A dynamic separable convolution RCNN for lubrication condition identification of planetary roller screw mechanism

Wei Cai¹, Geng Liu¹, Shangjun Ma¹, Xiaojun Fu¹, Jianxin Zhang¹,² and Maodong Niu¹

Abstract
Lubrication condition has a strong effect on the service ability of planetary roller screw mechanism (PRSM), so how to effectively identify the lubrication condition of PRSM is highly important in practical industrial applications. A dynamic separable convolution residual convolutional neural network (DSC-RCNN) method is proposed in this paper for the lubrication condition identification of PRSM. In the proposed method, a dynamic separable convolution (DSC) is developed by adopting depthwise separable convolution and dynamic convolution. To verify the learning competence of the proposed method, the PRSM failure test is carried out firstly and vibration data of the PRSM with and without grease are collected in multiple working conditions. Then, three experiments are implemented. The first one is to obtain the optimal number of the depthwise separable convolution in the DSC. The second one compares the effect of the DSC unit and the dynamic convolution unit on the diagnosis capacity of the proposed method. The last one compares SVM, BSA-SVM, AEs, LSSVM, LSTM, VGG-13, and the proposed method. The results reveal the best number of the depthwise separable convolution, the optimal unit of the proposed model and indicate that the DSC-RCNN has enormous recognition and transfer learning abilities.

Keywords
Planetary roller screw mechanism, lubrication condition identification, transfer learning, dynamic separable convolution, residual convolutional neural network

Date received: 23 November 2021; accepted: 4 March 2022

Handling Editor: Chenhui Liang

Introduction
Planetary roller screw mechanism (PRSM) is one of the most important and frequently used components in the electromechanical actuator (EMA), which plays a critical role in precision machine tools,¹,² robots,³ medical equipment,⁴ and so on. Effectively detecting the faults of PRSM can provide an assurance for the reliability of machine equipment.

Figure 1 shows the structure of PRSM, containing a screw, a nut, a group of rollers, two annular gears, two carriers, and two circlips. The screw, nut, and rollers

¹Shaanxi Engineering Laboratory for Transmissions and Controls, Northwestern Polytechnical University, Xi'an, China
²Hubei Jiangshan Heavy Industries Co., Ltd, Xiangyang, China

Corresponding authors:
Geng Liu, Shaanxi Engineering Laboratory for Transmissions and Controls, Northwestern Polytechnical University, No. 127 Youyi Xilu, Xi'an 710072, China.
Email: npuliug@nwpu.edu.cn
Shangjun Ma, Shaanxi Engineering Laboratory for Transmissions and Controls, Northwestern Polytechnical University, Youyi Xilu No. 127, Xi'an 710072, China.
Email: mashangjun@nwpu.edu.cn

Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (https://creativecommons.org/licenses/by/4.0/) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).
with precision threads are key parts that transfer motion and force. With the increase of external load and operating speed, the grease performance on the surface of those transmission parts will degrade gradually. It leads to accelerate wear of those components and the transmission precision loss of PRSM. Hence the lubrication condition has a crucial influence on the performance of the PRSM. In general, detecting the lubrication condition of PRSM is difficult because the transmission components are installed inside the EMA. Therefore, it is essential to establish an intelligent model to achieve the aforementioned purpose.

In the last decades, researches on the PRSM have mainly focused on load distribution,5–7 meshing principle,8 thermal characteristic analysis,9,10 transmission accuracy,11 and so on. Du et al.7 established a load distribution model considering the effect of incorporate radial load and machining error. The results showed the load distribution and fatigue life dramatically change with machining errors. Jones and Velinsky8 used the principle of conjugate surfaces to contact at the screw/roller and nut/roller interfaces in PRSM. It derives the radii of contact on the roller, screw, and nut bodies and makes us know the meshing position of PRSM. Qiao et al.9 founded a thermal model based on the thermal network method and confirmed the thermal characteristics of the PRSM in various working conditions experimentally. Nevertheless, limited literature has been reported in terms of lubrication condition identification of PRSM. Niu et al.13 developed a BSA-SVM based on a bird swarm algorithm (BSA) and support vector machine (SVM)14 to discern the lubrication condition of PRSM. Feature data were extracted artificially from the time domain, frequency domain, and time-frequency domain, respectively, as the input of SVM. BSA was applied to optimize main parameters in the SVM model. Although this method has good recognition ability, it depends strongly on feature data. It will cause that the performance of SVM is directly determined by the quality of the feature data. Therefore, an end-to-end intelligent recognition method which doesn’t rely on manually extracted features needs to be applied in recognizing lubrication condition of PRSM.

At present, plenty of works using machine learning methods to achieve the identification of different states only aim at the same distribution condition. Ezz-Eldin et al.15 invented a model based on hybrid convolutional neural networks (CNN) and feedforward deep neural networks to automatic speech emotional-speech recognition. Jin et al.16 created a decoupling attentional residual network for bearing compound fault diagnosis. In those works, the training, validating, and testing data were obtained from the same dataset. The models trained in the same distribution may not get good identification performance in others. Especially for the PRSM, the data in different working conditions have different distributions. It implies that the models trained in one working condition are not suitable for others. In addition, it is difficult for the PRSM to acquire data under various working conditions. Therefore, it is significant to handle the domain shift problem about recognizing lubrication condition of the PRSM. This knowledge learning process in different domains is denoted as transfer learning. Figure 2 displays the differences between machine learning and transfer learning. The nature distinction is that the training and test data in the transfer learning are from different domains in the different distributions. Consequently, how to improve the performance of transfer learning is crucial to an intelligent recognition method.

Therefore, to settle the above needs this paper develops a novel method based on dynamic separable convolution RCNN (DSC-RCNN) to identify the different states of grease in the PRSM. The dataset with and without lubrication is collected from the PRSM fault bench test to explore the proposed method’s discerning competence and transfer learning ability in different working conditions. The influence of the parameter $k$ in the dynamic separable convolution on the developed method is discussed. The performance of the developed method is compared with SVM, BSA-SVM, autoencoders (AEs),17 least square SVM (LSSVM),18 long short-term memory (LSTM),19 and VGG-13.20 The highlights of this paper are summarized below.

1. An end-to-end intelligent learning method is proposed for the lubrication condition monitoring of PRSM in different working conditions.

2. A dynamic separable convolution is developed to increase the representation capability without increasing the extra parameters.
(3) A basis unit based on a dynamic separable convolution and a shortcut layer is introduced into the proposed method in order to enrich the extracted characteristics.

(4) The transfer learning of the proposed method is validated for lubrication condition monitoring of PRSM and compared with the existing techniques in a real-world dataset.

After introducing the characteristics of PRSM and the importance of lubrication conditions to work performance of the PRSM in Section 1, Section 2 mainly proposes a dynamic separable convolution (DSC) instead of traditional convolution to improve the prediction performance. Section 3 briefly describes the DSC-RCNN structure based on DSC blocks and pooling layers. The experiment of PRSM in working conditions, including different loads and speeds, is carried out and provides data to verify the learning ability of the corresponding proposed method in Section 4. And the performance of the proposed method is experimentally validated. Finally, the conclusions are presented in Section 5.

**Dynamic separable convolution**

In this section, a novel dynamic separable convolution based on depthwise separable convolution and attention is introduced to provide better trade-off between network performance and computational burden. Figure 3 describes the inner structure of DSC in detail. The DSC mainly is made up of attention, point convolution, depth convolution, Batch Normalization (BN), and ReLU. The attention includes two standard
convolution layers, a dimensionality reduction layer and an activation function layer.

The coefficient vector \( C \) of DSC is defined as follows.

\[
C = g(r(W_1^T \cdot (W_1^T \cdot X + b_1) + b_2))
\]

\[
0 \leq C_k(x) \leq 1, \quad \sum_{k=1}^{k} C_k(x) = 1
\]

where \( X \) represents the input of the DSC, \( W_1^T, W_2^T \) and \( b_1, b_2 \) are weight matrix and bias vector, \( g \) and \( r \) are the softmax activation function and dimensionality reduction function, \( k \) is the number of the depthwise separable convolution.

\[
Y = h(w(C, W_p \cdot (C, W_d \cdot X + b_d) + b_p))
\]

where \( Y \) is the output of the DSC, \( W_p \) and \( W_d \) are weight matrix of point convolution and depth convolution, \( b_p \) and \( b_d \) are bias vectors correspondingly, \( h \) is normalized function in the BN, \( w \) is ReLU activation function in ReLU. Note that the coefficient \( C(x) \) is a function related to the input \( X \). Moreover, depth convolutions share the same coefficient \( C(x) \) with point convolutions presented in Figure 3.

In the DSC the depth convolution and the point convolution are 2D standard convolution. The number of the convolution kernels in the depth convolution is same as the channel number of the input \( X \). And the channel number of the depth convolution is 1. The kernel size of point convolution is \( 1 \times 1 \). The output channel of the point convolution is \( k \). The depthwise separable convolution in DSC consists of a depth convolution and a point convolution. In the DSC replacing standard convolution with depthwise separable convolution effectively reduce the computational cost and model parameters.

In addition, the DSC adopts two standard convolution operations to obtain coefficient vector \( C \) varied with the input \( X \). The output channel number of the first convolution Conv1 is the same as that of the input. The size of the convolution kernel is equal to that of the input. The output channel number of the first convolution Conv2 is \( k \). The convolution kernel size is \( 1 \times 1 \). The dimension of the output after two convolution operations must be reduced from three to one dimension by dimensionality reduction function \( r \). Those coefficients choose the optimal aggregation of linear models for the given input \( X \). Thus, the aggregated model is a non-linear function with excellent characteristic representation ability.

**Network architecture design**

This subsection introduces an end-to-end network structure for PRSM fault signal classification problems. The flow chart is presented in Figure 4. This RCNN method mainly includes eight DSC blocks, four pooling layers, a BN layer, and a fully connected (FC) layer.

**DSC block**

Convolution operation can extract and combine feature data hierarchically. With increasing convolution layers, the extracted information is richer and more relevant. However, the training of deep neural networks is not a process of stacking simply. The deeper the network, the more prone it is to gradient explosions and gradient disappearance. Consequently, this paper employs a shortcut layer in the DSC for addressing those aforementioned problems.

As shown in Figure 5, two kinds of basic units for the CNN method are taken into account. The DSC unit
comprises a shortcut layer, a DSC, a BN layer, and a ReLU layer.

The output of DSC unit is described as follows.

$$Y_{DSC} = w(h(w(C_{DSC}(X))W_p(C_{DSC}(X)\frac{W_s}{C_{15}}X + b_p)) + \frac{W_s}{C_{15}}X + b_s)$$  \hspace{1cm} (4)

where $C_{DSC}(X)$ is the weight coefficient in DSC operation, $W_s$ and $b_s$ represent the weight vector and bias vector of the shortcut layer, respectively, $w'$ and $h'$ are the normalized function and the ReLU activation function in the DSC unit.

In the DSC, a shortcut layer adds the raw low-level features directly across the multilayer network to the high-level features to enrich the extracted data features. The output of the shortcut layer is described as follows.

$$Y_s = W_sX + b_s$$  \hspace{1cm} (5)

Whereas the dynamic convolution (DC)21 unit in Figure 5 adopts a shortcut layer combined with a DC that replaces separable convolution with standard convolution.

The output of the DC unit is given as follows.

$$Y_{DC} = w(h(w(C_{DC}(X))W_{DC} + b_{DC}) + (W_sX + b_s))$$  \hspace{1cm} (6)

where $C_{DC}(X)$ is the weight coefficient in DC operation, $W_{DC}$ and $W_s$ is the weight vector in DC and shortcut layer, $b_{DC}$ and $b_s$ is the bias vector, respectively.

In Figure 5 it is clear that the output channels of the DSC or DC and the shortcut layer are the same to do addition. As far as we all know, the BN layer and the ReLU layer don’t change the output size. Hence, the output channels of those units are equal to that of the DSC or DC. In addition, the kernel size of the shortcut layer is $1 \times 1$ to obtain the raw low-level features as possible.

The DSC/DC block is made up of two those basic units in Figure 6. The size of input features is the same as that of output features. Only the channel number of input features changes from $C$ to $O$.

**Reshaped data**

Due to the two-dimensional convolution operation in this paper, the one-dimensional input must be reshaped to a two-dimensional signal. The processing methods, such as wavelet packet transform, wavelet transform, and Fourier transform, are adopted to artificially extract features and reshape data. Although those methods are applied successfully in video processing and fault diagnosis, the processing inevitably increases the computation time. Hence, in this paper the two-dimensional samples are formed by directly taking a
sort segmental data from a raw signal and organizing it in a row fashion.

**Pooling layer**

A pooling layer\(^{22,23}\) is used widely in neural networks to realize data volume decline and relieve computational pressure. In this paper, an average pooling layer is utilized because it doesn’t need parameters and can reduce the risk of overfitting in the training of the neural network.

**Batch normalization layer**

In neural network, a BN\(^{21}\) layer can make the distribution of eigenvalues normalized. This could not only speed up the training process but also accelerate the convergence of the network model. The process of BN layer is described as follows.

\[
\begin{align*}
\bar{x}_k &= \frac{x_k - E(x_k)}{\sqrt{\text{Var}(x_k)}} \\
y_k &= \gamma_k \bar{x}_k + \beta_k
\end{align*}
\]

where \(E(x_k)\) and \(\text{Var}(x_k)\) represent the mean and variance of the eigenvalues of a layer, respectively. \(\gamma_k\) and \(\beta_k\) are a pair of parameters which scale and shift the normalized value \(y_k\).

**Fully connected layer**

Generally, a FC layer\(^{24}\) is placed in the last layer in neural network to map the data features learned by the front layers into sample space. So a FC layer in this paper is utilized to reduce data dimensions and output classification results.

**Cross-entropy loss**

Cross-entropy loss\(^{25}\) is usually used to describe the distance between the real probability distribution and the predicted probability distribution in classification tasks.

\[
H(p, q) = - \sum_{i=1}^{n} p(x_i) \log(q(x_i))
\]

where \(p(x_i)\) is the real label for the training set and \(q(x_i)\) is the label value predicted by the network.

**Network parameters**

The network parameters of the proposed method and the feature map size after each layer are clearly exhibited in Table 1. \(K\) and \(C\) are kernel sizes and convolution channels.

**Experiments**

**Data collection**

To evaluate the performance of the proposed recognition method, the experiment of PRSM in two states is carried out. Figure 7 indicates the lubrication states of PRSM with and without grease.

The data is collected from the PRSM failure test bench, as shown in Figure 8. The test bench is composed of a servo motor, a reducer, a PRSM, a vibration acceleration sensor, and a hydraulic system. The servo motor is connected to the input end of the PRSM through the reducer to provide power for the PRSM.

| Input | Architecture parameters \((K/C)\) | Feature map size |
|-------|---------------------------------|-----------------|
| Layer 1 | 3 x 3 conv/3 stride = 1 | 3 x 32 x 32 |
| Layer 2 | 3 x 3 conv/6 stride = 1 | 6 x 32 x 32 |
| Layer 3 | 3 x 3 conv/12 stride = 1 | 12 x 16 x 16 |
| Layer 4 | 3 x 3 conv/24 stride = 1 | 24 x 8 x 8 |
| Layer 5 | 3 x 3 conv/48 stride = 1 | 48 x 4 x 4 |
| Layer 6 | 3 x 3 conv/96 stride = 1 | 96 x 4 x 4 |
| Layer 7 | 3 x 3 conv/192 stride = 1 | 192 x 2 x 2 |
| Layer 8 | 3 x 3 conv/384 stride = 1 | 384 x 2 x 2 |
| Layer 9 | 3 x 3 conv/768 stride = 1 | 768 x 2 x 2 |

Table 1. Network parameters of dynamic separable convolution RCNN.
The hydraulic system is directly connected with the output end of the PRSM to simulate the loading process of the PRSM. During the experiment, the PRSM drives the hydraulic system to produce reciprocating motion so as to realize active loading. The vibration acceleration sensor is placed on the nut of the PRSM to collect the vibration characteristics of the whole system as it moves. The rated power of the servo motor is 5000 W. The diameter of the screw in the PRSM is 25 mm.

The sampling frequency of the data acquisition card is set to 20,480 Hz. Data acquisition is carried on every 0.1 s. Each vibration signal sample contains 2048 data points. The collected data has three channels, respectively representing the three directions of PRSM. Eight working conditions are listed in Table 2. In the experimental process, the movement of the PRSM is controlled to make its displacement curve appear triangle wave so as to ensure that its speed is uniform in the reciprocating movement. The stroke of the reciprocating motion is set to 70 mm. The raw signal of PRSM without and with grease is collected as shown in Figures 9 and 10.

Data preparation

Due to the limited data samples, data augmentation is used to increase samples in the training process. This method makes use of overlapping samples and shift transformation. As shown in Figure 11, the process of overlapping sample is described vividly. Every two adjacent samples have an overlapping sample. The number of the overlapping samples is shift length. In this paper, the shift length is 128, the sample data length is set to 1024, and the sample number changes from 60 to 480. To avoid occasionality, 80% of the dataset in each working condition is randomly divided into the training samples, 10% of the dataset is validating samples, and the rest is testing samples. As a result, the training, validating, and testing samples are totally 6144, 768, and 768.

To test the transfer learning competence of the proposed model, one of the eight working conditions in turn is treated as a testing set. The rest is divided into a training set and validating set according to the ratio of 9:1. So the sample numbers of a training set, validating set and testing set are 6048, 672, and 960.

### Discussion and analysis

The effect of the number of the depthwise separable convolution

The effect of the number $k$ of the depthwise separable convolution on the proposed network is discussed because of having a great influence on the complexity of the model and the training efficiency. Table 2 lists the working conditions of the PRSM.

![Figure 7. The lubrication states of PRSM: (a) without grease and (b) with grease.](image)

![Figure 8. The PRSM failure test bench.](image)

| Working state | Load (kN) | Speed (r/min) | Samples | Working state | Load (kN) | Speed (r/min) | Samples |
|---------------|-----------|---------------|---------|---------------|-----------|---------------|---------|
| With grease   | 3         | 30            | 60      | Without grease | 3         | 30            | 60      |
|               | 90        | 60            |         |               | 90        | 60            |         |
| With grease   | 6         | 30            | 60      |               | 6         | 30            | 60      |
|               | 60        | 60            |         |               | 60        | 60            |         |
|               | 90        | 60            |         |               | 90        | 60            |         |
| With grease   | 9         | 30            | 60      |               | 9         | 30            | 60      |
|               | 60        | 60            |         |               | 60        | 60            |         |
|               | 90        | 60            |         |               | 90        | 60            |         |
Figure 9. The raw signal of PRSM without grease.
Figure 10. The raw signal of PRSM with grease.
of the model. Considering the scale and effectiveness of the network, predicted results of the model are analyzed when \( k \) is 2, 4, 6, and 8.

The models in different \( k \) are examined by testing set. Figure 12 shows the testing accuracy and the area under receiver operating characteristic (ROC) curves in different \( k \). It is evident in Figure 12 that this method can precisely discern the lubrication condition and the testing accuracies in different \( k \) are 100%, 100%, 100%, and 99.74%, respectively. In addition, Figure 12 shows the area under ROC curve (AUC) of the different \( k \) is 1. It suggests that the proposed method has good diagnostic authenticity in spite of the different \( k \).

In different working conditions, the testing accuracy of transfer learning in different \( k \) is shown in Table 3. The testing accuracies of the different working conditions are higher than 97%. It indicates that the proposed method also has good transfer learning ability. Particularly, when \( k \) is 8, the testing accuracies in different working conditions are all higher than 98% and the fluctuation of testing accuracies is smaller. Figure 13 describes the testing AUC of transfer learning in different \( k \). The values of AUC in different \( k \) are higher than 0.994. It implies that the models of the different \( k \) all have very reliable identification ability.

Although the average accuracy of \( k = 8 \) in transfer learning is 0.04% higher than that of \( k = 4 \), the testing accuracy of \( k = 8 \) is 0.24% lower than that of \( k = 4 \).

### Table 3. The transfer learning accuracy of DSC-RCNN in different \( k \).

| \( k \) | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | Average |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|---------|
| 2     | 98.21 | 99.27 | 98.65 | 100.00 | 97.29 | 100.00 | 99.90 | 99.79 | 99.14   |
| 4     | 99.69 | 99.58 | 99.27 | 100.00 | 98.85 | 99.58 | 100.00 | 99.69 | 99.58   |
| 6     | 98.15 | 99.58 | 99.69 | 99.90 | 98.75 | 100.00 | 99.90 | 99.69 | 99.46   |
| 8     | 98.96 | 98.85 | 99.79 | 100.00 | 99.48 | 99.90 | 100.00 | 100.00 | 99.62   |

![Figure 11. The process of overlapping sample.](image)

![Figure 12. The testing result of DSC-RCNN in different \( k \).](image)
And the network structure parameters of $k = 8$ is more than that of $k = 4$. As a result, the number $k$ of the depthwise separable convolution is set as 4 considering the network structure parameters and predicted results of the method.

The predicted effect of the DSC block and the DC block on the method

The DC-RCNN model is based on the DC block and the RCNN model. After training process, the testing results of the DC-RCNN model in different $k$ are compared with the DSC-RCNN model in Figure 14(a) that the testing accuracies of the DC-RCNN are decreasing with $k$. What’s more, the testing accuracies of the DC-RCNN in different $k$ are lower than that of the DSC-RCNN.

The AUC of DC-RCNN and DSC-RCNN in different $k$ is presented in Figure 14(b). The values of AUC in the DC-RCNN are all higher than 0.98 in different $k$ but lower than that of DSC-RCNN. It proves that the DC-RCNN model is not good as the DSC-RCNN model in the identification learning process.

Next, to further compare the transfer learning ability in the DC-RCNN and the DSC-RCNN, the predicted accuracy of DC-RCNN in transfer learning is illustrated in Table 4. When $k$ is 2 the testing accuracy of the DC-RCNN is the highest in those working conditions and all higher than 95%. At the same time, in Table 5, the transfer learning AUC of DC-RCNN in $k = 2$ is higher than that of other $k$ in all working conditions. However, by contrasting Table 3 with Table 4, the predicted accuracy of the DC-RCNN in $k = 2$ is less than the DSC-RCNN. And the AUC of the DC-RCNN is also inferior to that of the DSC-RCNN by comparing Figure 13 with Table 5. Therefore, the transfer learning ability of DSC-RCNN is superior to that of the DC-RCNN.

In a word, the DSC-RCNN has better recognition performance and transfer learning performance compared with the DC-RCNN.

Diagnosis results of different methods

In order to further evaluate the performance of the proposed method, different models are implemented for comparisons.

SVM. In the SVM, the kernel function is the Gaussian Radial Function kernel. Kernel parameter $g$ and penalty parameter $c$ are set as 0.5 and 1.

![Figure 13. The transfer learning AUC of DSC-RCNN.](image)

![Figure 14. The testing results of the DC-RCNN and the DSC-RCNN in different $k$: (a) the testing accuracy and (b) the testing AUC.](image)
BSA-SVM. In the BSA-SVM, the same kernel function as the SVM model is chosen. The range of kernel parameter $g$ and penalty parameter $c$ is set as $[0, 100]$, population number and iteration number are 30 and 50 in BSA-SVM.

AEs. The AEs model is employed to extract features and the output of the encoder block is input into a linear classifier. The output channels of the classifier layer are the numbers of the discerning classes.

LSSVM. In the LSSVM, the RBF kernel function is chosen. The regularization parameter and kernel parameter are 10 and 0.3.

LSTM. In the LSTM, the input feature is 32 and the hidden feature is 100. The LSTM can extract hidden feature about time from a time domain signal.

VGG-13. The VGG-13 is chosen mainly due to taking into account the influence of the input size and the model depth. Because of the difference of the input size and the identification variety, the linear classifier of the VGG-13 is replaced by those three FC layers that the weight of those is $512 \times 256, 256 \times 256, \text{and } 256 \times 2$.

The predicted results of different methods are shown in Table 6. Table 6 indicates testing accuracy of the proposed method is the highest. Although the running time of the SVM is the least, the identification result is terrible. The predicted results of the AEs, the VGG-13, and the LSTM are also bad. The BSA algorithm can improve the recognition ability of SVM. Nonetheless, the optimization will spend so much time on parameters selection of SVM. And the BSA-SVM optimization process implies that the classification ability of SVM has tremendous relevance to parameters of the kernel function. This trouble will exist all the time in the SVM model. The LSSVM can enhance the predicted accuracy. Whereas it costs more time than the proposed method and the accuracy of the LSSVM is also lower than that of DSC-RCNN in $k = 4$. As a result, the recognition competence of the proposed method is the best in those available algorithms.

Furthermore, the comparison of the transfer learning in above models is shown in Tables 7 and 8. Table 7 shows us that the DSC-RCNN in the different working conditions also has better transfer learning competence. The predicted average accuracy of the DSC-RCNN in the different working conditions is 99.62% and the highest compared with other methods. Although the LSSVM also has good identification competence, the predicted accuracies are lower than that of DSC-RCNN. In Table 8 the values of the AUC in DSC-RCNN and LSSVM are 1. It implies that the DSC-RCNN and LSSVM have good generalization abilities.

As a summary, compared with SVM, BSA-SVM, LSTM, AEs, LSSVM, and VGG-13, the developed

### Table 4. The transfer learning accuracy of DC-RCNN in different $k$.

| $k$  | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | Average |
|------|-----|-----|-----|-----|-----|-----|-----|-----|---------|
| 2    | 96.15 | 97.97 | 98.54 | 99.58 | 95.21 | 99.69 | 99.90 | 99.58 | 98.33   |
| 4    | 95.94 | 96.88 | 93.96 | 98.44 | 89.06 | 98.54 | 99.79 | 95.94 | 96.07   |
| 6    | 90.00 | 86.88 | 80.63 | 96.35 | 83.02 | 88.96 | 73.33 | 85.94 | 85.64   |
| 8    | 83.23 | 98.44 | 97.60 | 100.00 | 94.79 | 93.54 | 97.92 | 87.81 | 94.17   |

### Table 5. The transfer learning AUC of DC-RCNN in different $k$.

| $k$  | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | Average |
|------|-----|-----|-----|-----|-----|-----|-----|-----|---------|
| 2    | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00    |
| 4    | 0.99 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 0.96 | 0.97 | 0.96    |
| 6    | 0.97 | 0.99 | 0.88 | 1.00 | 0.99 | 0.96 | 0.95 | 0.97 | 0.96    |
| 8    | 0.95 | 1.00 | 0.99 | 1.00 | 1.00 | 0.99 | 1.00 | 0.99 | 0.99    |

### Table 6. Performance comparison of some methods and the proposed method.

|          | Time (s) | Accuracy (%) | AUC  |
|----------|----------|--------------|------|
| AEs      | 156.93   | 50.00        | 0.12 |
| SVM      | 152.91   | 54.69        | 0.62 |
| LSSVM    | $1.12 \times 10^4$ | 99.74 | 1.00 |
| BSA-SVM  | $3.25 \times 10^4$ | 99.22 | 1.00 |
| LSTM     | $2.65 \times 10^4$ | 54.56 | 0.57 |
| VGG-13   | 565.75   | 50.00        | 0.50 |
| DSC-RCNN | 746.97   | 100.00       | 1.00 |
DSC-RCNN achieves comprehensive performance and is able to identify the lubrication condition of the PRSM.

**The comparison of the different methods in gear fault data**

To further verify the property of the DSC-RCNN in other dataset, the fault data set of 3 MW wind turbine pinion gear by Eric Bechhoefer is utilized. This data set includes 11 fault data files and 13 good condition data files. The fault data is obtained from a fault 3 MW wind turbine pinion gear. Those good condition data are given from pinion gears of different wind turbines of the same model. Those data have two channels and is a record of the radial vibration accelerometer and the tachometer. In this training process, only the vibration data is used. The sample rate of the vibration data is 97656 Hz. The record length is 6s. The testing results comparison of the above methods is displays in Figure 15.

The predicted accuracy of the LSSVM, the LSTM, the VGG-13 and the proposed method all is 100%. And the values of AUC in those methods are 1. It indicates that the proposed method still has good identification competence.

**Visualization of the network learning process**

A convolutional neural network is similar to a black box, so the internal process cannot be reasonably explained. To clearly understand the internal learning process of the network, the T-SNE method is used to reduce the dimension of the middle output result and displays the middle result in the lower dimension. The visualization is depicted in Figure 16.

Figure 16 shows the visualization of the output result in validating process of the DSC-RCNN. The data with grease and without grease is chaotic and not identified before the training of the network. But after 40, 70, 100 iterations, it is obvious that the validating set can be easily divided. Consequently, the visualization directly confirmed the recognition capability of the proposed network.

---

**Table 7.** The transfer learning accuracy of the DSC-RCNN and those compared models in different working conditions.

|       | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | Average |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|---------|
| AEes  | 50.00 | 50.00 | 50.00 | 50.00 | 50.00 | 50.00 | 50.00 | 50.00 | 50.00   |
| SVM   | 50.00 | 50.00 | 50.00 | 50.00 | 50.00 | 50.00 | 50.00 | 85.31 | 54.41   |
| LSSVM | 97.40 | 99.69 | 99.79 | 100.00 | 94.79 | 99.39 | 100.00 | 99.58 | 98.83   |
| BSA-SVM | 45.00 | 100.00 | 99.06 | 99.79 | 92.19 | 98.33 | 100.00 | 98.44 | 91.60   |
| LSTM  | 50.00 | 53.65 | 74.17 | 52.40 | 50.00 | 50.00 | 57.60 | 50.73 | 54.82   |
| VGG-13 | 50.00 | 50.00 | 50.00 | 50.00 | 50.00 | 50.00 | 50.00 | 50.00 | 50.00   |
| DSC-RCNN | 98.96 | 98.85 | 99.79 | 100.00 | 99.48 | 99.90 | 100.00 | 100.00 | 99.62   |

**Table 8.** The transfer learning ROC of the DSC-RCNN and those compared models in different working conditions.

|       | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | Average |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|---------|
| AEes  | 0.18 | 0.05 | 0.07 | 0.01 | 0.11 | 0.08 | 0.05 | 0.06 | 0.08    |
| SVM   | 0.46 | 0.59 | 0.53 | 0.73 | 0.48 | 0.47 | 0.53 | 0.04 | 0.48    |
| LSSVM | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00    |
| BSA-SVM | 0.51 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00    |
| LSTM  | 0.57 | 0.63 | 0.76 | 0.75 | 0.41 | 0.51 | 0.63 | 0.95 | 0.65    |
| VGG-13 | 0.50 | 0.5  | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50    |
| DSC-RCNN | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00    |

**Figure 15.** The testing results of the DSC-RCNN and those compared models in the gear fault data set.
Conclusions

In this paper, an effective lubrication condition identification method, called DSC-RCNN, for the PRSM is proposed. In order to verify the diagnosis performance of the method, vibration acceleration data of PRSM with and without grease is collected from the PRSM failure test bench and preprocessed by data augmentation. Three analysis of the DSC-RCNN are conduct. The conclusions are present as follows.

1. The effect of the number of the depthwise separable convolution $k$ in the DSC on the predicted accuracy is discussed, when $k$ is 2, 4, 6, and 8. The results show the optimal number of the depthwise separable convolution is 4 in the DSC-RCNN.

2. The learning ability of the RCNN based on DC blocks is analyzed. It shows that the maximum difference of transfer learning accuracy between DSC-RCNN and DC-RCNN is 13.82%.

3. The effectiveness of the proposed PRSM lubrication condition identification method is confirmed by comparison with AEs, SVM, LSSVM, BSA-SVM, LSTM, and VGG-13. The recognition accuracies of DSC-RCNN are 50.00%, 45.31%, 0.26%, 0.78%, 45.44%, 50.00% higher than that of the above models, respectively. And the transfer learning accuracies of DSC-RCNN are also 49.62%, 45.23%, 0.79%, 8.02%, 44.80%, 49.62% higher than those.

As a result, the DSC-RCNN for lubrication condition identification of PRSM has robust recognition ability and fine facticity.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was supported by National Natural Science Foundation of China (Grant No. 51875458), Key Research and Development Program of Shaanxi (Program No.2021ZDLGY10-08), and Natural Science Basic Research Plan in Shaanxi Province of China (Grant No.2020JQ-178).

ORCID iDs
Wei Cai https://orcid.org/0000-0001-6590-2388
Shangjun Ma https://orcid.org/0000-0001-8177-2907

References
1. Brandenburg G, Bruckl S, Dormann J, et al. Comparative investigation of rotary and linear motor feed drive systems for high precision machine tools. In: 6th international workshop on advanced motion control. Proceedings, Nagoya, Japan, 2000, pp. 384-389.
2. Zhang W, Liu G, Tong R, et al. Load distribution of planetary roller screw mechanism and its improvement approach. Proc IMechE, Part C: J Mechanical Engineering Science 2016; 230: 3304–3318.
3. Fu X, Liu G, Ma S, et al. A comprehensive contact analysis of planetary roller screw mechanism. J Mech Des 2017; 139: 012302.
4. Andrade A, Nicolosi D, Lucchi J, et al. Auxiliary total artificial heart: a compact electromechanical artificial heart working simultaneously with the natural heart. Arif Organs 1999; 23: 876–880.
5. Abei F, Daidie A, Chaussumier M, et al. Static load distribution and axial stiffness in a planetary roller screw mechanism. J Mech Des 2016; 138: 012301.
6. Rys J and Lisowski F. The computational model of the load distribution between elements in a planetary roller screw. Appl Mech Mater 2014; 52: 699–705.
7. Du X, Chen B and Zheng Z. Investigation on mechanical behavior of planetary roller screw mechanism with the effects of external loads and machining errors. Tribol Int 2021; 154: 106689.
8. Jones MH and Velinsky SA. Contact kinematics in the roller screw mechanism. J Mech Des 2013; 135: 051003.
9. Qiao G, Liu G, Ma S, et al. Thermal characteristics analysis and experimental study of the planetary roller screw mechanism. Appl Therm Eng 2019; 149: 1345–1358.
10. Du C, Liu G, Qiao G, et al. Transient thermal analysis of standard planetary roller screw mechanism based on finite element method. Adv Mech Eng 2018; 10: 1–10.
11. Velinsky SA, Chu B and Lusky TA. Kinematics and efficiency analysis of the planetary roller screw mechanism. J Mech Des 2009; 131: 011016.
12. Fu X, Liu G, Ma S, et al. An efficient method for the dynamic analysis of planetary roller screw mechanism. Mech Mach Theory 2020; 150: 103851.
13. Niu M, Ma S, Cai W, et al. Fault diagnosis identification of planetary roller screw mechanism based on bird swarm algorithm and support vector machine. J Phys Conf Ser 2020; 1519: 012007.
14. Ji W, Cheng C, Xie G, et al. An intelligent fault diagnosis method based on curve segmentation and SVM for rail transit turnout. J Intell Fuzzy Syst 2021; 41: 4275–4285.
15. Ezz-Eldin M, Khalaf A, Hamed H, et al. Efficient feature-aware hybrid model of deep learning architectures for speech emotion recognition. IEEE Access 2021; 9: 19999–20011.
16. Jin Y, Qin C, Huang Y, et al. Actual bearing compound fault diagnosis based on active learning and decoupling attentional residual network. Measurement 2021; 173: 108500.
17. Polic M, Krajacic I, Lepora N, et al. Convolutional auto-encoder for feature extraction in tactile sensing. IEEE Robot Autom Lett 2019; 4: 3671–3678.
18. Hu K, Jiang M, Zhang H, et al. Design of fault diagnosis algorithm for electric fan based on LSSVM and Kd-Tree. Appl Intell 2021; 51: 804–818.
19. Zhou H, Ying Y, Li J, et al. Long-short term memory and gas path analysis based gas turbine fault diagnosis and prognosis. Adv Mech Eng 2021; 13: 1–12.
20. Simonyan K and Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2014.
21. Chen YP, Dai XY, Liu MC, et al. Dynamic convolution: attention over convolution kernels. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR). Seattle, WA, USA, 2020, pp.11027–11036.
22. Sun W, Yao B, Zeng N, et al. An intelligent gear fault diagnosis methodology using a complex wavelet enhanced convolutional neural network. Materials 2017; 10: 1–18.
23. Liu L, Shen C and Hengel AVD. Cross-convolutional layer pooling for image recognition. IEEE Trans Pattern Anal Mach Intell 2017; 39: 2305–2313.
24. Nakahara H, Fuji T and Sato S. A fully connected layer elimination for a binarize convolutional neural network on an FPGA. In: 2017 27th international conference on field programmable logic and applications, Ghent, Belgium, 2017, pp.1–4.
25. Zhao M, Kang M, Tang B, et al. Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes. IEEE Trans Ind Electron 2018; 65: 4290–4300.
26. Chen K, Zhou XC, Fang JQ, et al. Fault feature extraction and diagnosis of gearbox based on EEMD and deep briefs network. Int J Rotating Machinery 2017; 2017: 1–10.
27. Jia F, Lei Y, Lin J, et al. Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech Syst Signal Process 2016; 72–73: 303–315.
28. He W, Guo B, Chen B, et al. A data-driven group-sparse feature extraction method for fault detection of wind turbine transmission system. Meas Sci Technol 2020; 31: 074008.
29. Maaten L and Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008; 9: 2579–2625.