Remaining Useful Life Prediction of Machinery Equipment via Deep Learning Approach Based on Separable CNN and Bi-LSTM

İbrahim Eke[1](✉) and Ahmet Kara[2]

[1] Graduate Education Institute, Foreign Trade and Supply Chain Management, Hitit University, Çorum, Turkey
ibrahimeke05@gmail.com.tr

[2] Department of Industrial Engineering, Hitit University, Çorum, Turkey
ahmetkara@hitit.edu.tr

Abstract. Predictive maintenance occupies a significant role to drop the operation and maintenance costs in production systems. Remaining useful life (RUL) prediction is one of the most preferred tasks in predictive maintenance decisions. Recently, deep learning techniques are extensively employed to accurately and effectively predict remaining useful life (RUL) by examining the past deterioration data of machinery and equipment failures. In this study, a deep learning approach that includes multiple separable convolutional neural networks (CNN), a bidirectional long short-term memory (Bi-LSTM) and fully-connected layers (FCL) are proposed to ensure more effective predictive maintenance planning. Separable CNN layers are applied to learn the nonlinear and sophisticated dependencies from the raw degradation data while the Bi-LSTM layer is employed to capture the long-short temporal characteristics. Besides, the dropout method and L2 regularization are used in the training stage of the proposed deep learning approach to achieve more accurate learning. The effectiveness of the proposed approach is verified by the popular FEMTO-bearing dataset presented by NASA. Finally, it is aimed that the experimental results provide better prognostic prediction compared with the benchmark models.

Keywords: Predictive maintenance · deep learning · prognostic prediction · separable convolution

1 Introduction

Maintenance of machine equipment is of paramount importance in the industrial and manufacturing sectors, as it directly impacts the efficiency, productivity, and profitability of an organization. Through the implementation of regular and systematic maintenance practices, machinery and equipment can operate at their optimal performance levels, reducing downtime and minimizing the risk of unexpected breakdowns [1].

Predictive maintenance is an advanced technology-based approach that focuses on predicting the future health and performance of equipment or systems, as well as detecting and diagnosing faults and failures in real-time [2]. It is an integrated process that

Z. Şen et al. (Eds.): IMSS 2023, LNME, pp. 128–137, 2024.
https://doi.org/10.1007/978-981-99-6062-0_13

involves the collection, analysis, and interpretation of data from various sources, such as sensors, diagnostics, and modelling, to provide insights into the condition of equipment or systems. Predictive maintenance is a proactive approach to predicting the future performance and health of a system, such as a machine, based on real-time data analysis. This approach involves the use of sensors, data analytics, and machine learning to monitor the health of a system and predict when maintenance is needed, which allows maintenance personnel to take corrective action before a failure occurs. The goal of predictive maintenance is to improve the reliability, availability, and safety of systems by detecting and diagnosing problems early before they result in downtime or failure, which can significantly reduce costs and increase efficiency [3].

Traditional predictive maintenance techniques rely on statistical and machine learning algorithms to analyze historical and real-time data to predict equipment failures and recommend maintenance actions. However, these techniques can be limited by the complexity and variability of data, which can make it difficult to identify patterns and relationships. Deep learning can overcome these limitations by automatically discovering patterns and relationships in complex data sets, including data from sensors, logs, and other sources. By training deep neural networks on large data sets, Deep learning algorithms can learn to identify patterns and relationships that are not easily detected by traditional machine learning techniques. This can lead to more accurate predictions of equipment failures and better recommendations for maintenance activities. Another advantage of deep learning is its ability to adapt to changing conditions. These algorithms can learn from new data as it becomes available, allowing them to adapt to changes in equipment performance and environmental conditions. This can help ensure that predictive models remain accurate and effective over time.

Predicting impending failure and estimating remaining useful life (RUL) is essential to avoid abrupt breakdown and schedule maintenance [4]. Increasing the accuracy of RUL prediction depends on determining the fundamental relationship between bearing deterioration progression and the current state of health. Therefore, the relationship between the two is also very important. To determine this relationship, effective feature compression and optimum feature selection are required. Similarly, it is difficult to determine a failure threshold since the health indicators of different machines are often different at the time of a failure [5]. Shen and Tang [3] proposed a novel data-driven method to address the challenge of data redundancy and initial prediction time in RUL prediction. This method involves extracting time-frequency features of vibration signals, constructing a nonlinear degradation indicator, and applying an attention mechanism called Multi-Head Attention Bidirectional-Long-Short-Term-Memory (MHA-BiLSTM). In the model proposed by Jiang et al. [6], a convolutional neural network (CNN) and an attention-based long short-term memory (LSTM) are used to partition a time series into multiple channels and improve performance by different deep learning approaches. Ren et al. [7] introduced a new method for the prediction of bearing RUL based on deep convolution neural network (CNN) and a new feature extraction method called the spectrum-principal-energy-vector. Yang et al. [8] addressed a new deep learning-based approach for predicting the remaining useful life (RUL) of rolling bearings based on long-short term memory (LSTM) with uncertainty quantification. The proposed method includes a fusion metric and an improved dropout method based on nonparametric kernel density to accurately estimate the RUL. Gupta et al. [9] addressed a deep learning approach for the real-time condition-based monitoring of bearings. A CNN-BILSTM model with attention mechanism for predicting an automatic RUL of bearings is developed by Xu et al. [10].

Furthermore, Sun et al. [11] presented a hybrid deep learning-based technique combining the convolutional neural network (CNN) and long short-term memory (LSTM) network to predict the short-term degradation of a fuel cell system used for commercial vehicles. Chang et al. [12] developed a LSTM network RUL prediction algorithm that is based on multi-layer grid search (MLGS) optimization, which integrates feature data and optimizes network parameters to ensure accuracy and effectively predict the non-stationary degradation of the bearing. A new deep learning framework called MSWR-LRCN for predicting the RUL of rolling bearings is presented by Chen et al. [13]. The framework incorporates an attention mechanism, a dual-path long-term recurrent convolutional network, and polynomial fitting to improve the RUL prediction accuracy. The DSCN proposed by Wang et al. [14] directly takes monitoring data acquired by different sensors as inputs, automatically learns high-level representations through separable convolutional building blocks, and estimates RUL through a fully-connected output layer. A hybrid approach based on deep order-wavelet convolutional variational autoencoder and a gray wolf optimizer for RUL prediction is proposed by Yan et al. [15].

In this research, a deep learning approach including multiple separable convolutional neural networks (CNNs), a bidirectional long short-term memory (Bi-LSTM) and fully connected layers (FCL) is adopted to accurately and efficiently estimate the remaining useful life (RUL) and enable more effective predictive maintenance planning. The separable CNN layers are deployed to learn non-linear and complex dependencies from raw distortion data, while the Bi-LSTM layer is used to capture long-short temporal features. Moreover, the dropout method and L2 regularization are used in the training phase of the proposed deep learning approach to achieve more accurate learning. The performance of the proposed approach is validated on the popular FEMTO Bearing dataset provided by NASA.

The organization of the research is as follows. Section 2 describes the technical background of the proposed approach for RUL prediction of machinery equipment via deep learning approach. In Sect. 3, the experimental setting and results are presented. Lastly, the conclusion is drawn in Sect. 4.

2 Materials and Method

In the proposed deep learning-based approach, separable CNN networks, Bi-LSTM, attention mechanism and full-connected layers are used to learn spatial-temporal and complex features in historical data of deterioration progressions. Detailed information about the deep learning model is presented below.

2.1 Convolution Neural Network (CNN)

Convolutional Neural Network (CNN) is a class of deep learning algorithms widely used to learn highly representative features from multi-sensor data. Recently, it has been used in image recognition, natural language processing, signal processing, and object detection [16]. The architecture of the CNN consists of multiple layers, including convolutional and pooling layers. The convolutional layers extract features from the input data by convolving filters over it, while the pooling layers reduce the features and reduce the number of parameters [17]. The convolution process is formulated as follows:

$$f_i = \delta(w_f \otimes x_i + b_f) \tag{1}$$

In the above formula, f_i stands for the features extracted by CNN, w_f for the kernel weights, b_f for the bias parameters and δ for the activation function. In addition, the operator $\otimes$ covers the retrieval process.

2.2 Separable Convolution Neural Network

Deeply separable convolution, also called separable convolution, aims to efficiently extract temporal and cross-channel relationships from different sensor data. Deeply separable convolutions have been widely applied in different fields because they reduce the computation time and the number of network parameters and avoid unnecessary learning correlations [18]. Unlike the traditional convolutional network, the depth separable convolution consists of two parts, including depth convolution and point convolution, as shown in Fig. 1. After deep convolution, the number of input channels remains the same [19].

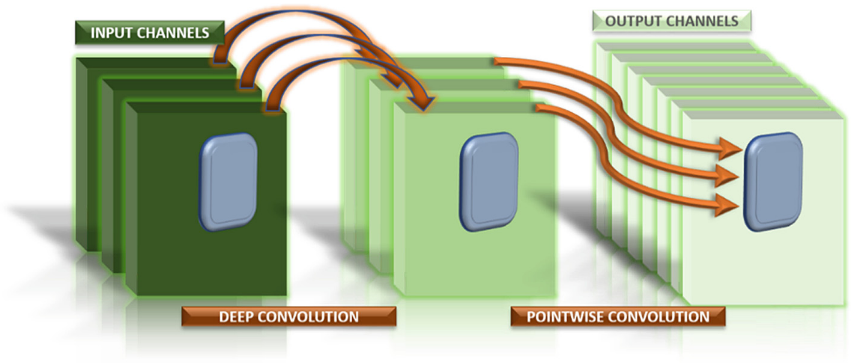


Fig. 1. Separable convolution network.

2.3 Bidirectional LSTM

LSTM uses a memory cell, an input gate, an output gate and a forget gate to control the flow of information through the network. The memory cell allows the network to selectively remember or forget information, while the gates help regulate the flow of information. In this study, unlike the traditional LSTM, bidirectional LSTM is used. Traditional LSTM can only utilize previous data for sequential input data. In other words, no future data of the sequential data are taken into account in the estimation of the current state. Bidirectional LSTM, on the other hand, utilizes the previous and future state of time series data simultaneously [20]. The final output of the network is obtained by combining the two hidden layers and can be calculated as follows:

$$\overrightarrow{h}_t = \tau(b_{\overrightarrow{h}} + \boldsymbol{W}_{\overrightarrow{\boldsymbol{h}},\overrightarrow{\boldsymbol{h}}} \cdot \overrightarrow{h}_{t-1} + \boldsymbol{W}_{\overrightarrow{\boldsymbol{h}},\boldsymbol{x}} \cdot x_t) \tag{2}$$

$$\overleftarrow{h}_t = \tau(b_{\overleftarrow{h}} + \boldsymbol{W}_{\overleftarrow{\boldsymbol{h}},\overleftarrow{\boldsymbol{h}}} \cdot \overleftarrow{h}_{t+1} + \boldsymbol{W}_{\overleftarrow{\boldsymbol{h}},\boldsymbol{x}} \cdot x_t) \tag{3}$$

$$h_t = \mathrm{g}(\boldsymbol{W}_{\boldsymbol{h},\overleftarrow{\boldsymbol{h}}} \cdot \overleftarrow{h}_t + \boldsymbol{W}_{\boldsymbol{h},\overrightarrow{\boldsymbol{h}}} \cdot \overrightarrow{h}_t) \tag{4}$$

In the above equations, $\overrightarrow{h}_t$ and $\overleftarrow{h}_t$ represent the state information in the forward and backward layers, respectively. The operator $\tau(\cdot)$ represents the LSTM processing steps, while $\mathrm{g}(\cdot)$ is the activation function.

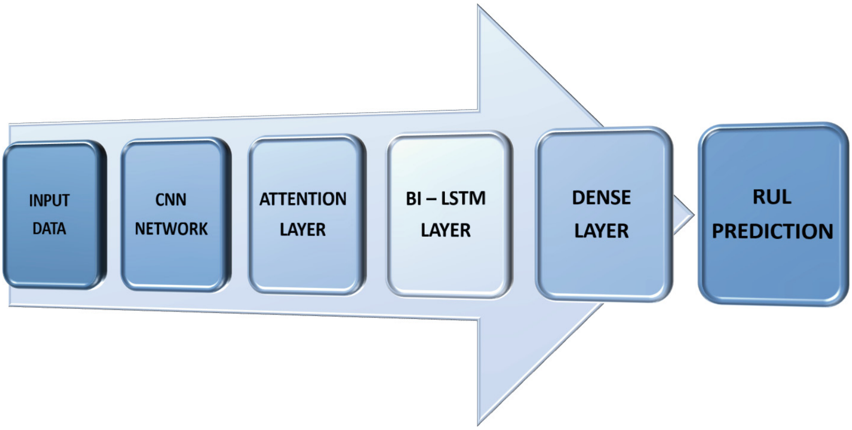


Fig. 2. General structure of the proposed approach.

The input data of the proposed approach is two-dimensional, $w_t \times f_t$. w_t, represents the time windows in the input data and f_t represents the predetermined number of features. The input data is first sent to two separable CNN networks with different kernel sizes. Through this process, it is planned to learn the complex and non-linear features in the input data. Then, the output of the CNN networks will be used by the self-attention mechanism. From the complex features and discriminative information to be obtained by CNN networks and attention mechanism, temporal dependencies will be extracted by Bi-LSTM network. Finally, the extracted features will be used by full-layer networks to predict the remaining lifetimes. Figure 2 shows the general structure of the proposed deep learning-based approach.

3 Experimental Setting and Results

3.1 Dataset

In this paper, we consider the FEMTO bearing dataset, which is widely used in the literature to predict the remaining life of machines and to evaluate the effectiveness of the proposed deep learning approach. The FEMTO dataset was collected by the PRONOSTIA test rig and made publicly available for the IEEE PHM 2012 prognostics competition [21]. The test rig consists mainly of an induction motor, a shaft, a speed controller, an assembly of two rollers and tested bearings. PRONOSTIA provides accelerated degradation of the bearings under three different operating conditions, and a total of up to seventeen failure operating datasets are provided, six training datasets and eleven test datasets.

3.2 Experimental Setting

The presented bearing RUL prediction approach deployed two separable CNN layers with kernel sizes of 5 and 3 as the first network component. The filter sizes of separable CNNs are set to 16 and 32, respectively. In addition, a Bi-LSTM with 16 units and two fully-connected layers with 32 and 1 units are used to accurately predict the bearing RUL. The dot product attention layer is adopted as the attention mechanism in the framework. To reduce the overfitting problem, a dropout method with a rate of 0.3 and an L2 regularization technique with a rate of 1e-4 are implemented. Mean Square Error (MSE) is handled as the loss function in this framework. The loss function minimization uses the Adam algorithm with a learning rate of 0.001.

A DNN with two fully-connected layers are used to verify the RUL prediction performance of the proposed method. Root mean square error (RMSE) is adopted as the assessment criterion. The experiments are carried out by means of Python v3.8.5 and TensorFlow v2.2.0.

3.3 Results

This section analysis the results of the proposed bearing RUL prediction approach based on deep learning by comparing with DNN benchmark. As an initial evaluation, the training loss curves of the proposed approach and DNN method are illustrated in Fig. 3. Considering the starting epochs, it is seen that the training loss of each model in the last epochs is at a low level. Therefore, the proposed framework and DNN produced less training loss at the end of the training period.

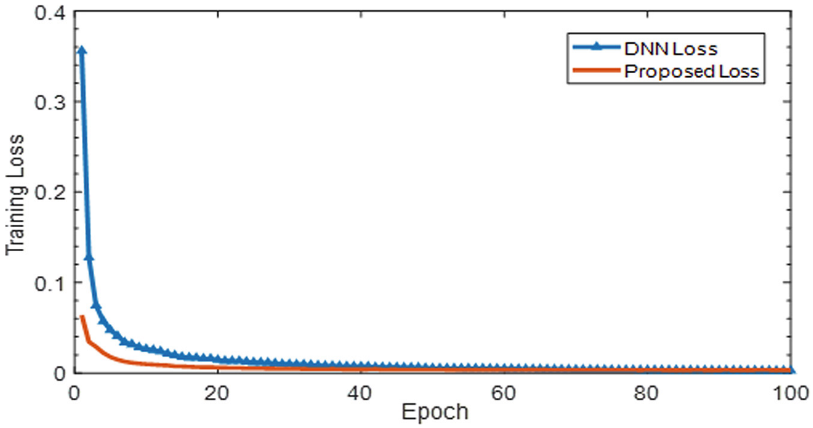


Fig. 3. Training loss curve derived by various techniques.

In this study, the effect of various time window sizes on the prediction accuracy of the proposed approach has been analyzed. In order to predict RUL of the bearings, the time window size is adjusted to 8, 16, and 32, respectively. Correspondingly, the box plots of the MAE score of the Bearing1_3 are demonstrated in Fig. 4. From this box plot, it is observed that MAE score of the proposed method at 16 gives better results compared with the different time window sizes. It was seen that both the mean and the variability of the MAE value were lower. The time window size in RUL prediction of bearings is set to 16 based on this result.

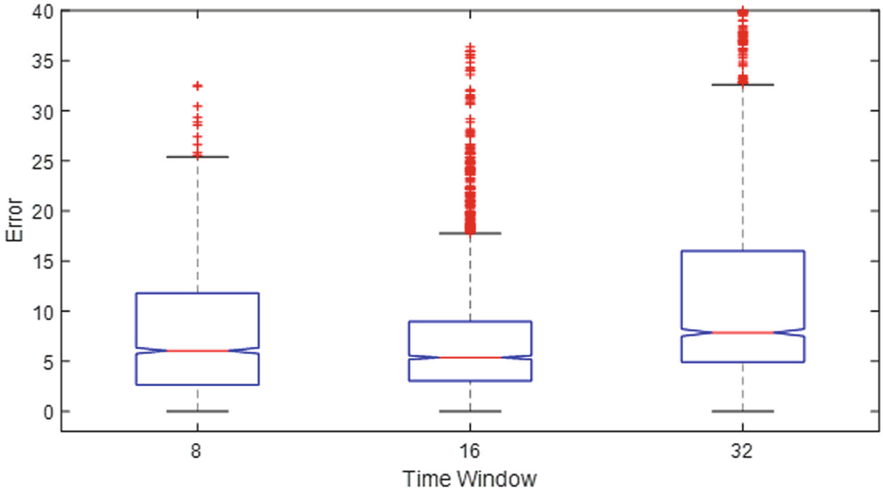
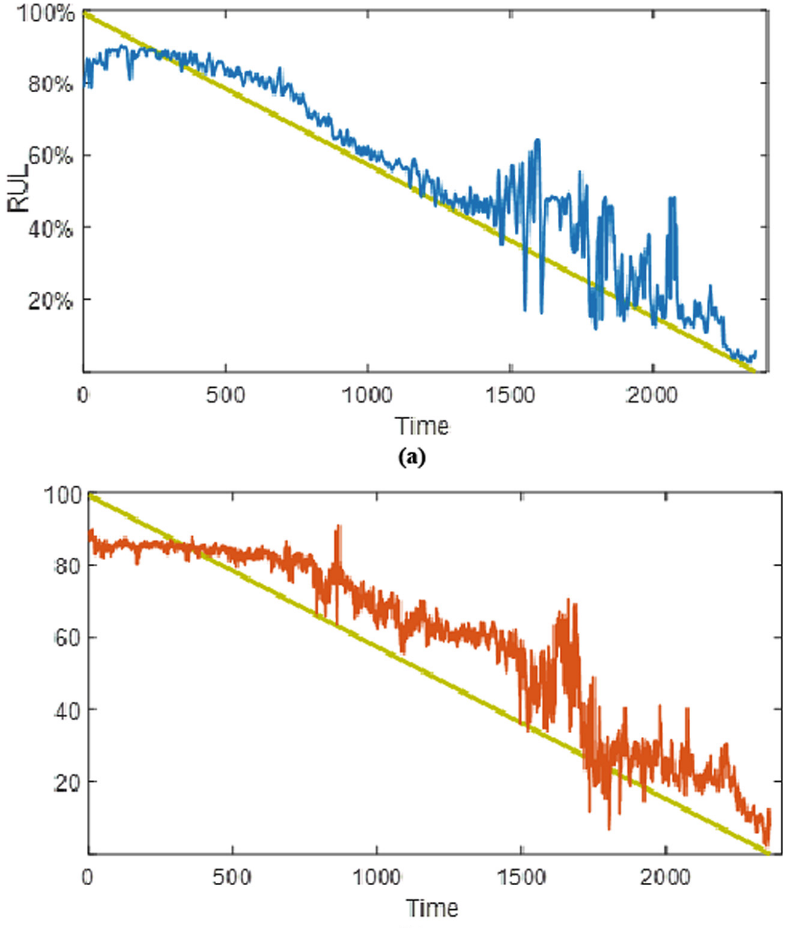


Fig. 4. Box plot of MAE scores under various time window sizes.

In Figs. 5(a) and (b), the RUL prediction results of the proposed and DNN approaches are compared with the actual RUL values of the Bearing1_3. In Fig. 5(a), it can be stated that, in spite of the local variations, the general pattern of degradation of the bearings can be represented by the proposed method. Moreover, compared to DNN method, the predictions of the proposed approach are very close to the actual values. On the other hand, it can be seen in both graphs that there is an increase in fluctuations towards the end of the time series.



(a)

(b)

Fig. 5. RUL prediction results of the different methods.

Furthermore, Table 1 reported the comparison of results obtained by the proposed and DNN methods in terms of RMSE and MAE scores. According to these RMSE and MAE values, the proposed method provides more effective prediction performance in the Bearing1_3, Bearing2_7, and Bearing3_3 compared DNN method. For other bearings, DNN is better. In general, the proposed framework for RUL prediction is able to capture the degradation behavior of the bearings, but an effective hyper-parameter tuning is needed for better results.

Table 1. Comparison of the prediction errors of different methods.

Testing Bearing	Proposed		DNN	
	RMSE	MAE	RMSE	MAE
Bearing1_3	9.61	7.28	12.34	10.72
Bearing2_3	41.18	33.96	39.81	34.4
Bearing2_5	46.85	37.17	42.58	33.96
Bearing2_7	16.9	13.3	18.05	14.86
Bearing3_3	13.35	10.58	14.45	11.22

4 Conclusion

In this research, with the aim of prediction RUL using FEMTO bearing dataset, a hybrid approach based on deep learning has been introduced. To extract the effective patterns from the raw degradation data, the introduced framework consists of the combination of two separable CNN layers, a Bi-LSTM layer and the fully connected layers. Comparisons with the DNN model, was performed to evaluate the effectiveness of the proposed approach. Taking into account the experimental results, although the presented approach gives remarkable results for bearing prognostics, hyperparameter tuning with a meta-heuristic algorithm is required for more effective results.

References

1. Wei, Y., Wu, D., Terpenny, J.: Bearing remaining useful life prediction using self-adaptive graph convolutional networks with self-attention mechanism. Mech. Syst. Signal Process **188**, 110010 (2023)
2. Ouadah, A., Zemmouchi-Ghomari, L., Salhi, N.: Selecting an appropriate supervised machine learning algorithm for predictive maintenance. Int. J. Adv. Manuf. Technol. **119**(7–8), 4277–4301 (2022)
3. Shen, Y., Tang, B., Li, B., Tan, Q., Wu, Y.: Remaining useful life prediction of rolling bearing based on multi-head attention embedded Bi-LSTM network. Measurement **202**, 111803 (2022)
4. Ahmad, W., Khan, S.A., Islam, M.M.M., Kim, J.M.: A reliable technique for remaining useful life estimation of rolling element bearings using dynamic regression models. Reliab. Eng. Syst. Saf. **184**, 67–76 (2019)
5. Rathore, M.S., Harsha, S.P.: An attention-based stacked BiLSTM framework for predicting remaining useful life of rolling bearings. Appl. Soft. Comput. **131**, 109765 (2022)
6. Jiang, J.R., Lee, J.E., Zeng, Y.M.: Time series multiple channel convolutional neural network with attention-based long short-term memory for predicting bearing remaining useful life. Sensors **20**(1), 166 (2019)
7. Ren, L., Sun, Y., Wang, H., Zhang, L.: Prediction of bearing remaining useful life with deep convolution neural network. IEEE Access **6**, 13041–13049 (2018)
8. Yang, J., Peng, Y., Xie, J., Wang, P.: Remaining useful life prediction method for bearings based on LSTM with uncertainty quantification. Sensors **22**(12), 4549 (2022)

9. Gupta, M., Wadhvani, R., Rasool, A.: A real-time adaptive model for bearing fault classification and remaining useful life estimation using deep neural network. Knowl. Based Syst. **259**, 110070 (2023)
10. Xu, Z., et al.: A novel health indicator for intelligent prediction of rolling bearing remaining useful life based on unsupervised learning model. Comput. Ind. Eng. **176**, 108999 (2023)
11. Sun, B., Liu, X., Wang, J., Wei, X., Yuan, H., Dai, H.: Short-term performance degradation prediction of a commercial vehicle fuel cell system based on CNN and LSTM hybrid neural network. Int. J. Hydrogen Energy **48**(23), 8613–8628 (2023)
12. Chang, Z.H., Yuan, W., Huang, K.: Remaining useful life prediction for rolling bearings using multi-layer grid search and LSTM. Comput. Electr. Eng. **101**, 108083 (2022)
13. Chen, Y., Zhang, D., Zhang, W.: MSWR-LRCN: a new deep learning approach to remaining useful life estimation of bearings. Control Eng. Pract. **118**, 104969 (2022)
14. Wang, B., Lei, Y., Li, N., Yan, T.: Deep separable convolutional network for remaining useful life prediction of machinery. Mech. Syst. Signal Process **134**, 106330 (2019)
15. Yan, X., She, D., Xu, Y.: Deep order-wavelet convolutional variational autoencoder for fault identification of rolling bearing under fluctuating speed conditions. Expert Syst. Appl. **216**, 119479 (2023)
16. Hammad, M., Pławiak, P., Wang, K., Acharya, U.R.: ResNet-Attention model for human authentication using ECG signals. Expert Syst. **38**(6), e12547 (2021)
17. Yu, J., Zhang, C., Wang, S.: Multichannel one-dimensional convolutional neural network-based feature learning for fault diagnosis of industrial processes. Neural Comput. Appl. **33**(8), 3085–3104 (2021)
18. Shang, R., He, J., Wang, J., Xu, K., Jiao, L., Stolkin, R.: Dense connection and depthwise separable convolution-based CNN for polarimetric SAR image classification. Knowl. Based Syst. **194**, 105542 (2020)
19. Huang, G., Zhang, Y., Ou, J.: Transfer remaining useful life estimation of bearing using depth-wise separable convolution recurrent network. Measurement **176**, 109090 (2021)
20. Dong, S., Xiao, J., Hu, X., Fang, N., Liu, L., Yao, J.: Deep transfer learning based on Bi-LSTM and attention for remaining useful life prediction of rolling bearing. Reliab. Eng. Syst. Saf. **230**, 108914 (2023)
21. Nectoux, P., et al.: PRONOSTIA: an experimental platform for bearings accelerated degradation tests. In: IEEE International Conference on Prognostics and Health Management, PHM 2012, pp. 1–8 (2012)