Received 2 October 2022, accepted 14 November 2022, date of publication 18 November 2022, date of current version 29 November 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3223387

RESEARCH ARTICLE

Residual Life Prediction of Bearings Based on SENet-TCN and Transfer Learning

YAN WANG, HUA DING, AND XIAOCHUN SUN

Shanxi Key Laboratory of Fully Mechanized Coal Mining Equipment, College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, China

Corresponding author: Hua Ding (dinghua2002@163.com)

This work supported in part by the Shanxi Province Unveiled the List of Outstanding Project under Grant 202101010101018, in part by the Special Project of Scientific and Technological Cooperation and Exchange in Shanxi Province under Grant 202104041101003, and in part by the National Natural Science Foundation of China under Grant 52174148.

ABSTRACT Accurate bearing remaining life prediction guarantees safety and continued profitability for the industry. Variable operating conditions of the bearing and difficulty in obtaining corresponding data labels in the industry result in low prediction accuracy of the model. To solve these problems, a bearing life prediction model based on an improved temporal convolutional network and transfer learning is proposed. First, the squeeze-and-excitation network is used to mine and recalibrate the deep features of source domain data. Second, the temporal convolutional network is used to calibrate the relationship between the features and lifetime, and the optimal source domain model is trained. Finally, the transfer learning training is conducted with the source domain model to obtain the transfer model, which can accurately predict the remaining life of the multi-operating condition signal. Comparative experiments were performed on IEEE PHM Challenge 2012 bearing life dataset. The results show that the proposed method can better mine the inherent degradation trend of bearings and effectively improve the prediction accuracy of the remaining useful life. Compared with the existing popular prediction methods, the prediction error was reduced by "20.8%" to "51.5%", which proves the effectiveness and feasibility of the proposed method.

INDEX TERMS Bearing, prediction of residual life, SENet, time convolution network, particle swarm optimisation, life prediction.

I. INTRODUCTION

Rolling bearings are widely used in various machines as a mechanical standard component [1]. Predicting the remaining useful life (RUL) of rolling bearings is extremely important for manufacturing engineering [2]. An accurate RUL can improve the maintainability, security and reliability of equipment while improving production efficiency and reducing production costs. These can help avoid personal injury and property damage caused by mechanical equipment damage.

RUL prediction methods can be roughly divided into two categories: mechanism-based models and data-driven methods [3]. Mechanism-based models establish an equipment degradation model according to the physical structure of the bearing to predict remaining life. The data-driven method uses machine or deep learning algorithms to fit main characteristic data that can reflect the bearing degradation state, thereby predicting the remaining service life of the bearing. The mechanism-based model requires a certain in-depth understanding of the bearing structure, and the mechanism of rolling bearings has a certain complexity. Thus, it is difficult to construct a mechanism model because the data-driven method can grasp the degradation of the bearing only through historical data. RUL prediction must be performed regularly. With the rapid development of artificial intelligence and deep learning in recent years, data-driven bearing life prediction methods have gradually become mainstream [4]. All the fully connected layers in the traditional convolutional neural network with convolutional and pooling layers should be replaced when predicting the life of a bearing; this is to reduce the parameters that the neural network requires for training [5]. Life prediction of rocker gearboxes is based on a long short-term memory model in a practical engineering environment [6]. Based on the

The associate editor coordinating the review of this manuscript and approving it for publication was Hongli Dong.

online monitoring data, the prediction model of an automatic encoder combined with a deep bidirectional gated recursive unit (bi-GRU) was constructed to predict a shearer RUL [7]. By combining DenseNet and squeeze-and-excitation network (SENet), a SE-DenseNet model is proposed to enhance the transmission of deep information, avoid the disappearance of gradients and recalibrate data features [8]. The remaining service life of the bearing is predicted by parallel multi-channel convolution long-and short-term memory network. [9]. The temporal convolution network (TCN) model was constructed to accomplish the RUL prediction by mining the inherent-time-series characteristics of the degradation trend using the TCN [10]. The vibration trend of the rolling bearing is predicted using the TCN, and the residual life of the rolling bearing is predicted by adding an attention mechanism to the TCN [11]. The original feature extraction in the above methods and models is unable to fully use the characteristics of the data and some methods ignore the time-series characteristics of the data. The machine contains different types of bearings in the actual engineering. The data collection is incomplete because of various working conditions and long-term operation of the equipment, and the data cannot be added to the label. The existing model is difficult to meet the actual situation.

In response to the above problems, a transfer learning method is introduced to the field of bearing RUL. Compared with traditional machine learning, which requires the same distribution of training and test data, this method can avoid the labour and material costs caused by re-labelling the obtained data in traditional machine learning [12]. Note that transfer learning is not a single method but a machine learning algorithm [13]. The main idea of transfer learning is to use similar information from the source domain to improve the task performance in the target domain [13]. In transfer learning, we first train a base network on a base dataset and task; then, we reuse or transfer the learnt features to a second target network for training on the target dataset and task. This process works when the features are generic [14].

Presently, transfer learning methods have shown promising effects in several fields, such as text [15], image [16], [17] and software defect classifications [18]. Furthermore, it was successfully applied to the bearing life prediction problem this year. These techniques include the LSTM-DNN transfer network method [19], cross-case method for time-series clustering [20], convolutional neural and long short-term memory networks [21], deep transfer metric learning for kernel regression [22] and bi-GRU network [23]. However, the research on the RUL of rolling bearings, which is based on transfer learning, is still in its infancy. Kang et al. [24] used a semi-supervised migration component analysis method to predict the RUL of rolling bearings under variable operating conditions. However, this method does not consider the influence of the bearing timing characteristics in the RUL during degradation. Chen [10] et al. used a deep time-series feature transfer model for RUL prediction, but they could not effectively mine the data features of the original vibration signal.

Based on the above discussion, a bearing residual-life prediction model based on improved SENet-TCN and transfer learning is proposed to improve the RUL prediction performance under different working conditions. The model combines the SENet and TCN and uses SENet to mine the data features of the original vibration signal, which is used to adaptively construct feature indicators. Furthermore, it uses multilayer TCN to mine the inherent-time series features of bearing degradation trend from the construction feature indicators to improve the RUL accuracy of the model. Furthermore, it is crucial to learn the time-series characteristics of different bearing variable working conditions through transfer learning; the learnt content should be used to train the transfer model. The purpose of improving the prediction effect of bearing RUL under different working conditions is achieved, and comparative experiments were performed on the IEEE PHM Challenge 2012 bearing life dataset to verify the effectiveness of the proposed method.

The contribution of this study can be summarised as follows:

1. The SENet network is improved. The improved SENet reduces the number of calculations, and it is more sensitive to vibration signals, which can effectively build feature indicators.
2. The residual structure of the TCN is optimised, and an attention mechanism is integrated into the TCN so that it can effectively use past information to improve its time-series prediction performance.
3. A transfer learning method based on improved SENet-TCN is proposed to predict the RUL of bearings, which improves the RUL prediction performance of rolling bearings under different working conditions and with fewer data.

The rest of this study is organised as follows: Section 2 introduces the theoretical background. Section 3 describes the technical process of the proposed method in detail. Section 4 presents relevant experiments to verify the effectiveness of the method. Finally, Section 5 concludes the study.

II. THEORETICAL BACKGROUND

A. SENet

The SENet learns the feature weight through the network according to the loss function so that the effective feature has a large weight, while the invalid feature has a small weight. It adaptively and selectively emphasises important features and suppresses unimportant ones. Fig. 1 shows the squeeze excitation structure.

Given an input, the number of feature channels is, W and H are the length and width of the feature map after a series of changes, such as full connection pooling. A feature with the number of feature channels is obtained, and the following three steps are used to recreate and calibrate the preceding features.

1. The squeeze operation performs feature compression on the channel information dimension and compresses the global spatial information into a channel descriptor with a global

receptive field, which is expressed as follows:

$$z_c = F_{sq}(u_c) = \frac{1}{W \times H} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \tag{1}$$

z_c, which represents the output of the squeeze operation, u_c is the C feature of the input matrix and F_{sq} is the compression operation.

2. The excitation operation is an adaptive adjustment, which automatically adjusts the correlation between feature channels and generates weights for each feature channel. The C feature output is s_c.

3. The reweight operation uses s_c weights, which are then multiplied and weighted to u_c channel-by-channel to complete the recalibration of the original features in the channel dimension.

$$\widetilde{x}_c = F_{scale}(u_c, s_c) = s_c \cdot u_c \tag{2}$$

$\widetilde{x}_c$ is the input data of the next level.

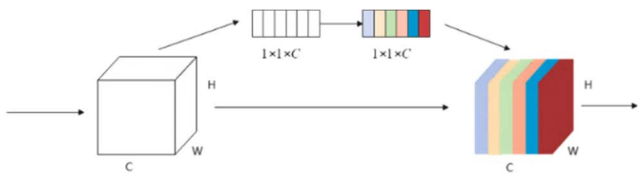

FIGURE 1. Squeeze excitation structure.

B. TCN

Shao-jie Bai et al. proposed the TCN in 2018 [25]. It has been proven in recent years that the sequence data are more accurate, simpler and clearer than standard cyclic networks, such as the LSTM and CNN. TCN consists of causal convolution, dilated convolution and residual connections. The causal convolution processes time-series data, dilates convolution for handling long-distance dependencies common in time-series models and uses residual connections to solve the problems of gradient disappearance and explosion, which may be caused by increasing network depth.

1) CAUSAL CONVOLUTION

The causal convolution is a strictly time-constrained model, and its formula is as follows:

$$P(x_t) = \prod_{t=1}^{T} P(x_t|x_1, x_2, \ldots, x_{t-1}) \tag{3}$$

Here, $P(x_t)$ is the predicted probability, T is the total time and $\prod$ is the quadrature operation. Causal convolution ensures that when predicting time-series data, the data (input) $x_1, x_2, \ldots, x_{(t-1)}$ before time t are used to predict the data (input) x_t at time t.

2) DILATED CONVOLUTION

The dilated convolution has a larger receptive field than the traditional convolution. Its expression calculation F is given as follows:

$$F(x_s) = (x \times f_d)(s) = \sum_{i=0}^{k-1} f(i) x_{(s-di)} \tag{4}$$

Here, $F(x_s)$ is the network output of the input x_s pair at time s, which is in the dilated-convolution calculation process, k is the size of the convolution kernel, d is the expansion coefficient and $s - d_i$ is the sequence corresponding to the elements in the convolution kernel, $i \in (0, 1, \ldots, k-1)$.

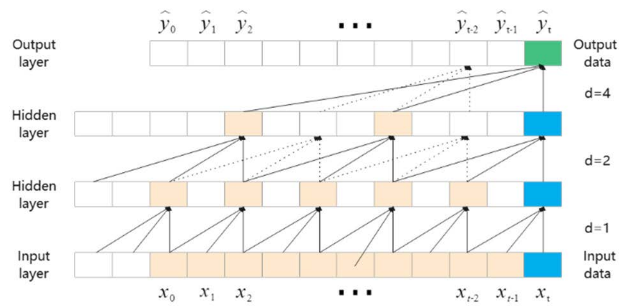

FIGURE 2. Expansion convolution structure of the TCN.

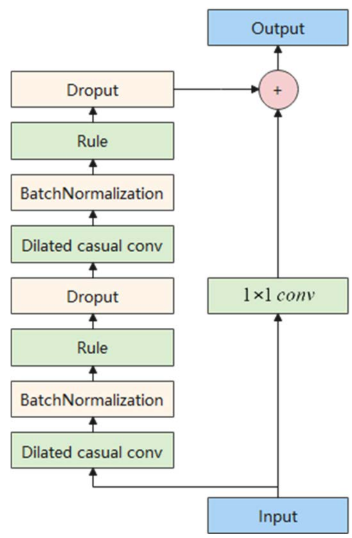

FIGURE 3. TCN residual-connection structure.

Fig. 2 shows the TCN structure for the dilated causal convolution with dilation coefficient d=1, 2 and 4 and convolution kernel size k=3. During the convolution operation, the input data are sampled orderly based on a certain interval, and the sampling rate is controlled by d; d=1 means that every point is collected and d=2 means that every two points are sampled during the input process. Take the first sample point as input such that the higher the level, the larger the value of d.

3) RESIDUAL CONNECTIONS

Fig. 3 shows the residual-connection structure. The input x of the model is weighed and fused into the output $F(x)$ of the model to obtain the final TCN output o. The formula is given as follows:

$$o = Activation[x + F(x)] \tag{5}$$

Here, *Activation* activates a function.

C. TRANSFER LEARNING

Transfer learning is a machine learning method that uses existing knowledge to solve problems in different related fields [12]. The learnt domain is defined as the source domain; D_S is defined as the field to be applied as the target domain and uses D_T means. T_S and T_T means the source and target domain tasks, respectively.

For a given target domain, based on the knowledge of the existing source domain and its task, a transfer model is established combined with the knowledge of the target domain to complete its task. Fig. 4 shows the principle of transfer learning.

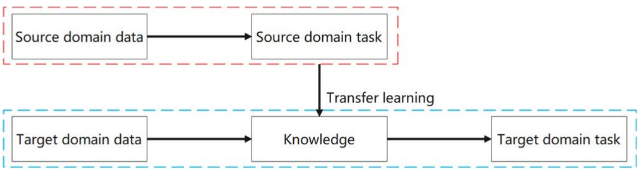

FIGURE 4. Squeeze excitation structure.

III. SENET-TCN MODEL

In this section, we describe the main steps of the SENet-TCN model and its improvements.

A. MODEL STRUCTURE

Fig. 5 (a) shows the basic structure of the proposed SNEet-TCN, which consists of two modules: feature adaptive calibration and RUL prediction modules. The former is composed of a normalisation and a SENet layer. It can effectively mine deep-level features from the original vibration signal while reducing the amount of calculation, and it can perform feature adaptive calibration. The latter consists of three stacked TCN blocks, an average pooling layer and a fully connected layer. It compares the output of the network with the labelled real RUL value and backpropagates the error. Fig. 5 (b) shows the TCN structure. The model input is a two-dimensional tensor that flows through the feature adaptive calibration module, dropout layer and RUL prediction module. First, the model recalibrates the corresponding relationship between the features and the full life. Dropout is used to prevent overfitting, and global average pooling is used to integrate global information for dimension reduction and average feature extraction. Finally, the current life stage according to calibrated sequence features before obtaining the remaining life is predicted. The feature extraction of the model and dimension transformation process are shown in the model architecture (Table 1). As the selection of super-parameters will affect the prediction results, in order to further improve the prediction accuracy of the model, the particle swarm optimization algorithm (PSO) is used to automatically find the optimal values of some super-parameters of the model. The specific process will be described in detail in the PSO section.

TABLE 1. Model architecture.

Layer	Hyperparameters	Input	Output
Input		1×2560×2	1×2560×2
BN		1×2560×2	1×2560×2
SEnet	Stride = 8	1×2560×2	1×40×8
Dropout		1×40×8	1×40×8
TCN	Stride = 1	1×40×8	1×40×4
TCN	Stride = 2	1×40×4	1×40×8
TCN	Stride = 4	1×40×8	1×40×16
GlobalAveragePooling		1×40×16	1×1×16
Dense	Units = 1	1×1×16	1×1

TABLE 2. Params comparison of different prediction methods.

Methods	Parameter quantity
SENet	2232
Improve SENet	1944

B. SENet NETWORK OPTIMISATION STRATEGY

To improve the sensitivity of the SENet network to vibration signals, two one-dimensional convolutional layers are used instead in the excitation process, and the two activation layers recalibrate the data features. Fig. 6 shows the improved excitation structure, and the comparison of the number of parameters is shown in Table 2. Using the nonlinear sigmoid function alone requires a large number of exponential calculations, and information loss will occur when the derivative reaches zero. Calculating the Relu function is very simple and there is no gradient saturation or disappearance. Adding a Relu function before the sigmoid function reduces the amount of network calculation; however, it will not cause information loss.

C. TCN NETWORK OPTIMISATION STRATEGY

The literature [26] compares five structures of residual networks, experiments in the literature show that the fully pre-activated structure is superior to other methods in improving the generalisation ability of the network and preventing overfitting. Therefore, causal convolution is applied after batch standardisation and activation function, and the complete pre-activation of residual connection is considered. Fig. 3 shows the improved residual connection.

The improved TCN uses Leaky Relu as the activation function and assigns a non-zero slope to all negative values, which can be expressed mathematically as follows:

$$y_i \begin{cases} x_i & if\ x_i \geq 0 \\ \dfrac{x_i}{a_i} & if\ x_i \prec 0 \end{cases} \tag{6}$$

Here, a_i is a fixed parameter in the interval $(1, +\infty)$. Leaky Relu has a small positive slope in the negative region; thus, it can perform backpropagation even for negative input values. Furthermore, it has the advantage of a rule activation function.

To make the TCN appropriately use past information, the attention mechanism is integrated into the TCN, and its calculation formula is shown in formulas 6–8:

$$a_t = LeakyRelu(M \cdot y_t^T) \tag{7}$$

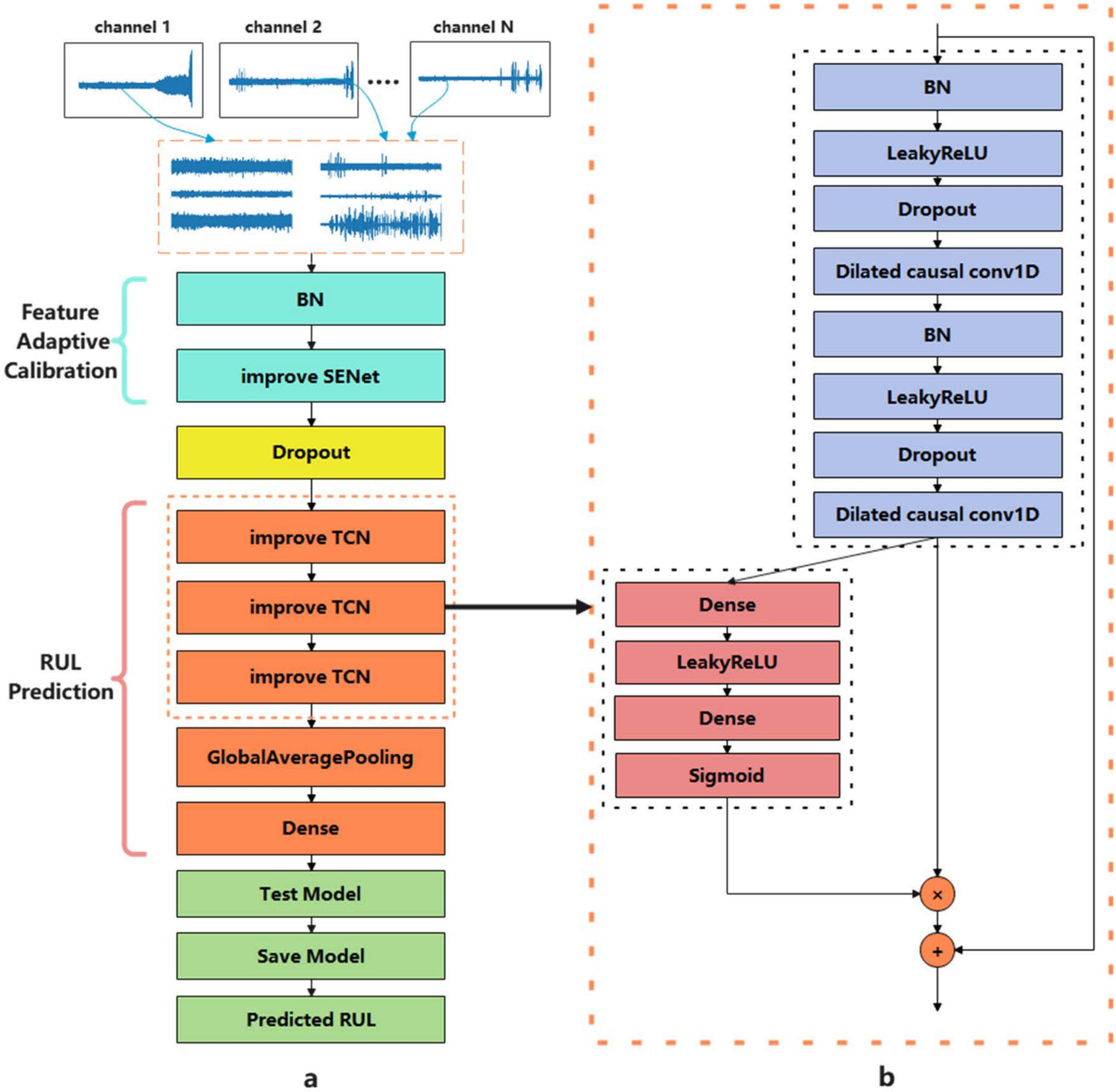

FIGURE 5. **(a) SNEet-TCN basic structure (b) TCN structure.**

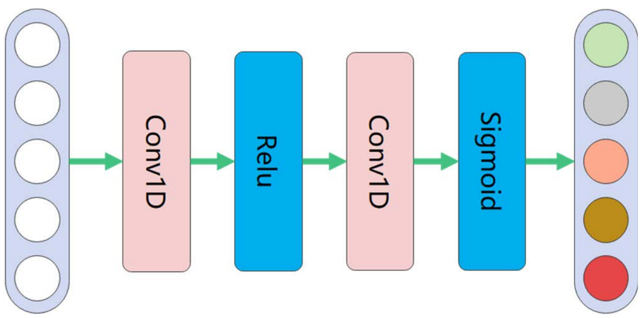

FIGURE 6. **Improved excitation structure.**

$$c_t = a_t^T M \tag{8}$$

$$y_r = soft\max(w[y_t; c_t]) \tag{9}$$

Here, W is a trainable parameter, $soft\max$ $LeakyRelu$ are all activation functions and t is transposed.

The output vector at the current time t is y_t when predicting the life of bearings. The output vector of $t-1$ TCNs before time t is $M = [y_0, y_1, \ldots, y_{t-2}, y_{t-1}]$; the attention score vector a_t is obtained through the relationship between y_t and M. According to a_t gets the vector c_t. Will c_t and y_t combined with life prediction for bearings.

D. HYPERPARAMETER SETTING

SENet and TCN must set hyperparameters, and the setting hyperparameters will affect the accuracy of the prediction results. The PSO solves the problem of hyperparameter selection, which is a probabilistic method of collective movement of biological organisms. It obtains the best solution by optimising the fitness function, which is simple and easy to achieve without numerous parameters. The important hyperparameters in the SENet-TCN are the SENet convolution kernel, number of TCN filters, TCN convolution kernel length and learning rate. According to the variation range of hyperparameters, the upper and lower bounds of the PSO search space are set as [10, 10, 128, 0.01] and [1, 1, 2, 0.0001]; the population size is 20, and the optimal one is automatically selected. As an adaptive optimisation algorithm, Adam can dynamically update the learning rate, and this study chooses an Adam optimiser for gradient optimisation.

IV. SENET-TCN TRANSFER MODEL

Because of the problems in practical engineering, such as the variable operating conditions of a machine, it is impossible to label the incomplete real collected signals. Thus, migration learning can be used to solve this problem. Bearing signals under different working conditions demonstrates good migration characteristics. Through the migration learning of vibration signals, the accuracy of the RUL prediction under multiple working conditions and few samples without labels is improved, thus migrating information features from the source to the target domain. The source and target domains D_S and D_T, respectively, have the same characteristic space $X_S = X_T$ and different probability distributions $P(X_S) \neq P(X_T)$. The specific process of the migration model (Fig. 7) is as follows:

1) Obtain the original vibration signal of the bearing under a certain working condition and label it as D_S (with the RUL label). Next, label the original vibration signal of the bearing under other working conditions as D_T (without the RUL label).
2) T_S trains the source domain model by improving the transferred TCN model. It verifies the model through the verification set and saves the optimal model as the source domain model.
3) T_T inputs the training data with the RUL label in the source domain and those without the RUL label in the target domain into the source domain model to generate a derived training set, which is used to minimise the distribution difference between D_S and D_T. Then, the derived training set is entered into the improved SENet-TCN model for transfer training to generate a transfer model. Finally, the transfer model is verified using the target domain validation set data.
4) Save the transfer model and perform the remaining life prediction on the target domain data.

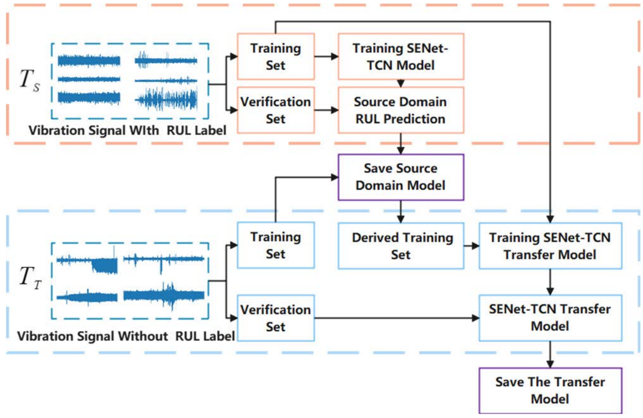

FIGURE 7. Transfer model flowchart.

V. EXPERIMENTAL RESULTS AND DISCUSSION

To verify the effectiveness of the proposed method, IEEE PHM 2012 Data Challenge bearing data [27] is used to experimentally verify the proposed bearing life prediction based on the SENet-TCN and transfer learning. The algorithm running environment is Python 3.7 and Keras 2.3.1; the configurations used in the experiment consist of AMD 4800H processor, NVIDIA 2060 graphics card and 16 GB of memory.

A. CASE INTRODUCTION

The IEEE PHM 2012 dataset was obtained from the Pronostia test bench. Through the accelerated-life degradation experiment of rolling bearings, the vibration acceleration data of the rolling bearing from normal to fault under different operating conditions were collected. The experiment stops when the amplitude of the vibration signal exceeds 20 g. Fig. 8 shows photos of normal and degraded bearings. Bearing failures can result from the failure of balls, rings and cages or a combination of these components. These bearing failure modes cause different degradation trends under different operating conditions.

FIGURE 8. Normal and degradation bearings.

The collected vibration signals are divided into horizontal and vertical directions. The sampling frequency of the data is 25.6 kHz. Data were recorded every 10 s, and the collection time was 0.1 s. Here, 2560 vibration data were collected each time.

The basic characteristics of the bearing in the IEEE PHM 2012 test are listed in Table 3. The test includes three working conditions (Table 4). For the first working condition, the motor speed and load are 1800 rpm and 4000 N, respectively. In the second working condition, the motor speed and load were 1650 rpm and 4200 N, respectively. For the third condition, the motor speed and load are 1500 rpm and 5000 N, respectively. Working Conditions 1 and 2 contain seven different bearings while working Condition 3 contains three different bearings. [23] calculates the theoretical failure frequency of the bearing under three working conditions according to bearing speed and characteristics; results are recorded in Table 5.

Fig. 9 shows the entire life vibration signal diagram of Bearing 1 under working conditions and the vibration signal of the 50th, 1166th and 2500th data points.

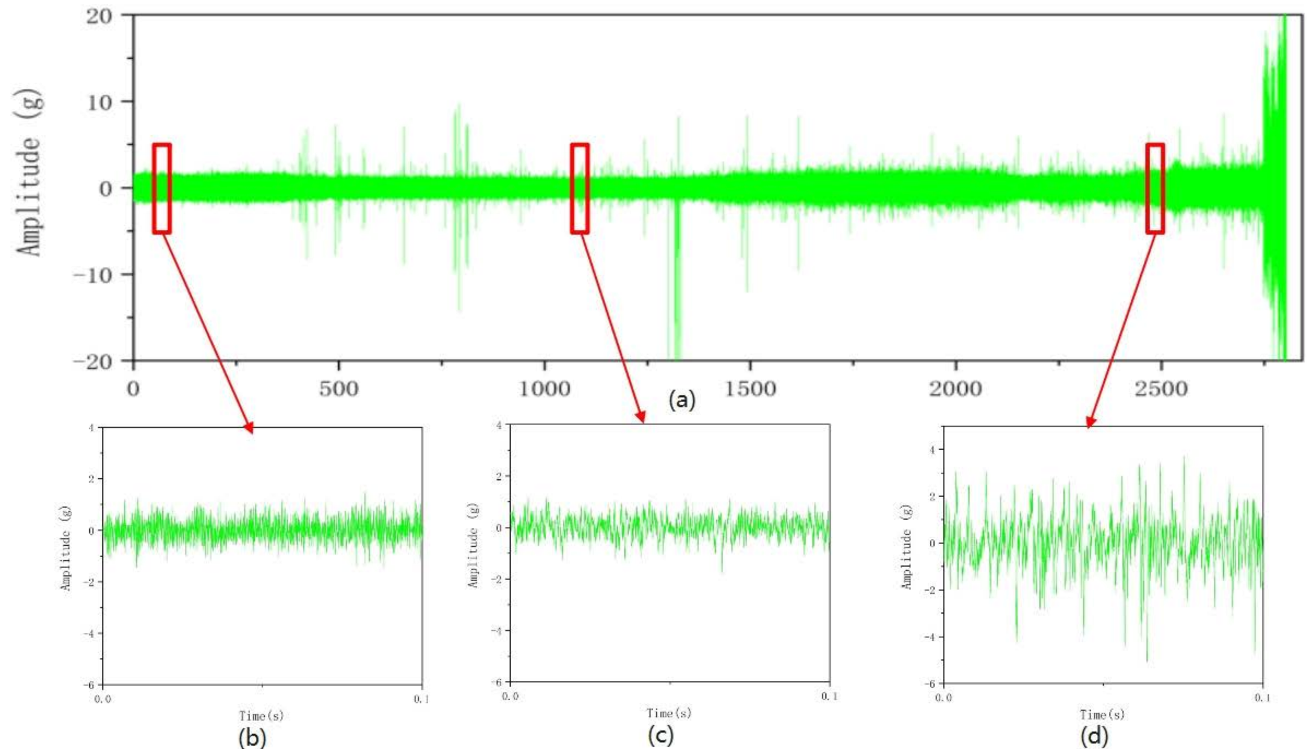

FIGURE 9. B1_1 Vibration signal diagram (a) full life, (b) 50th, (c) 1166 th and (d) 2500th.

TABLE 3. Characteristics of the bearings.

Pitch diameter (Dm)	Roller diameter (d)	Number of rollers (Z)	Contact angle (α)
25.6 mm	3.5 mm	13	0°

TABLE 4. PHM2012 data description.

	Condition 1	Condition 2	Condition 3
Load (N)	4000	4200	5000
Speed (rpm)	1800	1650	1500
Bearings	B1_1	B2_1	B3_1
	B1_2	B2_2	B3_2
	B1_3	B2_3	B3_3
	B1_4	B2_4	
	B1_5	B2_5	
	B1_6	B2_6	
	B1_7	B2_7	

TABLE 5. Theoretical fault frequency of each dataset.

	Condition 1	Condition 2	Condition 3
Inner ring	221.7 Hz	203.2 Hz	184.7 Hz
Outer ring	168.3 Hz	154.3 Hz	140.3 Hz
Cage	12.9 Hz	11.9 Hz	10.8 Hz
Ball	215.4 Hz	197.4 Hz	179.4 Hz

B. PERFORMANCE METRICS

To evaluate the performance of the model, three indicators, mean absolute error (MAE), root mean square error (RMSE) and a score function (score) are used to evaluate the prediction effect of the model. The formula is as follows:

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - y_i| \tag{10}$$

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \widehat{y}_i)^2} \tag{11}$$

$$Score = w_1\frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y}_i| + w_2\frac{1}{m-n}\sum_{n+1}^{m}|y_i - \widehat{y}_i| \tag{12}$$

During the lifetime of a machine, the accuracy of the RUL prediction in the later stages is more critical than in the earlier stages because the machine has a relatively low probability of failure in the earlier stages of its life cycle; thus, a larger weight should be assigned to the later stages. In the given 10,11 and12 formula, m is the number of forecast points; n is the percentage of the early stages to the total stage; i is the serial number of the forecast point; y_i is the actual value; $\hat{y}_i$ is the forecast value; w_1 and w_2 are the early and late weights, respectively. Here, $w_1 = 0.35$, $w_2 = 0.65$, $n = m \div 2$. MAE value range is $[0, +\infty)$, and the smaller the better; the smaller the RMSE value is, the higher the accuracy is; the value of score is in the range of (0, 1), higher values indicate better prediction performance.

C. SENET-TCN MODEL VALIDATION

1) DATA PARTITION

We used the cross-validation method in our experiment, and the data is divided as follows: 1) under working Condition 1, where bearings 1–7 are used as validation data in turn, and the

TABLE 6. Training time of four models.

Methods	Training time
SENet-TCN	1832 s
TCN	1826 s
Improved TCN	1641 s
Proposed model	1214 s

other six bearings are used as training data (with RUL label); 2) under working Condition 2, where bearings 1–7 are used as validation data in turn, and the other six bearings are used as training data (with RUL label).

2) IMPLEMENTATION DETAILS

The service life of the bearing may vary differently, even under the same working conditions. If the true RUL value is directly used as the output label, one RUL value may correspond to various degradation states from the bearings [28]. Significant differences in the life of bearings will cause underfitting [29]. Therefore, we use the life percentage of each bearing as the output label in the data preparation process. This implies that the actual RUL of each bearing is normalised in the range of 0%–100%. We entered the vibration signal and RUL label of the training set into the prediction model for training, trained the optimal model and entered the vibration signal of the verification set into the trained model to predict the verification set, and we obtained the prediction results. The abscissa in the figure is the time (min); the ordinate is the ratio of the remaining life corresponding to the current time of the total life (the full life value is 100, and the end of life is zero). The red and blue lines represent the predicted and actual life values, respectively. Fig. 10 shows the RUL prediction map of bearing 1_3; here, (a) is bearing 1_4, (b) is bearing 1_7, (c) is bearing 2_4, (d) is bearing 2_6, (e) is bearing 2_7 and (f) uses the improved SENet-TCN model.

3) COMPARE WITH RELATED MODELS

To illustrate the effectiveness and superiority of the proposed method, we compared three similar RUL prediction models in our experiments. These prediction models are SENet-TCN, TCN and improved TCN. The hyperparameters of the three models in the training process are the same as those of the proposed method, which uses PSO to find the optimal parameters. Fig. 11 shows the prediction results of the four methods for bearings 1_1 under working Condition 1. Table 7 shows the corresponding evaluation scores of the 14 bearing comparison test results, and Fig. 12 shows the comparison of the MAE and score based on different comparison purposes. The training time of the four methods is shown in Table 6. The comparison results are as follows:

(1) The proposed method is compared with the SENet-TCN to illustrate the effectiveness of improving the SENet. Comparing the prediction graphs and performance indicators, the SENet-TCN prediction results were very poor. The prediction result is worse than using TCN alone. It shows that SENet is insensitive to the bearing vibration signal and will have an adverse effect, but the improved SENet-TCN model showed a good RUL prediction effect under the same data conditions. The unimproved SENet is not sensitive to the bearing vibration signal features, which reduces the prediction accuracy of the TCN. However, the improved SENet can effectively construct feature indicators and improve the RUL accuracy. Combining Tables 2 and 6 reveals that the proposed method has fewer training parameters and shorter training time than SENet-TCN.

(2) Comparing the improved TCN of the proposed method with the unimproved TCN, the improved TCN prediction curve of the proposed method has a better fitting effect and lower error than the real RUL curve. It has been proven that improving the TCN can make better use of its past information to improve its prediction performance for time series.

(3) The comparison between the proposed method and the improved TCN is to verify that the combination of the SENet and the proposed TCN can better improve the RUL prediction accuracy based on the improved TCN. The results show that the proposed method has higher accuracy, and the prediction error is reduced by 20.8%–51.5% compared with the other three models.

4) COMPARE WITH RELATED WORKS

In the experiments, the bearing RUL prediction results are compared with related studies [30], [31] on the same dataset to further verify the performance of the proposed method. Table 8 shows the evaluation score of the comparison results, and Fig. 13 shows the comparison of the MAE.

Table 7 and Fig. 12 show that the method proposed in this study has the lowest percentage error, absolute percentage error and highest score. This result demonstrates the usability of the proposed RUL prediction method. In conclusion, the results show that the improved SENet-TCN has better accuracy than other methods in dealing with the RUL prediction of bearings.

5) MODELS WITH AND WITHOUT PSO COMPARISON

To verify the advantage of using the PSO to automatically select hyperparameters in the proposed method, the proposed method is compared with itself to remove the PSO. The hyperparameter SENet convolution kernel was set to two after removing the PSO; the number of TCN filters was one and the TCN convolution kernel length was 0.008. Table 9 shows the comparison results.

Furthermore, Table 9 shows that the performance parameters of the proposed method under PSO-adding conditions are significantly higher than those of the model without the PSO, which proves the advantage of adding the PSO to automatically select the optimal hyperparameters.

D. TRANSFER MODEL VALIDATION

1) TRANSFER EXPERIMENT DATA PARTITION

When verifying the migration model experiment to simulate the changing operating conditions of the machine in the real project, the actual signal acquisition is incomplete and the

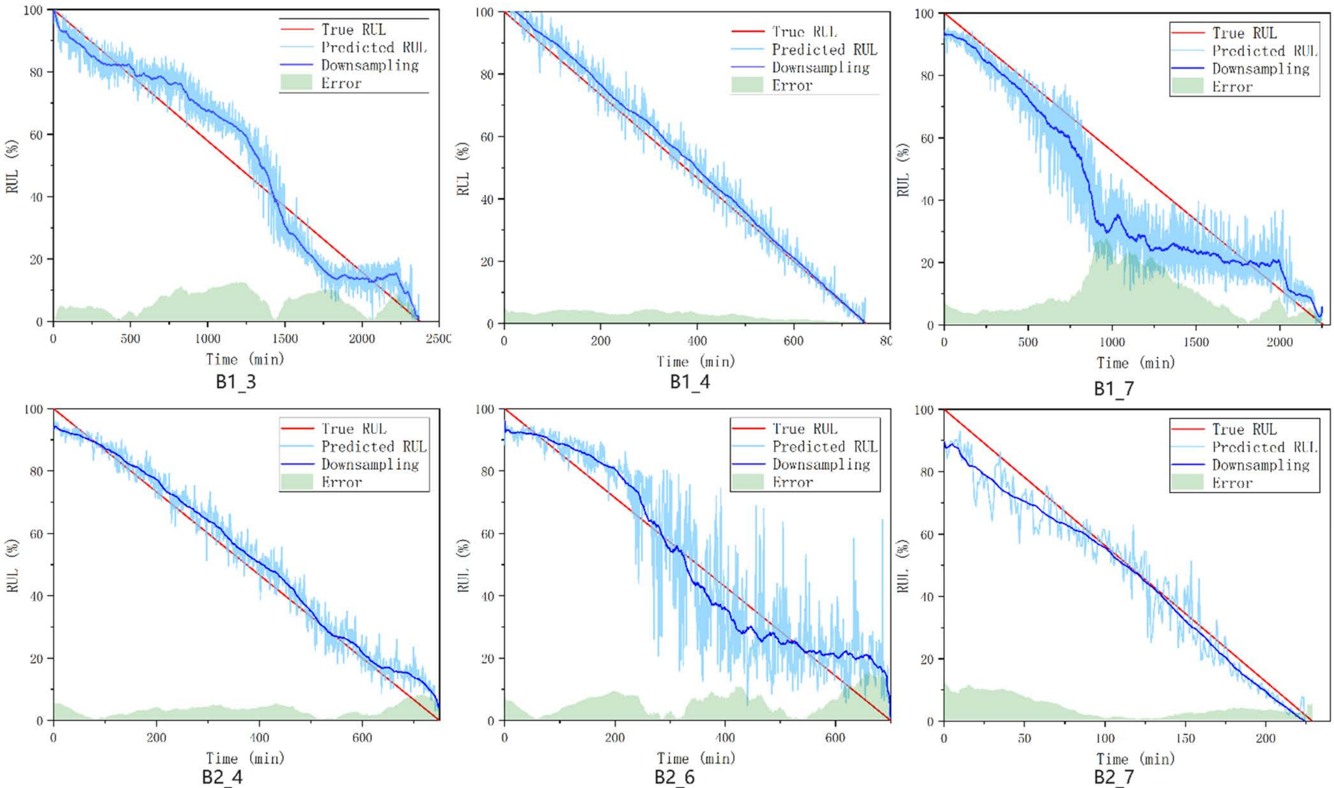

FIGURE 10. The bearing of an RUL prediction diagram of the proposed model.

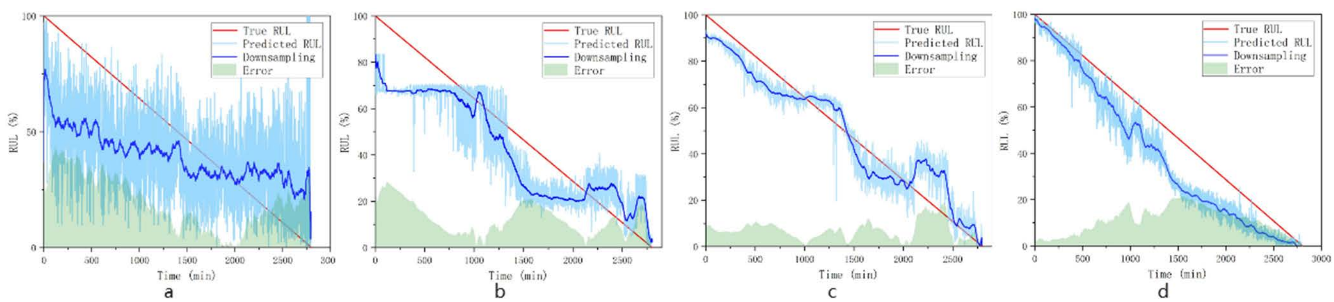

FIGURE 11. RUL prediction results of bearings B1_1 (a) SENet-TCN, (b) TCN, (c) Improved TCN and (d) Proposed model.

TABLE 7. Performance comparison of the four models.

Bearing	SENet-TCN			TCN			Improved TCN			Proposed model		
	MAE	RMSE	Score	MAE	RMSE	Score	MAE	RMSE	Score	MAE	RMSE	Score
B 1_1	29.5	31.6	0.38	21.8	27.4	0.67	7.1	8.6	0.82	6.0	7.2	0.91
B 1_2	18.9	22.8	0.47	36.1	43.7	0.27	19.7	24.8	0.41	9.3	11.7	0.71
B 1_3	34.3	56.9	0.69	24.0	38.0	0.50	12.1	15.3	0.65	5.4	6.8	0.85
B 1_4	25.6	26.7	0.91	5.9	7.5	0.90	5.6	6.6	0.86	4.2	5.7	0.91
B 1_5	24.6	30.7	0.38	21.4	26.7	0.58	12.1	16.2	0.70	6.9	11.	0.82
B 1_6	51.7	57.0	0.58	21.5	27.7	0.58	11.2	15.4	0.59	6.9	8.9	0.79
B 1_7	35.7	40.4	0.68	16.9	21.4	0.65	14.2	18.9	0.83	8.8	10.8	0.75
B 2_1	26.8	31.9	0.61	21.3	26.4	0.61	16.8	20.7	0.71	13.4	15.3	0.83
B 2_2	20.1	26.2	0.61	20.1	26.2	0.50	25.6	33.0	0.35	9.8	14.8	0.70
B 2_3	25.5	30.7	0.60	20.5	25.7	0.59	28.2	35.1	0.74	15.9	20.9	0.73
B 2_4	24.7	24.4	0.87	6.1	7.5	0.87	7.8	9.5	0.71	4.1	5.0	0.9
B 2_5	27.5	32.4	0.48	27.2	32.8	0.63	18.6	20.4	0.34	9.1	11.8	0.82
B 2_6	21.5	25.1	0.50	19.3	25.6	0.58	17.8	22.9	0.45	7.4	9.7	0.83
B 2_7	24.9	29.2	0.56	27.8	34.7	0.39	20.7	24.8	0.40	6.1	7.7	0.90

collected signals have no labels. Table 10 shows the transfer of the experimental data, and working Condition 1 is used as the source domain data (label). Case 2 (unlabelled) and Case 3 (unlabelled) are used as target domain data, and the

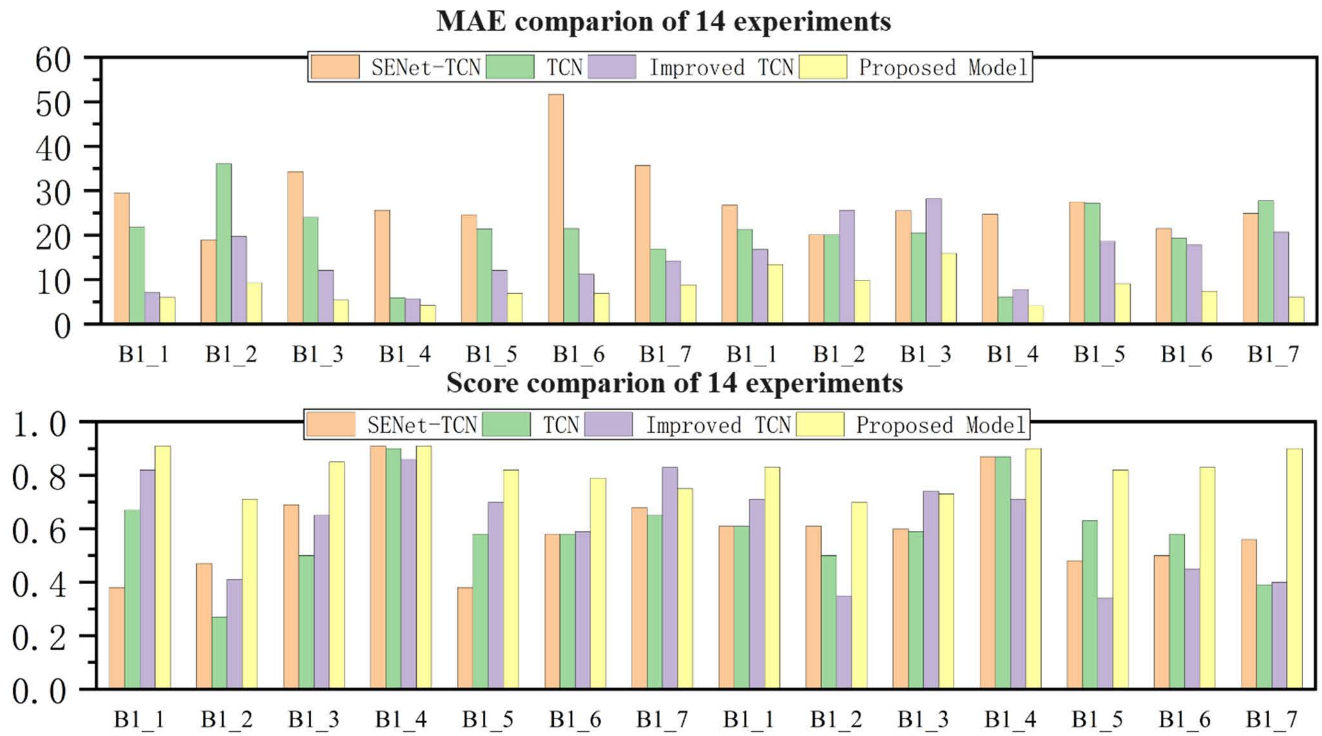

FIGURE 12. MAE and score comparison of the four models.

TABLE 8. Performance comparison of related works.

Bearing	The method in [18]			The method in [19]		Proposed model		
	MAE	RMSE	Score	MAE	RMSE	MAE	RMSE	Score
B 1_1	10.4	11.3	0.87	13.3	16.0	6.0	7.2	0.91
B 1_2	13.0	16.9	0.80	15.3	17.2	9.3	11.7	0.71
B 1_3	9.7	11.7	0.81	28.4	33.5	5.4	6.8	0.85
B 1_4	7.1	8.5	0.73	22.6	25.1	4.2	5.7	0.91
B 1_5	9.11	13.0	0.83	13.4	15.6	6.9	11.	0.82
B 1_6	8.3	11.9	0.68	13.5	17.7	6.9	8.9	0.79
B 1_7	9.0	12.9	0.75	15.2	19.3	8.8	10.8	0.75
B 2_1	14.3	18.1	0.74	16.3	19.0	13.4	15.3	0.83
B 2_2	14.5	18.6	0.44	26.3	29.4	9.8	14.8	0.70
B 2_3	17.9	23.0	0.77	13.5	16.8	15.9	20.9	0.73
B 2_4	5.18	6.42	0.91	12.0	14.3	4.1	5.0	0.9
B 2_5	13.3	16.1	0.78	22.7	27.5	9.1	11.8	0.82
B 2_6	10.9	13.8	0.83	20.6	23.3	7.4	9.7	0.83
B 2_7	17.5	25.0	0.53	12.6	13.9	6.1	7.7	0.90

source domain data and source domain task knowledge are migrated to the target domain task. The specific operations are as follows:

1) Under working Condition 1, bearing 1_1 is the source domain data to be tested, and the remaining six bearings under working Condition 1 are source domain training data (with RUL labels) used to train the source domain model.
2) Select one of bearings 2_1, 2_2 and 2_3 under working Condition 3 as the target domain to test data (with RUL label), and the remaining two bearings are the target domain training data (without RUL label). Then, train the transfer model according to the steps in IV and verify the model with the data to be tested in the target domain, thereby leading to three verifications.
3) Select one of bearings 3_1, 3_2 and 3_3 under working Condition 3 as the target domain to test data (with RUL label), and the remaining two bearings are the target domain training data (without RUL label). Then, train the transfer model according to the steps in IV, and verify the model with the data to be tested in the target domain, thereby leading to three verifications.

2) COMPARISON WITH AND WITHOUT TRANSFER LEARNING

To prove the advantages of adding transfer learning to the article, the proposed transfer model and method without transfer learning are compared under the same dataset conditions; the results are shown in Table 11.

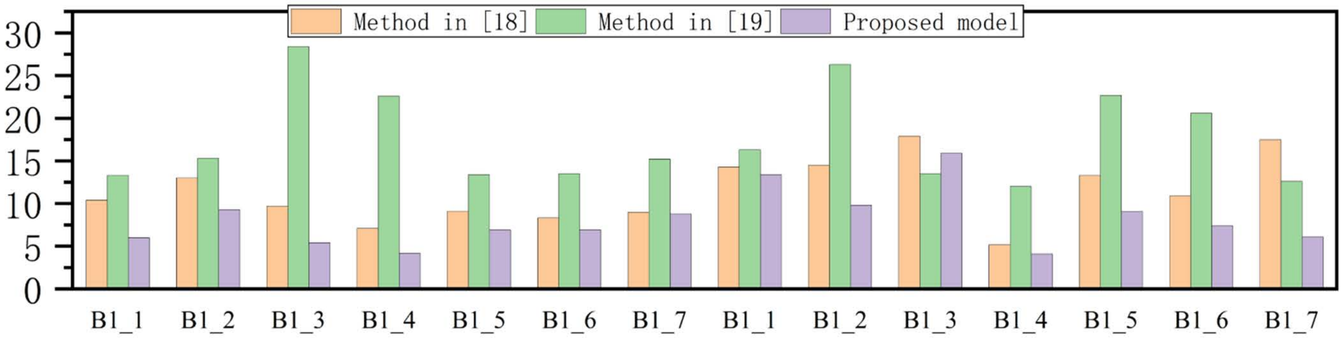

FIGURE 13. The MAE and score comparison of related works.

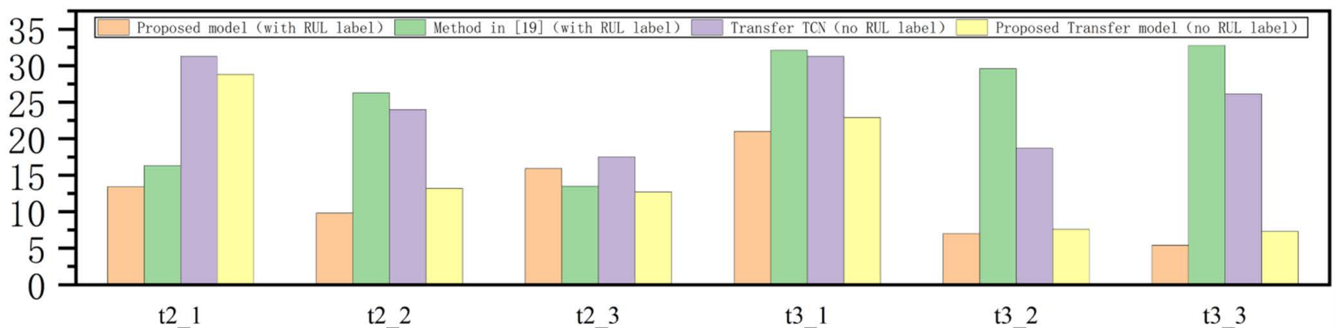

FIGURE 14. Transfer experiment MAE performance comparison.

TABLE 9. With or without PSO performance comparison.

Beari ng	Proposed model no pro			Proposed model		
	MAE	RMSE	Score	MAE	RMSE	Score
B 1_1	15.4	17.9	0.78	6.0	7.2	0.91
B 1_2	13.2	15.0	0.71	9.3	11.7	0.71
B 1_3	9.1	11.3	0.83	5.4	6.8	0.85
B 1_4	5.0	6.2	0.89	4.2	5.7	0.91
B 1_5	10.8	13.7	0.56	6.9	11.	0.82
B 1_6	10.7	12.4	0.73	6.9	8.9	0.79
B 1_7	12.1	15.8	0.73	8.8	10.8	0.75
B 2_1	21.7	26.8	0.66	13.4	15.3	0.83
B 2_2	13.6	16.8	0.67	9.8	14.8	0.70
B 2_3	17.6	21.9	0.67	15.9	20.9	0.73
B 2_4	7.5	9.1	0.86	4.1	5.0	0.9
B 2_5	19.3	25.0	0.50	9.1	11.8	0.82
B 2_6	12.7	15.2	0.78	7.4	9.7	0.83
B 2_7	8.4	10.5	0.77	6.1	7.7	0.90

TABLE 10. The transfer experiment data classification.

Source domain with RUL Label （ Condition 1）	Target domain without RUL label1 （Condition 2）	Target domain without RUL label2 （Condition 3）
t1_1	t2_1	t3_1
t1_2	t2_2	t3_2
t1_3	t2_3	t3_3
t1_4		
t1_5		
t1_6		
t1_7		

Table 11 shows that the error of each performance index of the proposed transfer model is lower and score is higher, which is better than that of the proposed method without transfer learning. Thus, the advantages of adding transfer learning are demonstrated.

TABLE 11. Comparison of performance with or without transfer learning.

Bearing	Proposed model without transfer			Proposed transfer model		
	MAE	RMSE	Score	MAE	RMSE	Score
t2_1	30.3	37.8	0.70	28.8	25.1	0.69
t2_2	26.6	32.4	0.65	13.2	16.6	0.75
t2_3	20.9	27.5	0.40	12.7	14.7	0.69
t3_1	31.4	36.4	0.24	22.9	29.0	0.37
t3_2	11.6	15.9	0.59	7.6	10.0	0.73
t3_3	11.0	13.1	0.79	7.3	8.8	0.82

3) COMPARISON OF RELATED TRANSFER MODELS

The proposed transfer model was compared with the TCN transfer model to illustrate its effectiveness and superiority.

Fig. 15 shows the prediction result of bearing 3_2 using the improved SENet-TCN transfer model; here, the predicted RUL curve generated by the unlabelled data through the bearing remaining life prediction method, which is based on the SENet-TCN and transfer learning, still fits well with the real RUL curve.

Table 12 shows the evaluation score of the compared results of the migration experiment, and Fig. 14 shows the MAE performance comparison. Comparing the performance indicators of the two models, the prediction performance of the proposed migration model is better than that of the TCN migration model. Thus, the feasibility and superiority of the proposed transfer model are proved.

TABLE 12. **Transfer experiment performance comparison.**

Bearing	The proposed model (with RUL label)			The method in [19] (with RUL label)		Transfer TCN (no RUL label)			Proposed transfer model (no RUL label)		
	MAE	RMSE	Score	MAE	RMSE	MAE	RMSE	Score	MAE	RMSE	Score
t2_1	13.4	15.3	0.83	16.3	19.0	31.3	38.7	0.67	28.8	25.1	0.69
t2_2	9.8	14.8	0.70	26.3	29.4	24.0	28.3	0.71	13.2	16.6	0.75
t2_3	15.9	20.9	0.73	13.5	16.8	17.5	21.5	0.45	12.7	14.7	0.69
t3_1	17.0	20.8	0.60	32.1	36.6	31.3	36.8	0.26	22.9	29.0	0.37
t3_2	7.0	9.2	0.79	29.6	34.8	18.7	24.7	0.44	7.6	10.0	0.73
t3_3	5.4	6.9	0.89	32.8	38.3	26.1	32.6	0.39	7.3	8.8	0.82

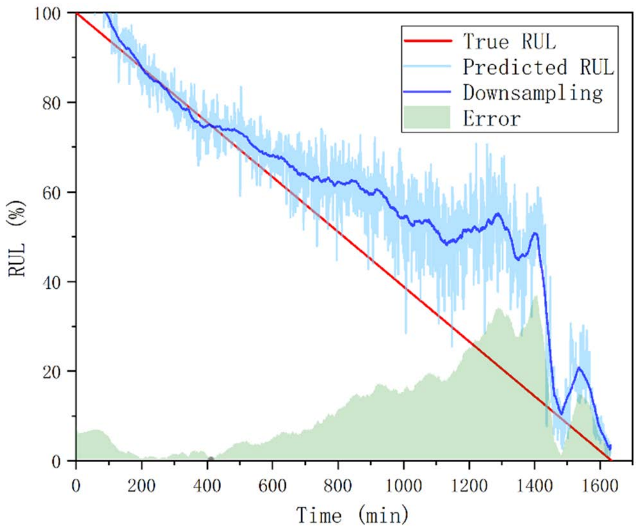

FIGURE 15. **The prediction result of the improved SENet-TCN transfer model for bearing 3_2.**

4) COMPARISON WITH RUL-LABELLED MODELS

In the experiments, the RUL prediction results of the proposed SENet-TCN transfer model are compared with the proposed SENet-TCN and the same dataset in the study [31] under the condition that the training data have RUL labels, which is used to further verify the performance of the proposed transfer method. Table 12 and Fig. 14 show the comparison results.

By comparing the RUL prediction results of the proposed SENet-TCN model under the condition that the training data has RUL labels, we observed that the SENet-TCN training with RUL labels has the best prediction results and its results are the closest to the real values. However, tags cannot be added to practical projects in most cases.

Through the proposed SENet-TCN migration model, the prediction results of the model with the RUL label on the same data [31] are compared. The prediction results of other bearings were better except for the prediction error of bearing 2_1, which was higher than the prediction results of the study [31]. The study [31] has a large error in the prediction performance index under the working Condition 3. The proposed SENet-TCN transfer model has a small error in the prediction performance index under working Conditions 2 and 3 of the target domain, which can meet the actual needs. The proposed SENet-TCN transfer model can still achieve excellent prediction results for multi-mode target tasks under Conditions 2 and 3 using Condition 1 as the source data.

Therefore, under the condition of no labels in the training data, the proposed bearing residual-life prediction method based on improved SENet-TCN and transfer learning can still achieve good results in predicting bearing RUL under multiple working conditions.

VI. CONCLUSION

A bearing residual-life prediction method based on improved SENet-TCN and transfer learning was proposed to address the problems of bearing operating conditions under multiple operating conditions because the actual signal cannot be added to the label and the prediction accuracy is low. Through the RUL prediction comparison experiment and the transfer model RUL prediction comparison experiment on the IEEE PHM Challenge 2012 bearing life data set, the following conclusions are drawn:

1) The experiments show that by improving the structure of SENet to make it more sensitive to vibration signals, it can better construct characteristic indicators. The residual structure of the TCN is improved, and the attention mechanism is integrated into the TCN. Furthermore, the experiments show that it can make better use of past information to improve the prediction performance of time series.
2) Compared with other existing popular models, the proposed improved transmitted TCN model has a better fitting effect, higher accuracy and 20.8%–51.5% reduction in prediction error under the same data conditions.
3) Introducing transfer learning can effectively solve the problem of limited RUL prediction results in the case of multiple operating conditions, fewer data and no labels. Compared with other existing popular models, the results show the feasibility and superiority of the transfer model in this research.

REFERENCES

[1] Y. Lei, N. Li, and S. Gontarz, "A new method of bearing life prediction based on feature clustering and evaluation," *J. Vib. Shock*, vol. 41, no. 5, pp. 141–150, May 2022.

[2] H. Ding, L. Yang, and Z. Cheng, "A remaining useful life prediction method for bearing based on deep neural networks," *Measurement*, vol. 172, no. 1, pp. 108878–108895, 2021.
[3] Z. Zhi-Hong, L. Qing, and Y. Shao-Pu, "Remaining useful life prediction based on BiLSTM and attention mechanism," *J. Vib. Shock*, vol. 41, no. 6, pp. 44–50, Apr. 2022.
[4] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, May 2018.
[5] Z. Ji-Dong, Z. Yi-Sheng, and D. Jia-Lin, "Bearing remaining life prediction based on full convolutional layer neural networks," *China Mech. Eng.*, vol. 30, no. 18, pp. 2231–2235, Oct. 2019.
[6] L. Wen-Guang, Y. Lan-Zhu, and Q. Bo, "Application of LSTM network in life prediction of shearer rocker gearbox," *Mech. Electr. Eng. Technol.*, vol. 49, no. 11, pp. 168–170, Jan. 2021.
[7] H. Ding, L. Yang, and Z. Yang, "A predictive maintenance method for shearer key parts based on qualitative and quantitative analysis of monitoring data," *IEEE Access*, vol. 7, pp. 108684–108702, 2019.
[8] W. Ren-Biao, Z. Ting, and Q. Jing-Yi, "Flight delay prediction model based on deep SE-DenseNet," *J. Electron. Inf. Technol.*, vol. 41, no. 6, pp. 1510–1517, Jun. 2019.
[9] Z. Da-Yi, Y. Ji-Hong, and Z. Yi-Sheng, "Bearing life prediction method based on PMCCNN-LSTM," *China Mech. Eng.*, vol. 31, no. 20, pp. 2454–2462, Nov. 2020.
[10] C. Jia-Xian, W. Mao, J. Liu, and X. Zhang, "Remaining useful life prediction of bearing based on deep temporal feature transfer," *Control Decis.*, vol. 36, no. 7, pp. 1699–1706, Mar. 2021.
[11] L. Ya-Ping, L. Su-Jie, and M. Bo, "Research on vibration trend and remaining life prediction of rolling bearings based on TCN," *Mod. Manuf. Eng.*, vol. 9, pp. 1671–3133, Dec. 2021.
[12] F. Z. Zhuang, P. Luo, Q. He, and Z. Shi, "Survey on transfer learning research," *J. Softw.*, vol. 26, no. 1, pp. 26–39, 2015.
[13] S. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, pp. 1345–1359, 2010.
[14] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.
[15] C. Wang and S. Mahadevan, "Heterogeneous domain adaptation using manifold alignment," in *Proc. 22nd Int. Joint Conf. Artif. Intell.*, vol. 2, 2021, pp. 1–6.
[16] Y. Gao and K. M. Mosalam, "Deep transfer learning for image-based structural damage recognition," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 9, pp. 748–768, Apr. 2018.
[17] M. Danish, M. M. Khan, and S. Parveen, "Real-time driver's drowsiness detection using transfer learning," in *Application of Communication Computational Intelligence and Learning.*, vol. 1, no. 1. Evanston, IL, USA: Routledge, 2022, pp. 73–82.
[18] J. Nam, W. Fu, S. Kim, T. Menzies, and L. Tan, "Heterogeneous defect prediction," *IEEE Trans. Softw. Eng.*, vol. 44, no. 9, pp. 874–896, Sep. 2018.
[19] N. Ding, "A novel method for journal bearing degradation evaluation and remaining useful life prediction under different working conditions," *Measurement*, vol. 177, no. 1, pp. 109273–109290, Jun. 2021.
[20] W. Mao, J. He, B. Sun, and L. Wang, "Prediction of bearings remaining useful life across working conditions based on transfer learning and time series clustering," *IEEE Access*, vol. 9, pp. 135285–135303, 2021.
[21] W. Xin-Gang, H. Kai-Zhong, and W. Chao, "Bearing remaining useful life prediction method based on transfer learning," *J. Northeastern Univ.*, vol. 42, no. 5, pp. 665–672, May 2021.
[22] Y. Ding, "Remaining useful life estimation using deep metric transfer learning for kernel regression," *Rel. Eng. Syst. Saf.*, vol. 212, no. 1, pp. 107583–107590, Aug. 2021.
[23] Y. Cao, "Transfer learning for remaining useful life prediction of multi-conditions bearings based on bidirectional-GRU network," *Measurement*, vol. 178, no. 1, pp. 109287–109290, Jun. 2021.
[24] K. Shou-Qiang, H. Ming-Wu, and W. Yu-Jing, "Fault diagnosis method of a rolling bearing under variable working conditions based on feature transfer learning," *Proc. CSEE*, vol. 3, no. 1, pp. 764–772, May 2019.
[25] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," 2018, *arXiv:1803.01271*.
[26] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision—ECCV 2016*, vol. 9908. Cham, Switzerland: Springer, Sep. 2016, pp. 630–645.
[27] P. Nectoux, R. Gouriveau, and K. Medjaher, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (PHM)*, Jun. 2012, pp. 1–8.
[28] B. Wang, "Deep separable convolutional network for remaining useful life prediction of machinery," *Mech. Syst. Signal Process.*, vol. 134, no. 1, pp. 106313–106330, Dec. 2019.
[29] Y. Gu, "An optimal sample data usage strategy to minimize overfitting and underfitting effects in regression tree models based on remotely-sensed data," *Remote Sens.*, vol. 8, no. 11, pp. 943–956, Nov. 2016.
[30] Y. Wang, L. Deng, L. Zheng, and R. X. Gao, "Temporal convolutional network with soft thresholding and attention mechanism for machinery prognostics," *J. Manuf. Syst.*, vol. 60, pp. 512–526, Jul. 2021.
[31] X. Li, W. Zhang, H. Ma, Z. Luo, and X. Li, "Data alignments in machinery remaining useful life prediction using deep adversarial neural networks," *Knowl.-Based Syst.*, vol. 197, no. 7, pp. 105843–105856, Jun. 2020.

YAN WANG received the B.S. degree in vehicle engineering from the Haibin College, Beijing Jiaotong University. He is currently pursuing the M.S. degree in mechanical engineering with the Taiyuan University of Technology. His research interests include deep learning and remaining useful life estimation.

HUA DING received the Ph.D. degree with a major in mechanical design and theory from the Taiyuan University of Technology, Shanxi, China, in 2011. She is currently a Professor and a Doctoral Supervisor with the Department of Mechanical Design, College of Mechanical and Vehicle Engineering, Taiyuan University of Technology. She was with the Postdoctoral Workstation, Taiyuan Heavy Coal Machine Company Ltd., from 2013 to 2016, and U.S. Pacific University, as a Visiting Scholar, for one year. Her research interests include intelligent design, deep learning, and virtual reality.

XIAOCHUN SUN received the B.S. degree in mechanical design and manufacturing from the North China Institute of Science and Technology. He is currently pursuing the Ph.D. degree in mechanical engineering with the Taiyuan University of Technology. His research interests include deep learning and fault diagnosis.