China Instrument and Control Society
Instrumentation
OPEN ACCESS
INSTRUMENTATION, Vol. 12, No. 1, March 2025 https://doi.org/10.15878/j.instr.202500217

Article

Unsupervised Domain Adaptive Migration Learning-Based Approach to Bearing Remaining Useful Life Prediction

Haitao Wang, Ruihua Wang, Jie Yang

School of Mechanical and Electrical Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China

* Corresponding author email: wangruihua@xauat.edu.cn.

Abstract: Accurate predictions of the Remaining useful life (RUL) of mechanical equipment are vital for lowering maintenance costs and maintaining equipment reliability and safety. Data-driven RUL prediction methods have made significant progress, but they often assume that the training and testing data have the same distribution, which is often not the case in practical engineering applications. To address this issue, this paper proposes a residual useful life prediction model that combines deep learning and transfer learning. In this model, called transfer convolutional attention mechanism for early-life stage time convolutional network (TCAM-EASTCN), an unsupervised domain adaptation strategy is introduced based on the characterization of subspace distances and orthogonal basis mismatch penalties in the convolutional attention mechanism for early-life stage time convolutional network (CAM-EASTCN). This approach minimizes the distribution differences between different domains, enhancing the learning of cross-domain invariant features and effectively reducing the distribution gap between the source and target domains, thereby improving the accuracy of RUL prediction under varying conditions. Experimental results demonstrate that TCAM-EASTCN outperforms other models in terms of RUL prediction accuracy and generalization.

Keywords: Deep learning; Temporal convolutional network; Representation subspace distance; Orthogonal basis mismatch penalty

Citation: Haitao Wang, Ruihua Wang, Jie Yang."Unsupervised Domain Adaptive Migration Learning-Based Approach to Bearing Remaining Useful Life Prediction." Instrumentation 12, no.1 (March 2025). https://doi.org/10.15878/j.instr.202500217

1 Introduction

With the rapid development of computational methods and information technology[1], the complexity of modern production systems is increasing. Prognostics and Health Management (PHM) have become effective methods for improving work availability[2], enhancing the maintainability, supportability, reliability, and safety of modern industrial equipment, and reducing lifecycle costs. Consequently, they have garnered widespread attention from both academia and industry professionals. Within the framework of Fault Detection and Prognostics Health Management, multiple processes are involved, with Remaining Useful Life (RUL) prediction technology being one of the key components. RUL prediction serves as the foundation for decision-making in management activities. By forecasting the remaining operational time of systems or components and providing early warnings of impending failures, RUL prediction can significantly mitigate the occurrence of accidents. Therefore, constructing accurate and reliable RUL prediction models holds paramount importance in the industrial domain and is a focal point of research for experts both domestically and internationally.

Over the past decade, RUL prediction technology has seen significant development, primarily categorized into three main approaches[3,4]: model-based methods, data-driven methods, and hybrid approaches combining both mechanistic models and data-driven techniques. Model-based approaches typically require personnel with extensive domain knowledge to utilize dynamic mathematical models for system state estimation or prediction. However, due to the need for real-time updates of model parameters and data information, this method may struggle to accurately model highly complex and dynamic systems[5,6]. In contrast, methods for predicting RUL based on data alone don't require pre-existing expert insights and can automatically uncover underlying causal relationships in the data. Furthermore, with the availability of large datasets and reduced demand for specialized expertise, data-driven prediction methods hold promise for handling extensive monitoring data and providing more accurate RUL predictions. These approaches mainly include conventional machine learning methods as well as techniques based on deep learning.

Standard machine learning models for RUL prediction generally consist of three phases: firstly, data is collected from various sensors with different functionalities. These sensors measure parameters such as vibration, temperature, and pressure, which are integrated into the equipment; secondly, feature extraction from the collected data, choosing the characteristics that most accurately represent the equipment's degradation pattern for future modeling; finally, using the extracted features as inputs to estimate the RUL of the equipment. At present, machine learning techniques, including Support Vector Machines (SVM) [7], Random Forests[8], and Artificial Neural Networks(ANN) [9], have achieved certain success in estimating the remaining Useful Life of Mechanical apparatus. As deep learning technology continues to advance and expand, it has gradually replaced the position of traditional machine learning methods in many fields. Deep learning constructs deep neural network models that can learn deep feature representations from data, thereby achieving precise handling of complex tasks. Deep learning, a specialized subset of machine learning, comprises a network of intricately interconnected nonlinear processing units. Its powerful nonlinear mapping and feature extraction capabilities have made deep learning highly regarded in the domains of Remaining Useful Life prediction and health monitoring. Recurrent Neural Networks (RNNs) and their derivative variants have found extensive applications owing to their distinctive sequential processing abilities. For example, Wang[10] proposed a framework for predicting mechanical failures utilizing a Recursive Convolutional Neural Network (RCNN). This framework utilizes recursive convolutional layers to learn time dependencies in degradation states. It integrates variable reasoning to quantify uncertainty in RCNN life prediction. Liu[11] introduced a feature-attention mechanism in their study. This mechanism dynamically modifies the weights of various features in the input data. It aims to enhance the extraction of long-term time series information. They input the features into bidirectional gated recurrent units to extract long-term time series information and combine convolutional neural networks to capture local features. Luo[12] employed Bidirectional Long Short-Term Memory (BiLSTM) networks, demonstrating their efficacy in accurately predicting the deterioration patterns of rolling bearings. However, RNNs often suffer from the problems of vanishing gradients and exploding gradients, greatly affecting the accuracy of their training. As technology continues to evolve, Convolutional Neural Networks (CNNs) provide a new perspective for predicting time series data. Leveraging their advantages in parallel computation, CNNs can efficiently handle large-scale data and capture more historical information while increasing the receptive field. For example, Wang[13] utilized Deep Separable Convolutional Networks (DSCN) to prognosticate the Remaining Useful Life of bearings, establishing a direct correlation in RUL predicting devoid of antecedent data. Ge[14] introduced a method for short-term traffic speed prediction grounded on Graph Attention Convolutional Networks, yielding noteworthy predictive efficacy. Furthermore, Lin[15] devised a Trend-adaptive Fully Convolutional Network (TaFCN) to enhance prediction accuracy even further.

Although deep learning-based approaches have shown promising advancements in the realm of Remaining Useful Life prediction for bearings. However, these methodologies often necessitate the assumption that the testing data (target domain) and training data (source domain) adhere to identical distributions. This assumption poses a challenge in practical engineering scenarios, as machinery frequently operates under diverse conditions, leading to notable distribution disparities in the data. Machines undergo an extended deterioration phase from regular functioning to breakdown, involving diverse modes of failure[16]. It is noteworthy that machines, even when subjected to identical operating conditions, can demonstrate notably distinct degradation trends[17]. To bolster the model's adaptability across varying operational settings, deep transfer learning[18] models have been widely applied in the fault diagnosis domain. Introducing transfer learning into the RUL regression domain is a challenging task that is still in the exploratory stage. To address the challenge of cross-condition RUL prediction, Pan[19] introduced a two-stage approach for predicting the RUL of rolling bearings based on the Extreme Learning Machine. This method segments the operational stages of the bearings into two states to predict RUL, albeit focusing solely on short-term predictive objectives for a single operational condition. Lv[20] introduced a sophisticated deep subdomain adaptive regression network designed to predict the RUL of bearings amidst changing operational conditions, successfully demonstrating cross-condition RUL predicting. However, it retains a lower degree of preservation for local spatial features, resulting in the loss of significant information. E Tzeng[21] introduced an adversarial discriminative domain adaptation model to mitigate distribution differences between the training and testing domains, thereby enhancing generalization performance. However, this method also loses relevant information, leading to a decrease in prediction performance. Sun[22] presented a Deep Domain Adaptation (DDA) technique for RUL prediction challenges, employing Long Short-Term Memory (LSTM) for feature extraction and leveraging a reverse gradient approach to alleviate domain shift issues. Li[23] developed a progressive domain alignment strategy to cater to two distinct domains, utilizing a shared codebook to align feature distinctions and progressively diminish inter-domain variations. The majority of these approaches endeavor to ascertain domain-invariant characteristics shared across the source and target domains. However, imposing constraints on the similarity between target and source features without limitations may result in the exclusion of specific valuable information in the target domain For instance, the mutual information quantifies the relationship between the target dataset and the features extracted. This limitation could impede the domain adaptation efficacy in RUL prediction tasks.

To address the disparity in distribution between the original and intended domains and to guarantee feature scale consistency, this study introduces a model that fuses deep learning and transfer learning methodologies. The proposed model combines an efficient adaptive shrinkage model with a convolutional attention network[24] and incorporates an unsupervised domain adaptation transfer learning approach involving Representation Subspace Distance (RSD) and Basis Mismatch Penalization (BMP). This integration aims to bolster the predictive capabilities of the model specifically for the target domain. The main research contributions are as follows:

(1) The introduction of an unsupervised domain adaptation transfer learning strategy rooted in RSD and BMP.

(2) The resolution of bearing life prediction challenges across diverse operating conditions, coupled with the reduction of distribution disparities between the target and source domains enhances the accuracy of RUL prediction.

2 RSD and BMP

2.1 RSD

Representation subspace distance[25] is a method for measuring the differences in distributions between different datasets in a specific feature space. The fundamental idea is to quantify their dissimilarity by comparing the geometric distance of two subspaces on the Grassmann manifold.

In high-dimensional spaces, datasets are typically represented by a set of feature vectors, which form a subspace. The Grassmann manifold comprises all k-dimensional subspaces, providing a natural framework for comparing and analyzing these subspaces. By computing the principal angles between two subspaces (essentially the "angles" between them), we can obtain a measure of their dissimilarity. Smaller principal angles indicate greater similarity between two subspaces, while larger angles suggest greater dissimilarity.

RSD is commonly calculated based on these principal angles between subspaces. Principal angles represent the "angles" between two subspaces, analogous to the angles between two vectors. Two k-dimensional subspaces principal angles can be found by solving an optimization problem that involves maximizing the inner products between vectors from the two subspaces.

Assuming we have two subspaces, each represented by an orthogonal basis matrices ($U \in \mathbb{R}^{n\times k}$) and ($V \in \mathbb{R}^{n\times k}$) ,where (n) is the dimensionality of the feature space and (k) is the dimensionality of the subspace. The column vectors of matrices (U) and (V) are unit vectors and orthogonal to each other.

The principal angles (θ_i) can be found as follows:

$$\cos(\theta_i) = \max_{\mathbf{u}\in U, \mathbf{v}\in V} \frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{u}||\mathbf{v}|} \tag{1}$$

where $u$and$v$ respresent the column vectors of matrices (U) and (V).

Once we have found all the principal angles ($\theta_1, \theta_2, \ldots, \theta_k$) , RSD can be computed using some function of these principal angles.

$$RSD(U, V) = \sqrt{\sum_{i=1}^{k} \sin^2(\theta_i)} \tag{2}$$

In other words, RSD provides a quantitative measure of the difference between two subspaces. A lower RSD value indicates more similarity between the two subspaces, whereas a higher RSD value suggests more dissimilarity.

RSD is a measurement method based on principal angles. It can be determined precisely by adding the sine values of each principal angle, or by taking the square root of the total sum of the squares of these sine values. Since the sine function monotonically increases between 0 and π/2, RSD effectively reflects the dissimilarity between subspaces. When two subspaces are completely identical, meaning all principal angles are 0, RSD is also 0. Conversely, as the dissimilarity between two subspaces increases, the value of RSD also increases.

2.2 BMP

The orthogonal basis mismatch penalty[26] is a regularization technique aimed at mitigating the discrepancy between feature spaces of different domains in the realm of transfer learning and domain adaptation. By introducing a penalty term that quantifies the mismatch between the bases of the source and target domains, this technique seeks to minimize the difference during the optimization process. The essence of the orthogonal basis mismatch penalty lies in finding a new basis matrix (W) that aligns the feature representations of the source(U) and target domains (V)as closely as possible. The orthogonal basis mismatch penalty is typically formulated as follows:

$$Peanlty(U,V,W)=|U-W|_F^2+|V-W|_F^2 \quad (3)$$

In this context, $(|\cdot|_F)$ denotes the Frobenius norm, utilized to quantify the difference between matrices. This penalty term encourages (W) to simultaneously approximate (U) and (V). By integrating the orthogonal basis mismatch penalty into the objective function of transfer learning or domain adaptation tasks, we aim to optimize task performance while reducing the discrepancy between feature spaces of different domains.

3 TCAM-EASTCN Model

3.1 Unsupervised Domain Adaptive Transfer Learning Strategy

This section aims to enhance the predictive accuracy of the model on the target domain through the integration of an unsupervised domain adaptive transfer learning approach within the Channel Attention Mechanism-Enhanced Adaptive Shrinkage Temporal Convolutional Network(CAM-EASTCN). This strategy involves transferring labeled knowledge from the source domain to unlabeled data in the target domain. Denoting the source domain information as $D_s=\{x_s^i,y_s^i\}_{i=1}^{n_s}$ and the target domain information as $D_t=\left\{x_t^j\right\}_{i-1}^{n_t}$, with x_s^i and x_t^j representing the vibration data from the respective domains, and y_s^i denoting the remaining useful life label associated with the source domain. n_s and n_t denote the sample sizes in the source and target domain data, respectively. CAM-EASTCN is utilized to extract profound degradation features in the source and target domains denoted as $F_s=\left[f_s^1,\cdots,f_s^b\right]$ and $F_t=\left[f_t^1,\cdots,f_t^b\right]$, respectively, with b indicating the batch size. Each feature matrix is regarded as a point within the Grassmann space utilizing a base consisting of unit vectors. Through basis matching, the discrepancy between subspaces is minimized without altering the feature scale. Consequently, the transition from the source domain to the target domain is accomplished by aligning orthogonal bases while maintaining the feature scale.

This study applies the singular value decomposition technique to decompose F into F_s and F_t orthogonal bases and singular values, facilitating the extraction of standard representations.

$$F_s=U_s\Sigma_s(V_s)^T,F_t=U_t\Sigma_t(V_t)^T(4)$$

The orthogonal bases in the matrix, along with the representation subspaces of the source domain, the representation subspaces of the target domain, are constituted by orthogonal bases $U_s=\left[u_s^1,u_s^2,\cdots,u_s^b\right]$ and $U_t=\left[u_t^1,u_t^2,\cdots,u_t^b\right]$. In Grassmann space, principal angles are commonly used to measure the difference between U_s and U_t The definition of principal angles is :

$$\theta_i=min\arccos\left(\frac{(u_s^i)^T u_t^i}{\|u_s^i\|\cdot\|u_t^i\|}\right) \quad (5)$$

In this instance, When $\Theta=[\theta_1,\cdots,\theta_b]$ equals zero, it indicates that the subspaces formed by the two sets of orthogonal bases are identical, implying that the source domain F_s and target domain F_t share similar distributional characteristics.

In this study, RSD is defined as the geometric distance based on principal angles. The computation involves summing the sine values of all principal angles. Metrics of variation in subspace distribution are as follows:

$$dis_{RSD}(U_s,U_t)=\|\sin\Theta\|_1=\sum_{i=1}^{b}\sin\theta_i \quad (6)$$

$$(U_s)^T U_t=P_s(diag(\cos\Theta))(P_t)^T \quad (7)$$

With $\cos\Theta$ denoting the cosine value of the principal angle and $\|\cdot\|_1$ representing the Frobenius norm of matrix 1.

To uphold the geometric characteristics of deep features within the spatial domain, it is plausible to introduce a penalty for any discrepancies in the bases, thereby guaranteeing the congruence of orthogonal bases within the feature subspaces. Various orthogonal bases exhibit distinct degenerate characteristics, and in various domains, similar degenerate features often correspond to similarly arranged orthogonal bases. Match orthogonal bases of equal importance from two subspaces together, the calculation formula is:

$$reg_{BMP}(U_s,U_t)=\||P_s|-|P_t|\|_F^2 \quad (8)$$

The Frobenius norm of the matrix $\|\cdot\|_F$, as well as the matrices P_s and P_t calculated through equations (8), play a crucial role in this process.

3.2 Training Optimization Objective

The TCAM-EASTCN model's structural layout is presented in Figure 1.

The objective function consists of three main components: the basic RUL regression loss function for source domain data, the RSD loss function comparing deep features between the source and target domains, and the BMP loss function. Both the feature extractor and Remaining Useful Life predictor undergo training through the minimization of the RUL regression loss function using the source domain dataset. The equations for the calculations are detailed below:

$$Loss_{RUL}=\frac{1}{b}\sum_{i=1}^{b}(\hat{y}_i-y_i)^2 \quad (9)$$

Here, $\hat{y}$ and y_i represent the predicted and actual remaining useful life, respectively. b signifies the batch size of samples from the source domain data. After extracting a new set of samples b from the target domain data, the RSD function and the BMP function are determined across the feature subspaces of both the source and target domains.

$$L_{RSD}(F_s,F_t)=\|\sin\Theta\|_1 \quad (10)$$

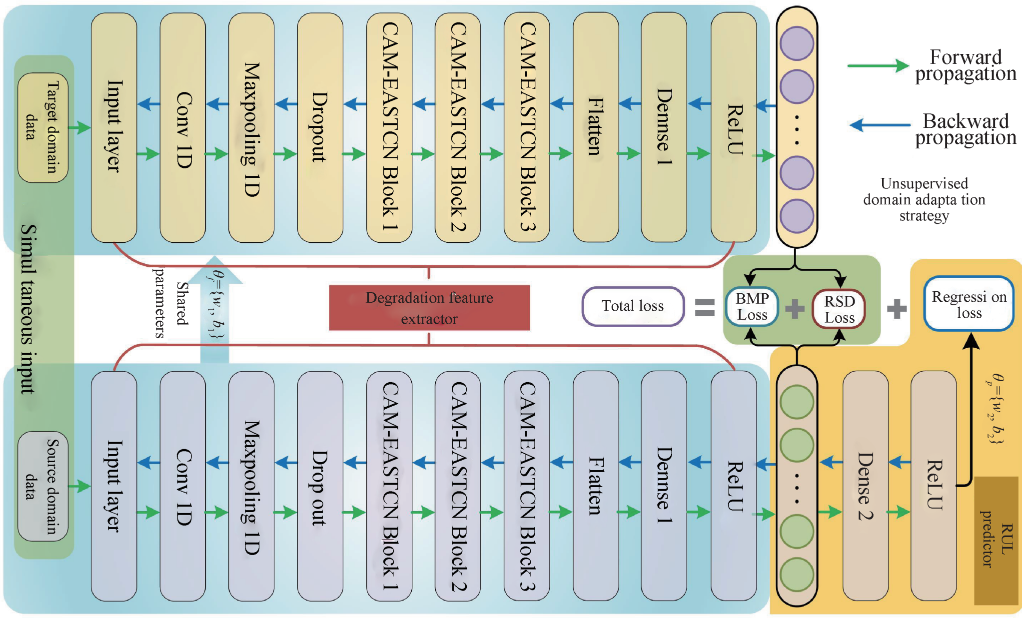

Fig.1 Model architecture diagram of TCAM-EASTCN

$$L_{BMP}(F_s,F_t)=\ \| \ |P_s|-|P_t| \ \|_F^2 \tag{11}$$

This loss function facilitates the learning of transferable features, enhancing domain-adaptive regression. The optimization function is formulated by combining equations (10) through (11).

$$L_{total}=L_{RUL}+\alpha L_{RSD}+\beta L_{BMP} \tag{12}$$

where α and β are penalty coefficients used to balance the L_{RSD} and L_{BMP} terms . Minimizing the feature space misalignment between distinct domains, following the determination of the ultimate optimization loss function in TCAM-EASTCN, facilitates the acquisition of cross-domain invariant features by the model, thereby bolstering domain-adaptive regression capabilities. Equation (12) is adjusted as follows:

$$L_{total}\left(\theta_f,\theta_p\right)=minL_{RUL}(\theta_f,\theta_p)+\alpha L_{RSD}(\theta_f)+\beta L_{BMP}(\theta_f) \tag{13}$$

Here, $\theta_f=\{W_1,b_1\}$ represents the parameters of the degradation feature extractor, and $\theta_p=\{W_2,b_2\}$ denotes the parameters of the RUL prediction variable. The model parameters are iteratively adjusted utilizing the Adam optimization algorithm until the loss function reaches the specified convergence criteria. The algorithmic updating procedure is delineated as follows:

$$\theta_f\leftarrow\theta_f-\eta\left(\frac{\partial L_{oss_{RUL}}}{\partial\theta_f}+\alpha\frac{\partial L_{RSD}}{\partial\theta_f}+\beta\frac{\partial L_{BMP}}{\partial\theta_f}\right) \tag{14}$$

$$\theta_p\leftarrow\theta_p-\eta\frac{\partial L_{oss_{RUL}}}{\theta_p} \tag{15}$$

In this context, η stands for the learning rate. The methodology for training the TCAM-EASTCN model is illustrated in Fig. 2. Post-training with TCAM-EASTCN, the model's capability to acquire cross-domain invariant features is bolstered. Therefore, improving the ability to predict RUL of bearings under various operational circumstances.

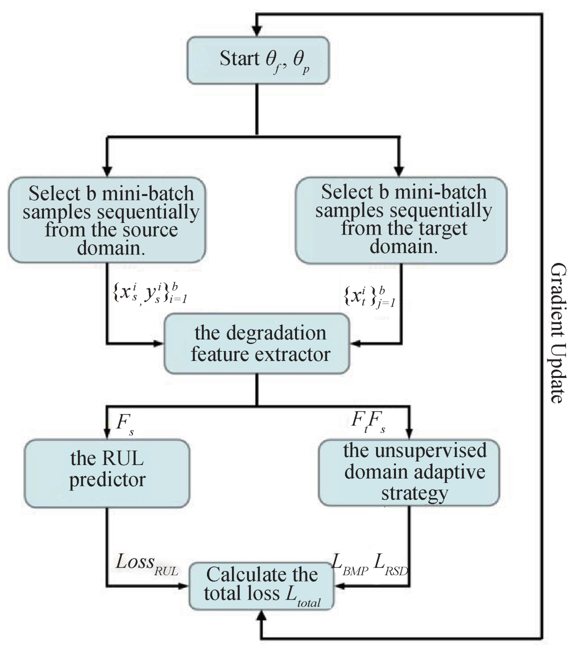

Fig.2 The exhaustive training process of the TCAM-EASTCN

4 Rolling Bearing RUL Prediction based on TCAM-EASTCN

4.1 Method for RUL Prediction

The structure for predicting the RUL of rolling bearings utilizing the devised approach primarily comprises two key components: offline training and online RUL estimation, as delineated in Fig. 3. Data concerning the complete life cycle of bearings exposed to two distinct operational environments was initially obtained from mechanical apparatus. The data that has been annotated under a specific operational context is identified as the source domain data, whereas the unannotated data collected under a different operational context is recognized as the target domain data. The TCAM-EASTCN model is subsequently trained to utilize data from both the source and target domains in order to obtain trained feature extractors and Remaining Useful Life predictors. The testing data from the specified domain must be incorporated into the degradation feature extractor and the Remaining Useful Life predictor in order to realize accurate predictions for bearing RUL.

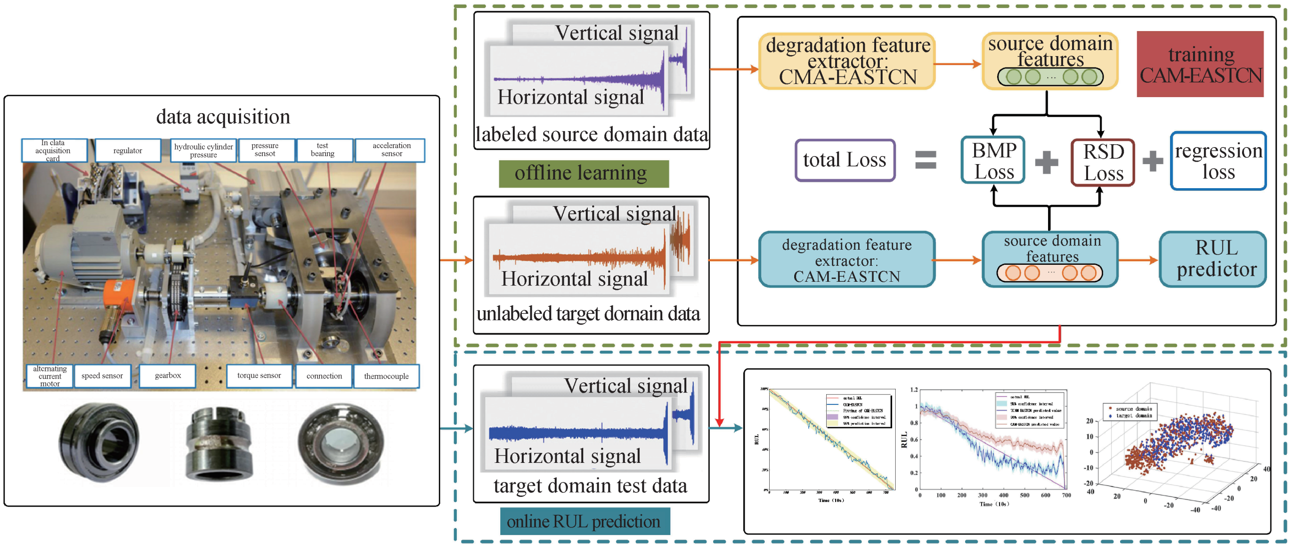

Fig.3 Comparison of MAE and RMSE for XJTU-SY Bearing Dataset

4.2 Experimental Validation

The FEMTO dataset serves as the primary reference dataset for the IEEE PHM 2012 Predictive Maintenance Challenge, accompanied by comprehensive details as outlined in citation[27]. To evaluate the domain adaptation efficacy of the novel approach under a range of varied operational circumstances, transfer datasets have been established to reflect distinct scenarios, as detailed in Table 1. The source domain comprises labeled bearing data from one operating condition, while the target domain encompasses unlabeled bearing data from a distinct operating condition.

Table 1 The task of RUL prediction for bearing migration datasets

Task	Training Dataset(source domain (labeled)→target domain (unlabeled))	Test Dataset
1	B1–1~B1–7→B2-1 and B2-2	B2-6
2	B1–1~B1–7→B3-1 and B3-2	B3-3
3	B2–1~B2–7→B1-1 and B1-2	B1-7
4	B2–1~B2–7→B3-1 and B3-2	B3-3
5	B3–1~B3–2→B1-1 and B1-2	B1-7
6	B3–1~B3–2→B2-1 and B2-2	B2-6

The test data is gathered within identical operational parameters to those of the specified target domain. Figure 4 demonstrates the comparison of RUL prediction outcomes with and without knowledge transfer, corresponding to tasks one through six (labeled as a to f). Compared to scenarios without knowledge transfer, the TCAM-EASTCN model demonstrates RUL curves that closely approximate real-world situations. This suggests a significant performance enhancement of the TCAM-EASTCN model by addressing data distribution discrepancies across different operating conditions.

Furthermore, a visual examination of the identified deterioration characteristics is carried out to demonstrate the effectiveness of transfer learning in practice. An examination is conducted in this study to analyze the degradation characteristics that have been extracted. In Task One, the delineated features obtained from the originating and destination domains are transformed into two-dimensional and three-dimensional representations through t-SNE, illustrated in Fig. 5. The comparison between Figure 5(a) and 5(b) shows that The TCAM-EASTCN model adeptly mitigates the distribution incongruity observed among the features derived from both the source and target domains, thereby validating the efficacy of the unsupervised domain adaptation approach proposed in this research.

4.3 Comparative Experiments

The efficacy of our proposed method was assessed through a comparative analysis involving three distinct approaches. Among them, two were rooted in domain adaptation: CAM-EASTCN-CMD (Central Moment Discrepancy) and CAM-EASTCN-MMD (Maximum Mean Discrepancy). The third approach, CAM-EASTCN, did not incorporate domain adaptation. Analysis of the data presented in Table 2 revealed that TCAM-EASTCN demonstrated superior performance when compared to the alternative methods. Furthermore, the results displayed in Table 3 highlighted that TCAM-EASTCN

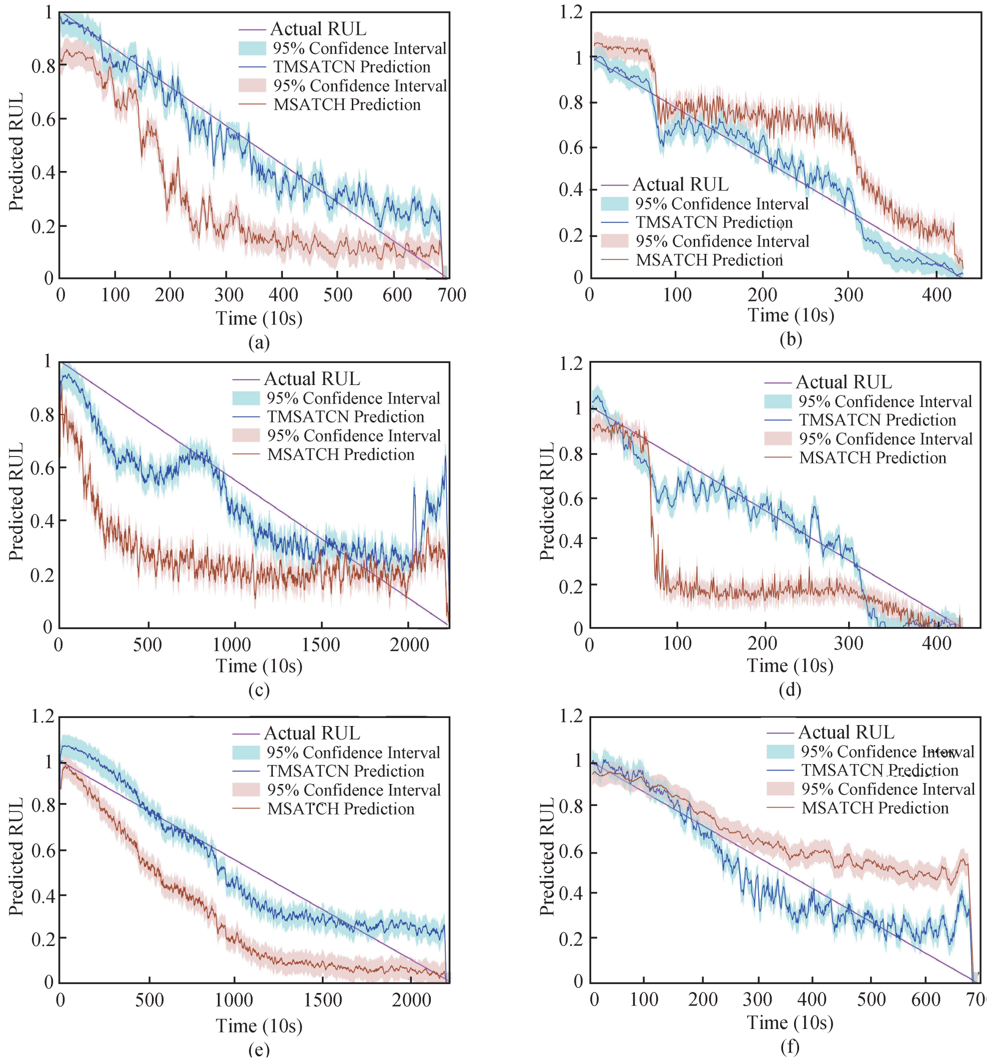

Fig.4 Comparing TCAM-EASTCN's RUL Prediction Accuracy with and without Transfer Learning on Bearing Migration Datasets

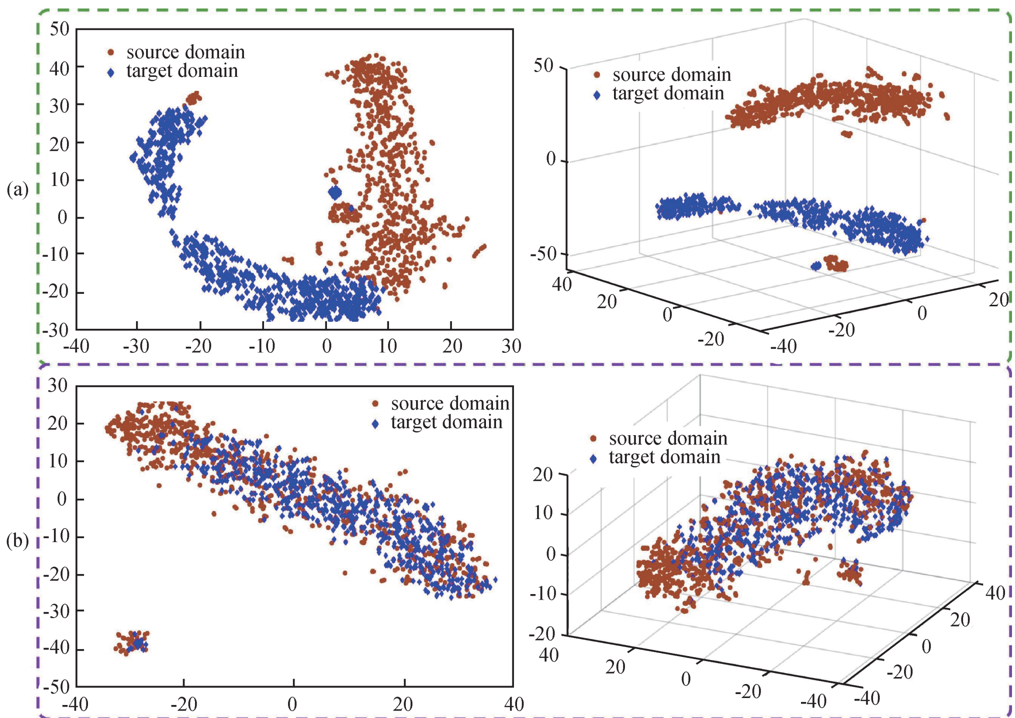

Fig.5 Analysis of visual characteristics through the utilization of t-SNE

achieved a notable decrease of 44.4% and 47.8% in MAE and RMSE, respectively, as opposed to CAM-EASTCN. Additionally, TCAM-EASTCN outperformed CAM-EASTCN-MMD and CAM-EASTCN-CMD by reducing MAE and RMSE by up to 28.6% and 29.4%, respectively. The results emphasize the efficacy of the suggested approach in facilitating the model to attain cross-domain invariant characteristics.

Table 2 Comparison of outcomes between the method put forth and similar approaches.

Test Set	CAM-EASTCN		CAM-EASTCN-CMD		CAM-EASTCN-MMD		TCAM-EASTCN	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Task1	0.16±0.01	0.20±0.01	0.11±0.01	0.14±0.01	0.11±0.01	0.15±0.01	0.09±0.01	0.12±0.01
Task2	0.12±0.01	0.17±0.01	0.08±0.01	0.10±0.01	0.08±0.01	0.09±0.01	0.07±0.01	0.09±0.01
Task3	0.20±0.01	0.25±0.01	0.13±0.01	0.16±0.01	0.14±0.01	0.16±0.01	0.10±0.01	0.13±0.01
Task4	0.19±0.01	0.24±0.01	0.14±0.01	0.16±0.01	0.15±0.01	0.20±0.01	0.09±0.01	0.12±0.01
Task5	0.29±0.01	0.34±0.01	0.22±0.01	0.26±0.01	0.23±0.01	0.27±0.01	0.13±0.01	0.15±0.01
Task6	0.12±0.01	0.15±0.01	0.13±0.01	0.16±0.01	0.13±0.01	0.17±0.01	0.10±0.01	0.12±0.01

Table 3 Comparison of Mean Results between the Proposed Method and Relevant Approaches

Different methods	Average RMSE	Discrepancy	Average MAE	Discrepancy
TCAM-EASTCN	0.12	—	0.10	—
CAM-EASTCN-MMD	0.17	↑ 0.05	0.14	↑ 0.04
CAM-EASTCN-CMD	0.16	↑ 0.04	0.14	↑ 0.04
CAM-EASTCN	0.23	↑ 0.11	0.18	↑ 0.08

A comparative analysis was undertaken against four modern methodologies for predicting RUL to bolster the credibility of the proposed approach. These methodologies encompass the transfer-gated recurrent unit introduced by Cao[28], the multi-layer perceptron transfer learning method by Zhu[29], the transferable convolutional neural network by Cheng[30], and the domain adversarial neural network by Costa[31]. Table 4 and Table 5 present the findings from the experiments. Based on the findings delineated in Table 4, the data suggests that TCAM-EASTCN displays superior performance in predicting Remaining Useful Life (RUL). Examination of Table 5 reveals that TCAM-EASTCN outperforms Cao, Zhu, Cheng, and Costa in terms of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The graphical representation provided in Figure 6 illustrates that TCAM-EASTCN consistently maintains the lowest MAE and RMSE values across all six tasks. This indicates the robust domain adaptation capabilities of TCAM-EASTCN in analyzing vibration signals from diverse operational conditions. By incorporating the convolutional attention sub-network and the adaptive shrinkage sub-network, TCAM-EASTCN streamlines regression tasks on vibration signals. The novel methodology adeptly rectifies the disparity in feature distribution observed across the source and target domains, leading to a significant improvement in the precision of prognosticating RUL. In summary, the proposed methodology demonstrates remarkable efficacy and superiority in predicting bearing RUL under varied operational contexts.

Table 4 Comparison Results of the Proposed Method against Existing Approaches

Test Set	Costa[31]		Cheng[30]		Zhu[29]		Cao[28]		TCAM-EASTCN	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
Task 1	0.14±0.01	0.16±0.01	0.31±0.01	0.33±0.01	0.18±0.01	0.20±0.01	0.15±0.01	0.17±0.01	0.09±0.01	0.12±0.01
Task 2	0.17±0.01	0.19±0.01	0.29±0.01	0.31±0.01	0.25±0.01	0.27±0.01	0.13±0.01	0.15±0.01	0.07±0.01	0.09±0.01
Task 3	0.27±0.01	0.28±0.01	0.22±0.01	0.24±0.01	0.30±0.01	0.32±0.01	0.21±0.01	0.23±0.01	0.10±0.01	0.13±0.01
Task 4	0.22±0.01	0.24±0.01	0.25±0.01	0.27±0.01	0.37±0.01	0.40±0.01	0.20±0.01	0.22±0.01	0.09±0.01	0.12±0.01
Task 5	0.73±0.01	0.75±0.01	0.58±0.01	0.60±0.01	0.77±0.01	0.79±0.01	0.62±0.01	0.64±0.01	0.13±0.01	0.15±0.01
Task 6	0.71±0.01	0.73±0.01	0.52±0.01	0.54±0.01	0.67±0.01	0.69±0.01	0.38±0.01	0.40±0.01	0.10±0.01	0.12±0.01

Table 5 Comparison findings of the Average between the novel methodology and current strategies.

Different methods	Average RMSE	Discrepancy	Average MAE	Discrepancy
TCAM-EASTCN	0.12	—	0.10	—
Cao[28]	0.30	↑0.18	0.28	↑0.18
Zhu[29]	0.45	↑0.33	0.42	↑0.32
Cheng[30]	0.38	↑0.26	0.36	↑0.26
Costa[31]	0.39	↑0.27	0.37	↑0.27

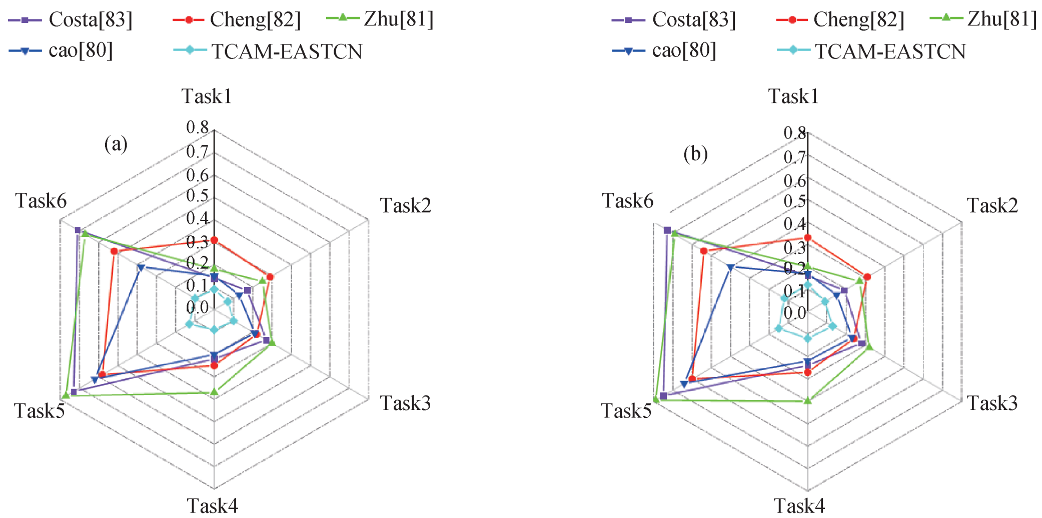

Fig.6 Diagram of a tubular reactor

5 Conclusion

This paper presents a novel methodology for Remaining Useful Life (RUL) prediction, integrating deep learning with transfer learning to address practical challenges in real-world applications. The proposed approach incorporates an unsupervised domain adaptation mechanism, leveraging Representation Subspace Distance (RSD) and Bases Mismatch Penalization (BMP) within a convolutional attention mechanism to align feature distributions across domains. This integration results in the TCAM-EASTCN model, specifically designed to overcome the challenges posed by cross-domain feature distribution disparities. A key innovation of this method lies in its unsupervised domain adaptation strategy, which minimizes discrepancies between source and target domain feature distributions. By utilizing RSD and BMP, the model enhances its ability to extract domain-invariant features, ensuring robust generalization across diverse operational scenarios. Comprehensive evaluations on benchmark datasets, including XJTU-SY and FEMTO, demonstrate that TCAM-EASTCN significantly outperforms existing methods, achieving superior prediction accuracy, generalization, and robustness under varying operational conditions.

Author Contribution:

Contributions: Haitao Wang: Project administration , Supervision, Writing - review & editing. Ruihua Wang: Writing - original draft and Writing - review & editing.Jie Yang:Writing - review & editing.

Foundation Information:

This work is supported in part by the Key Research and Development Program of Shaanxi Province under Grant 2020GY-104; in part by the Key Laboratory of Highway Construction Machinery of Shaanxi Province, Key Laboratory of Road Construction Technology and Equipment (Chang'an University), MOE, under Grant 300102250503; and in part by the Fundamental Research Funds for the Central Universities under Grant CHD 300102250503.

Data Availability:

The data that support the results of this study are available upon reasonable request by contacting the corresponding author.

Conflicts of Interest:

The authors declare no competing interests.

Dates:

Received 6 August 2024; Accepted 8 February;

Published online 31 March 2024

References

[1] WEN Chenglin, LV Feiya, et al. Fault Diagnosis Method Synthesis Based on Deep Learning[J]. *Electronics and Informatics* . **2020**,42(1):234-248.DOI:10.11999/JEIT190715

[2] LIU Hui, Liu Zhenyu, Jia Weiqiang, et al. Current Research and Challenges of Deep Learning for Equipment Remaining Useful Life Prediction[J]. *Computer Integrated Manufacturing Systems* . **2021**, 27(1): 34-52. DOI: 10.13196/j.cims.2021.01.003

[3] PEI Hong, HU Changhua, SI Xiaosheng, et al. Review of Machine Learning based Remaining Useful Life Prediction Methods for Equipment[J]. *Journal of Mechanical Engineering*. **2019**,55(8):1-13. DOI:10.3901/JME.2019.08.001

[4] LEI Yaguo, JIA Feng, Kong Detong. et al. Opportunities and Challenges of Mechanical Intelligent Fault Diagnosis under Big Data[J]. *Journal of Mechanical Engineering* .**2018**,54(05): 94-104.DOI:10.3901/JME.2018.05.094.

[5] SINGlETON R K,STRANGAS E G,AVIYENTE S. Extended Kalman Filtering for Remaining-Useful-Life Estimation of Bearings. *Industrial Electronics IEEE Transactions on* , **2015**, 62(3):1781-1790.DOI:10.1109/TIE.2014.2336616.

[6] QIN Y, CHEN D, XIANG S, ZHU C. Gated Dual Attention Unit Neural Networks for Remaining Useful Life Prediction of Rolling Bearings. *IEEE Transactions on Industrial Informatics*, **2020**,PP(99):1-1. DOI:10.1109/TII.2020.2999442.

[7] CORTES C, VAPNIK V. Support-Vector Networks[J]. *Machine Learning* , **1995**,20(3):273-297.

[8] BREIMAN L. Random Forests. *Machine Learning* , **2001**,45 (1):5-32.

[9] DONGXIAO N,HUI S,JIANQING L,et al. Research on Short-term Power Load Time Series Forecasting Model based on BP Neural Network[C]. 2010 2nd International Conference on Advanced Computer Control. IEEE, **2010**. DOI: 10.1109/ICACC.2010.5486899.

[10] WANG B, LEI Y, LI N, WANG W. Multiscale Convolutional Attention Network for Predicting Remaining Useful Life of Machinery[J]. *IEEE Transactions on Industrial Electronics* , **2021**,68(8):7496-7504, Aug.

[11] LIU H,LIU Z,JIA W, et al. Remaining Useful Life Prediction Using A Novel Feature-Attention-Based end-to-end Approach [J]. *IEEE Transactions on Industrial Informatics* , **2020**, 17(2), 1197-1207.

[12] LUO J,ZHANG X. Convolutional Neural Network based on Attention Mechanism and Bi-LSTM for Bearing Remaining Life Prediction[J]. *Applied Intelligence* , **2022**, 52, 1076 – 1091.

[13] WANG B,LEI Y,LI N,YAN T. Deep Separable Convolutional Network for Remaining Useful Life Prediction of Machinery. *Mechanical systems and signal processing* , **2019**,134(Dec.1): 106330.1-106330.18.

[14] GUO G, YUAN W. Short-term Traffic Speed Forecasting based on Graph Attention Temporal Convolutional Networks [J]. *Neuro computing* , **2020**,410,387 – 393.

[15] FAN L,CHAI Y,CHEN X. Trend Attention Fully Convolutional Network for Remaining Useful Life Estimation[J]. *Reliability Engineering & System Safety* , **2022**,225, 108590.

[16] TIAN J, HAN D, LI M, SHI P. A Multi-Source Information Transfer Learning Method with Subdomain Adaptation for Cross-Domain Fault Diagnosis[J]. *Knowledge-Based Systems* , **2022**,243,108466.

[17] LI B, ZHAO Y P, CHEN Y B. Learning Transfer Feature Representations for Gas Path Fault Diagnosis Across Gas Turbine Fleet[J]. *Engineering Applications of Artificial Intelligence* , **2022**,111,104733.

[18] LI W, HUANG R, LI J, et al. A Perspective Survey on Deep Transfer Learning for Fault Diagnosis in Industrial Scenarios: Theories, Applications and Challenges[J]. *Mechanical Systems and Signal Processing* , **2022**,167, 108487.

[19] PAN Z, MENG Z, CHEN Z, GAO W, SHI Y. A Two-stage Method Based on Extreme Learning Machine for Predicting the Remaining Useful Life of Rolling-element Bearings. *Mechanical Systems and Signal Processing* , **2020**, 144, 106899. DOI:10.1016/j.ymssp.2020.106899.

[20] LV Mingzhu, et al. Prediction of the Remaining Service Life of Bearings under Variable Operating Conditions based on Deep Sub Domain Adaptive Regression Network[J]. *Journal of Bearings* . **2023**, (09): 80-94. DOI: 10.19533/j. issn1000-3762.2023.09.014.

[21] TZENG E, HOFFMAN J, SAENKO K, et al. Adversarial Discriminative Domain Adaptation [C]. Conference on Computer Vision and Pattern Recognition. Honolulu, HI,USA: [s.n.],**2017**:2962-2971.

[22] SUN B, FENG J, SAENKO K. Correlation Alignment for Unsupervised Domain Adaptation. **2017**. DOI: 10.1007/978-3-319-58347-1_8.

[23] LI J, LU K, HUANG Z, ZHU L, SHEN H T. Heterogeneous Domain Adaptation through Progressive Alignment. *Neural Networks and Learning Systems* , IEEE Transactions on, **2019**, 30(5):1381-1391.DOI:10.1109/TNNLS.2018.2868854.

[24] WANG H, YANG J, WANG R, et al. Remaining Useful Life Prediction of Bearings based on Convolution Attention Mechanism and Temporal Convolution Network[J]. *Ieee Access* , **2023**, 11: 24407-24419.

[25] CHEN Xinyang, WANG Sinan, WANG Jianmin, LONG Mingsheng. Representation Subspace Distance for Domain Adaptation Regression. *International Conference on Machine Learning.* **2021** *PMLR* .

[26] QIU X, BAI Y, WANG S. A Novel Unsupervised Domain Ddaptation-based Method for Lithium-ion Batteries State of Health Prognostic[J]. *Journal of Energy Storage* , **2024**, 75, 109358,

[27] NECTOUX P, GOURIVEAU R, MEDJAHER K, RAMASSO E, CHEBEL-MORELLO B, ZERHOUNI N, VARNIER C. PRONOSTIA: An Experimental Platform for Bearings Accelerated Degradation Tests[C]. Proc. IEEE Int. Conf. Prognostics Health Manage, Jun.**2012**, 1-8.

[28] CAO Y, JIA M, DING P, DING Y. Transfer Learning for Remaining Useful Life Prediction of Multi-conditions Bearings based on Bidirectional-GRU Network[J]. *Measurement* , **178**, Article 109287, 2021.

[29] ZHU J, CHEN N, SHEN C. A New Data-driven Transferable Remaining Useful Life Prediction Approach for Bearing under Different Working Conditions[J]. *Mech. Syst. Signal Process* , **139**, Article 106602, 2020.

[30] CHEN H, KONG X, CHEN G, WANG Q, WANG R. Transferable Convolutional Neural Network based Remaining Useful Life Prediction of Bearing under Multiple Failure Behaviors[J]. *Measurement* , **168**, Article 108286, 2021.

[31] P R O Costa, A Akçay, Y Zhang, U Kaymak Remaining Useful Lifetime Prediction via Deep Domain Adaptation Reliability Engineering & System Safety, **195**, 2020, Article 106682.