X. An interpretable deep transfer learning-based remaining useful life prediction approach for bearings with selective degradation knowledge fusion. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 3508616. [\[CrossRef\]](#)
16. Xie, B.; Duan, Z.; Zheng, B.; Liu, L. Research on target object recognition based on transfer-learning convolutional SAE in intelligent urban construction. *IEEE Access* **2019**, *7*, 125357–125368. [\[CrossRef\]](#)
17. Li, C.; Zhang, S.; Qin, Y.; Estupinan, E. A systematic review of deep transfer learning for machinery fault diagnosis. *Neurocomputing* **2020**, *407*, 121–135. [\[CrossRef\]](#)
18. Han, T.; Zhou, T.; Xiang, Y.; Jiang, D. Cross-machine intelligent fault diagnosis of gearbox based on deep learning and parameter transfer. *Struct. Control Health Monit.* **2021**, *29*, e2898. [\[CrossRef\]](#)
19. Wang, F.; Gomez, W.; Amogne, Z.E.; Rahardjo, B. Transfer learning based deep learning model and control chart for bearing useful life prediction. *Qual. Reliab. Eng. Int.* **2023**, *39*, 837–852. [\[CrossRef\]](#)
20. Wang, L.; Liu, H.; Pan, Z.; Fan, D.; Zhou, C.; Wang, Z. Long Short-Term Memory Neural Network with Transfer Learning and Ensemble Learning for Remaining Useful Life Prediction. *Sensors* **2022**, *22*, 5744. [\[CrossRef\]](#)

21. Yan, H.; Bai, H.; Zhan, X.; Wu, Z.; Wen, L.; Jia, X. Combination of VMD Mapping MFCC and LSTM: A New Acoustic Fault Diagnosis Method of Diesel Engine. *Sensors* **2022**, *22*, 8325. [[CrossRef](#)] [[PubMed](#)]
22. Borgwardt, K.M.; Gretton, A.; Rasch, M.J.; Kriegel, H.P.; Schölkopf, B.; Smola, A.J. Integrating structured biological data by Kernel Maximum Mean Discrepancy. *Bioinformatics* **2006**, *22*, e49–e57. [[CrossRef](#)]
23. Gretton, A.; Borgwardt, K.; Rasch, M.; Schölkopf, B.; Smola, A. A Kernel method for the two-sample problem. *arXiv* **2008**, arXiv:0805.2368.

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