IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 September 2023, accepted 2 October 2023, date of publication 9 October 2023, date of current version 16 October 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3322733

RESEARCH ARTICLE

Remaining Useful Life Prediction via a Data-Driven Deep Learning Fusion Model–CALAP

MINGYAN WU[1], (Student Member, IEEE), QING YE[1], JIANXIN MU[2], (Student Member, IEEE), ZUYU FU[2], (Student Member, IEEE), AND YILIN HAN[3]

[1]School of Computer Science, Yangtze University, Jinzhou 434023, China
[2]School of Mechanical Engineering, Yangtze University, Jinzhou 434023, China
[3]School of Electronic Information, Yangtze University, Jinzhou 434023, China

Corresponding author: Qing Ye (yeqing@yangtzeu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 62006028, and in part by the Natural Science Foundation of Hubei Province of China under Grant 2023AFB909.

ABSTRACT As one of the key technologies in the field of Prognostic and Health Management (PHM), Remaining Useful Life (RUL) prediction technology plays an important role in equipment health maintenance and fault detection. For complex devices, the degradation process of the remaining useful life of the device is often difficult to be described with mathematical or physical models, so data-driven methods has become an important and feasible method in the field of RUL prediction. This paper proposes a data-driven deep learning fusion model based on CNN-ATTENTION-LSTM-ATTENTION-PARALLEL (CALAP) model, in which Convolutional Neural Networks (CNN) and Long Short-Term Memory Networks (LSTM) are adopted to extract the spatial features and temporal features of the data in parallel, and both of them combine the corresponding attention mechanism allowing the network to focus on important factors. The CNN path fuses CBAM and the LSTM path fuses attention mechanism. We evaluate the proposed model on the C-MAPSS dataset released by NASA and compare it to the state-of-the-art. The RMSE of the proposed model is reduced by nearly 2-5%, and the score is reduced by nearly 8-10% under simple conditions. Experimental results prove that the model has relatively high prediction accuracy and good robustness.

INDEX TERMS RUL prediction, convolutional neural networks (CNN), convolutional block attention module (CBAM), long short-term memory (LSTM), attention mechanism, turbofan engine.

I. INTRODUCTION

Prognostic and Health Management (PHM) refers to the use of sensor technology to obtain equipment operation data and information, and to evaluate, predict, and manage the health status of equipment with the help of mathmatical algorithms and artificial intelligence technology. PHM can provide equipment maintenance plans based on the prediction results to prevent potential failures from having a major impact on equipment later, thereby reducing human and financial losses and providing a safe production environment, also make machinery and equipment run more safely and reliably. Remaining Useful Life (RUL) prediction technology is an indispensable part of PHM. RUL can be defined as the number of operational cycles from the current moment to the moment of machine failure. The prediction of equipment RUL can be realized by analyzing the operation data of the equipment or establishing a suitable equipment degradation model.

There are two main types of RUL prediction techniques [1], one is a model-based method, and the other is a data-driven method. However, in general, it is difficult to establish an accurate model to describe the degradation process of equipment. Data-driven approaches have become increasingly popular thanks to recent advances in sensor systems and big data technologies [2], [30], [35]. Compared with model-based RUL prediction technology, the data-driven method relies less on prior knowledge [31]. Through data analysis and

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

feature extraction of the monitoring data collected by the system or components, the implicit information of the data can be obtained, which provide more useful information for the prediction of RUL. So data-driven methods play an important role in the current RUL prediction technology [34].

In industrial activities, mastering the health status and the RUL of equipment is conducive to timely updating the maintenance plan to ensure the orderly progress of production activities [22], while reducing maintenance costs and time costs, and recovering certain economic and personnel losses. **Overall, the main contributions of the article are as follows:**

1) **At present, most data-driven prediction models use one or a single variant of neural network for feature extraction, which is difficult to maintain good generalization ability in complex prediction environments. CALAP combines the advantages and disadvantages of the current mainstream deep learning models and popular algorithms, allowing CNN and LSTM to play their respective data feature extraction functions to improve the model's RUL prediction accuracy. The fusion model can adapt to different data distributions and different data characteristics, and is more capable of processing outliers and noise data, and has higher generalization ability.**
2) **We add CBAM to the CNN path and add attention mechanism to the LSTM path, so that the network can adaptively adjust the weight according to the importance of the input data, and focus on important factors, thereby improving the accuracy of the model, reducing the amount of calculation, and improving the efficiency of calculation.**
3) **In the experimental section, we provide detailed and sufficient ablation experiments to verify the effectiveness of the proposed fusion model.**

II. LITERATURE REVIEW

In this section, we review the mainstream deep learning prediction models in recent years.

Fully Connected Neural (FCN) Networks is an artificial neural network with good data fitting ability. Gebraeel et al. [3] developed neural-network-based models in bearing failures prediction, and used the vibration information obtained from a large number of bearings from operation to failure to train the model to predict the bearing operating time. Tian [4] introduced a validation mechanism in training process to improve the prediction acurracy. Because the FCN model has a weak ability to extract spatial and temporal features, FCN is rarely used as a prediction model alone, and is usually combined with other deep learning networks for prediction. Badu et al. [6] first applied Convolutional Neural Networks (CNN) to RUL estimation in the field of prediction. Convolution filters and pooling filters are applied along the time dimension to muti-channel sensor data for automatic feature learning from raw sensor signals. Ren et al. [7] proposed deep convolution neural network for the prediction for bearing RUL and used spectrum-principal-energy-vector in feature extraction process. Yoo et al. [8] proposed a new time-frequency image feature to construct HI and predicted the RUL. Huang et al. [38] developed a deep convolutional neural network-multilayer perceptron to extract the one-dimensional temporal features and two-dimensional spatial features of bearings, enabling the quantification of the prediction interval of RUL. Recurrent Neural Networks (RNN) is also widely used in the field of RUL prediction due to its good sequential modeling ability. In industrial production, the data obtained by sensors is almost all time-series data. It is not surprising that RNN is applied in RUL prediction. Heimes [9] utilized the recurrent neural network model using the EKF algorithm to estimate the RUL of the system and produced relatively high accuracy. As a variant of the RNN model, Long Short-Term Memory (LSTM) solves the problem of gradient disappearance and can effectively process long sequences. Zheng et al. [10] applied LSTM to RUL estimation and pointed out that LSTM outperforms traditional methods as well as CNN in predicting RUL. Wu et al. [11] used dynamic differential for inter-frame information extraction and proposed vanilla LSTM to get good RUL prediction accuracy. Deng et al. [37] proposed a hybrid GRU-PF method, combining data-driven and physics-based models into the PF framework to quantify the uncertainty of the RUL of the ball screw. Kong et al. [48] and Chen et al. [49] combined LSTM with other network models to predict the remaining useful life of turbofan engines.

In recent years, Autoencoder (AE) has been favored by researchers in the field of RUL prediction because of its powerful deep feature expression ability [12], [45], [46]. Ren et al. [50] combined deep autoencoders and deep neural networks (DNN) for RUL prediction without increasing DNN size. SAE is formed by stacking multiple AEs, and is trained by an unsupervised greedy pre-training method. Ma et al. [13] presented a deep approach for RUL prediction based on SAE and logistic regression. Performance degradation features from multiple sensors were extracted by stacked sparse autoencoder and multiple features were fused by multilayer self-learning. Focusing on high-precision RUL prediction for reliability, Wei et al. [59] proposed a SAE-GMR-based RUL prediction framework. Ye et al. [60] proposed a deep learning model based on parallel convolutional autoencoder (PCAE) and applied it to the feature generation phase of the diagnostic framework. Deep Belief Networks (DBN) is derived from a fast greedy algorithm proposed by Hinton et al. in [15], which can learn a layer of deep directed belief network, provided that the top two layers form an undirected associative memory. Zhang et al. [16] proposed a multiobjective deep belief networks ensemble (MODBN) method. MODBNE uses a multi-objective evolutionary algorithm combined with traditional DBN training techniques to develop multiple DBNs while subjecting accuracy and diversity as two conflicting goals. Moreover, the introduction of the optimization algorithm can also improve the prediction accuracy of DBN. Li et al. [17] presented a method based on the bispectrum

TABLE 1. The main advantages and disadvantages of each representative method.

	representative works	advantages	disadvantages
Convolutional Neural Networks	[6] [7] [8] [38]	Good at processing image data and mining spatial features	No memoization, cannot handle sequence data
Recurrent Neural Networks	[9] [10] [11] [37] [48] [49]	Good at processing sequence data and can maintain dependencies in sequence data very well	Long training time
Autoencoder	[13] [50] [59] [60]	Reduce data dimension, can be used for data denoising	There may be information loss problem
Deep Belief Networks	[15] [16] [17]	Can handle high-dimensional data and can be used for unsupervised pre-training	The training time is long, and there may be a local optimal solution problem
Transfer learning	[14] [36] [51-58]	Strong generalization ability, effective use of existing knowledge, adaptable to different tasks and data	There may be problems such as negative transfer, domain differences, and difficulties in knowledge transfer
Fault prediction under limited data	[39] [40] [41] [61]	Not dependent on large amounts of data	Due to the insufficient amount of data, it is difficult to comprehensively mine the information between the data

entropy and DBN, and introduced the Quantum Parricle Swarm Optimization (QPSO) for searching the optimal value of initial parameters of the network.

How to transplant the trained network to new target objects is a problem that researchers are concerned about. Transfer learning is also more and more important in the field of RUL prediction [51], [52], [53], [54]. Ding et al. [36] proposed a method combining deep metric learning and transfer learning to solve regression problems. Sun et al. [14] proposed a deep tranfer learning (DTL) model based on SAE, which can transplant the trained model to the new prediction object. The SAE network is first trained by running on fault data, and then the trained network is transferred to a new tool for online RUL prediction. Qian et al. proposed to use deep discriminative transfer learning network [57], relationship transfer domain generalization network [56], ensemble weighting subdomain adaptation network [58], and maximum mean square discrepancy [55] for fault prediction, and verified the effectiveness of the proposed method through relevant experiments, which promoted the development of transfer learning-based methods in prediction models. The problem of RUL prediction under limited data is also paid attention by researchers. Kong et al. [61] developed a first prediction time (FPT) identification method, combined with degeneracy factors and SBiLSTM, and proposed a multi-step ahead rolling prediction method to predict RUL. Li et al. [39] used first prediction time identification and time-series feature windows to predict the RUL of rolling bearings under limited data. Zhu et al. [40] proposed a Bayesian semi-supervised transfer learning intelligent fault prediction framework based on active query, which can perform RUL prediction across machines under limited data. Zhou et al. [41] made effective use of historical data and proposed a time series prediction method based on dynamic transfer learning under limited data.

In TABLE 1, we summarise the main strengths and weaknesses of current representative deep learning prediction models and cite the corresponding representative work. As can be seen form the references mentioned above, most of the data-driven predictions are one or a single variant neural network model. The fusion model can combine the advantages of different networks, make up for the shortcomings, and adapt to complex systems and changing working conditions. The RUL prediction model of this development strategy has become a new research direction. This paper aims to fill the gap in this direction and proposes a data-driven deep learning fusion model (CALAP), namely the CNN-Attention-LSTM-Attention-Parallel algorithm model. CALAP extracts the spatial features of the data through the CNN path, and uses the LSTM path to extract the temporal features of the data. Both paths are combined in parallel. The convolutional layer in CNN can perform dimensionality reduction processing on high-dimensional input data, and has the advantage of automatically extracting effective features. The pooling layer samples data of each dimension, which can further reduce the data scale and improve the generalization ability of the network. LSTM can learn long-term dependencies in time series data, and make full use of the characteristics of sensor signal time series. At the same time, we embed CBAM in the CNN path, and embed the attention mechanism in the LSTM path, so that the network can focus on the key features in the data, give greater weight to important features, and make up for the shortcomings caused by manual feature extraction. In summary, the CALAP model makes up for the lack of CNN's feature extraction ability in temporal data and the lack of LSTM's feature extraction in spatial data, while the introduction of the two attentional mechanisms also reduces the amount of computation and improves the computational efficiency. Experimental studies show that the fusion model

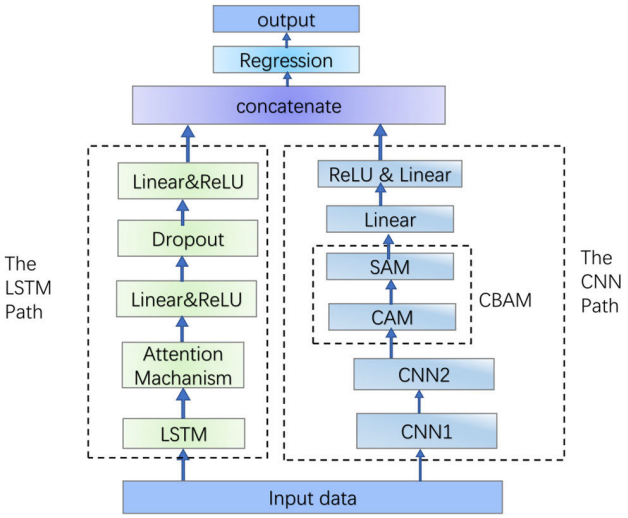

FIGURE 1. The overall architecture of the CALAP model.

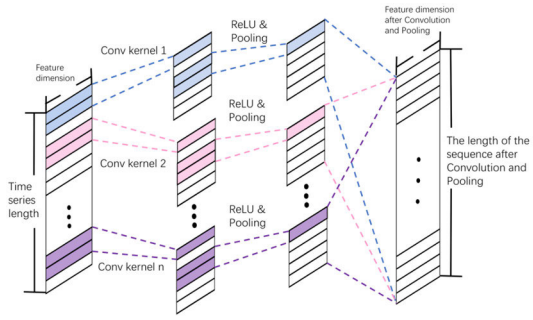

FIGURE 2. One of the one-dimensional neural networks of the CALAP model.

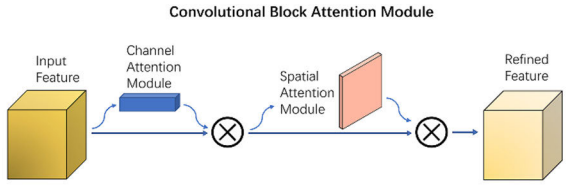

FIGURE 3. The overall architecture of CBAM.

has better prediction accuracy and better robustness than the single model.

III. METHODOLOGY

A. PROPOSED NETWORK ARCHITECTURE

In this section, we introduce the architecture of CALAP model (as shown in **FIGURE 1**). CALAP is mainly developed based on two important deep neural networks. Deep neural network has a great advantage in the era of big data because of its ability to mine useful information from massive data and its powful nonlinear data fitting ability.

The proposed CALAP model consists of two main paths, which are developed based on two deep neural networks, i.e., CNN and LSTM, respectively. The CNN is responsible for extracting the spatial features of the data, and the LSTM is responsible for extracting the temporal features of the data. At the same time, the two are combined with the corresponding attention mechanism to allow the network to focus on important factors. The CNN path fuses CBAM and the LSTM path fuses attention mechanism. The two paths extract features in parallel, and the extracted features are concatenated as predicted features. We will describe the important components of the CNN path and the LSTM path in detail below.

B. THE CNN PATH

This path mainly includes two one-dimensional CNN layers and one CBAM layer.

1) CONVOLUTIONAL NEURAL NETWORKS (CNN)

CNN is a type of feedforward neural network that includes convolution calculations and has a deep structure. CNN is proposed by the inspiration of biological receptive field mechanism. In recent years, many studies have successfully applied CNN to sequence data processing, such as RUL prediction, natural language processing, etc. The convolutional layer and the pooling layer are two important components of CNN. The special structure of CNN enables it to effectively extract the spatial features of the input data [42], [43], [44]. The convolution layer uses the convolution filter to convolve the input data to obtain the feature map, so as to achieve the effect of data dimensionality reduction. The pooling layer can further reduce the data size and improve the generalization ability of the network.

In this paper, the data to be processed is usually the sequence features from the sensor, and the spatial features in the time series data are not as obvious as the spatial features of the image data. Therefore, the CALAP model uses one-dimensional convolutional neural network. **FIGURE 2** shows one of the one-dimensional neural networks of the CALAP model.

2) CONVOLUTIONAL BLOCK ATTENTION MODULE (CBAM)

CBAM was first proposed by Woo et al. [19] and it is a lightweight attention module. CBAM can be embedded into any convolutional network with almost no overhead. **FIGURE 3** shows the overall architecture of CBAM. It contains two continous sub-modules, namely Channel Attention Module (CAM) and Spatial Attention Module (SAM), which act on the channel and space respectively.

a: CHANNEL ATTENTION MODULE (CAM)

As shown in **FIGURE 4**, the input feature map passes through the MaxPool layer and AvgPool layer in parallel, and the size changes from 128*H*W to 128*1*1. In Shared MLP module, it first reduces the number of channels to the original 1/r times (r means reduction ratio). The r in this experiment is 2. During this period, a ReLU activation is required, and then it is restored to the original number of channels. The two results are added to the corresponding elements, and activated by sigmoid to obtain the output result of CAM.

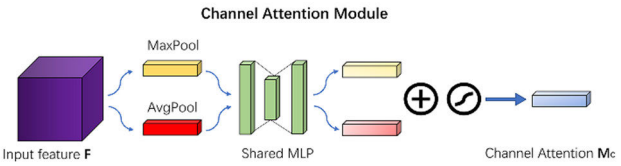

FIGURE 4. The structure of CAM.

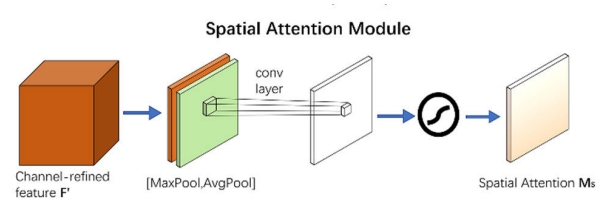

FIGURE 5. The structure of SAM.

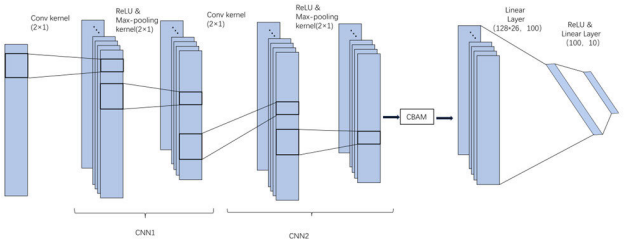

FIGURE 6. The specific structure of the CNN path in CALAP.

It can be computed as:

$$M_c(F) = \sigma\left(MLP\left(AvgPool\left(F\right)\right) + MLP\left(MaxPool\left(F\right)\right)\right) \tag{1}$$

b: SPATIAL ATTENTION MODULE (SAM)

The output result of CAM is restored to 128∗H∗W size by multiplying with the source data (as shown in **FIGURE 3**) and the result is call channel-refined feature. As shown in **FIGURE 5,** the channel-refined feature obtains two 1∗H∗W feature maps through MaxPool and AvgPool, then concatenated and convolved by a convolution layer. Finally, the output result of SAM is obtained through a sigmoid function.

It can be computed as:

$$M_s(F) = \sigma\left(f^{7\times7}\left(\left[AvgPool\left(F\right); MaxPool\left(F\right)\right]\right)\right) \tag{2}$$

where $f^{7\times7}$ represents a convolution operation with the filter size of 7×7 [19].

Finally, the specific structure of the CNN path is shown in **FIGURE 6**.

C. THE LSTM PATH

This path mainly includes one LSTM layer and one attention mechanism layer.

1) LONG SHORT-TERM MEMORY (LSTM)

The input of the feedforward network is independent of each other, and the current input has nothing to do with the past or the future. However, in some tasks, such as RUL prediction, it often occurs that the input of the model is not only related to the current input, but also related to the past states. This requires more powerful models for processing sequence data. Therefore, the LSTM network with the ability to capture long-term data relationships has a very good application in RUL prediction. Hochreiter et al. [18] proposed a new architecture called LSTM, which solved the problem of gradient disappearance in RNN through the design of memory information flow and gate structure.

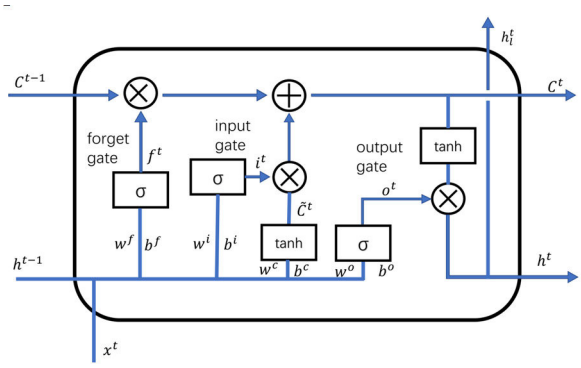

FIGURE 7. The cell structure of LSTM.

FIGURE 7 shows the cell structure of LSTM. It can be expressed as:

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{3}$$
$$i_t = \sigma\left(W_i \cdot [h_{t-1}, x_t] + b_i\right) \tag{4}$$
$$\tilde{C}_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{5}$$
$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{6}$$
$$o_t = \sigma\left(W_o\left[h_{t-1}, x_t\right] + b_o\right) \tag{7}$$
$$h_t = o_t * \tanh(C_t) \tag{8}$$

where x_t is the input at time step t, h_t is the hidden state at time step t, C_{t-1} is the previous cell state, C_t is the updated cell state, W and b are the weights and biases of the corresponding gated cells, and $\sigma()$ and $\tanh()$ are sigmoid and tanh functions.

Although LSTM has good time series data modeling capabilities, it does not distinguish the importance of the learned features. For inputs with too long time steps, some previous information may be lost. Therefore, we design an LSTM that incorporates an attention mechanism, by assigning different weights to different features, so that LSTM can more efficiently use historical information to generate the output of each time step. We'll go into more detail in the next section.

2) ATTENTION MECHANISM

The attention mechanism [24], [25], [26] is inspired by humans. When our vision perceives the scene in front of us, it scans the overall image to obtain key areas of focus, and then devotes more attention. When the neural network introduces the attention mechanism, it can increase the computing power in limited circumstances, focus on key information, reduce attention to other information, and allocate computing resources to important factors.

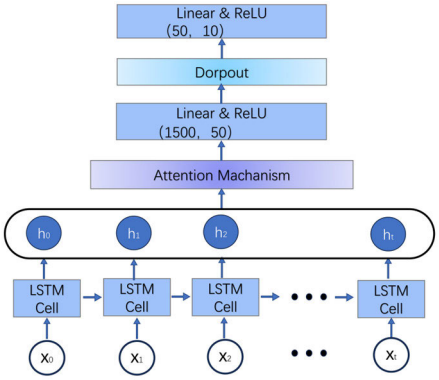

FIGURE 8. The specific structure of the LSTM path in CALAP.

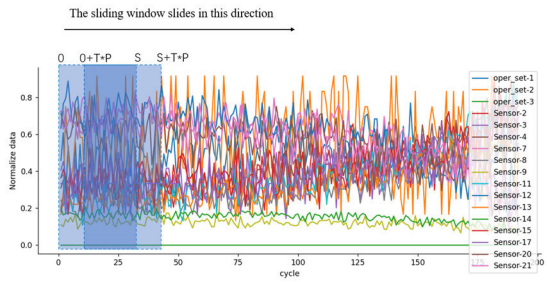

FIGURE 9. The sliding window example.

Assume that the features learned by the LSTM network for one sample can be expressed as $H = \{h_0, h_1, \ldots, h_t\}^T$. If the self-attention mechanism is introduced [21], the attention weights for different time steps t of h_t can be expressed as:

$$Attention = \Phi\left(W^T h_t + b\right) \tag{9}$$

where W and b are the weight matrix and bias vector, respectively, and $\Phi()$ is the activation function.

Then use the softmax function to normalize the attention matrix, and finally perform element-by-element multiplication with the original data to obtain the output of the attention mechanism.

Finally, the specific structure of the LSTM path is shown in **FIGURE 8**.

D. THE SLIDING WINDOW

we use the sliding window method to collect sample data. Reference [5] mentioned this method. Assume that the original data can be expressed as a matrix A_{ij} of size $m \times n$, the size of the sliding window is S, and the sliding step is P. The sliding window starts to collect data from the first row, and the first sample data $A^0_{ij}(i = 0, 1, \ldots, S; j = 0, 1, \ldots, n)$ with a size of $s \times n$ is obtained. Then the sliding window moves the step size P along the direction of the column to get the second sample data $A^1_{ij}(i = 0+P, 1+P, \ldots, S+P; j = 0, 1, \ldots, n)$ with a size of $s \times n$. Similarly, we can get the T-th sample data as $A^T_{ij}(i = 0+T*P, 1+T*P, \ldots, S+T*P; j = 0, 1, \ldots, n)$. **FIGURE 9** shows the sliding process of the sliding window.

TABLE 2. The overview of the dataset.

Data Set		FD001	FD002	FD003	FD004
Train number	engines	100	260	100	249
Test number	engines	100	259	100	248
Conditions		1	6	1	6
Fault modes		HPC Degradation	HPC Degradation	HPC Degradation,Fan Degradation	HPC Degradation,Fan Degradation

IV. EXPERIMENT

All experiments are performed on a PC with Intel Core i7 CPU, 8 GB RAM and GEFORCE GTX 1650 GPU.

A. DATASET INTRODUCTION

Our research uses the turbofan engine degradation simulation data set [28] provided by NASA, which is simulated by commercial modular aviation propulsion system simulation (C-MAPSS) and is a widely used benchmark data [23]. The data has four sub-datasets, namely from FD001 to FD004, and each sub-dataset has different types of operating conditions and fault conditions. The overview of the dataset is shown in TABLE 2.

The dataset consists of multiple multivariate time series. Each sub-dataset is further divided into a training subset and a testing subset [27], in which the training set records the sampling values of each time series of multiple state parameters of the aeroengine from normal to fault in the complete cycle. The test set contains the state parameter values before a certain point in time before the failure. The true RUL corresponding to the test data set is also provided.

Each row is the instantaneous data captured by the sensor in a single cycle, and each column represents a different variable, a total of 26 columns, consisting of engine unit number and cycle number, three operating settings, as well as 21 sensor values. For the meaning of the specific sensor representatives, please refer to [20].

B. DATA PREPROCESSING

In order to better understand the data, we visualized the three operational setting data and 21 sensor data in the dataset. **FIGURE 10** and **FIGURE 11** show the FD001 training data as an example.

We found that sensors 1, 5, 6, 10, 16, 18 and 19 always have constant values while the engine is running, which means that these sensor data have little impact on engine life and do not contribute to engine RUL prediction. So we can delete these sensor-related data, and for the remaining operational setting data and sensor data, we use the following formula to normalize the data. Generally, the features extracted from the original signal have different scales of input values, and normalization is necessary to map the features extracted from

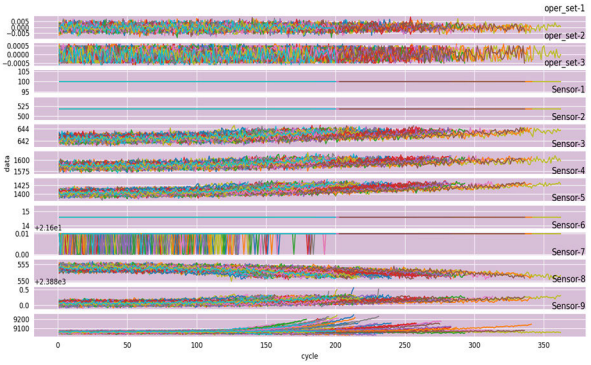

FIGURE 10. Sensor data for all engines (three operational setting data and data from sensor 1 to sensor 9).

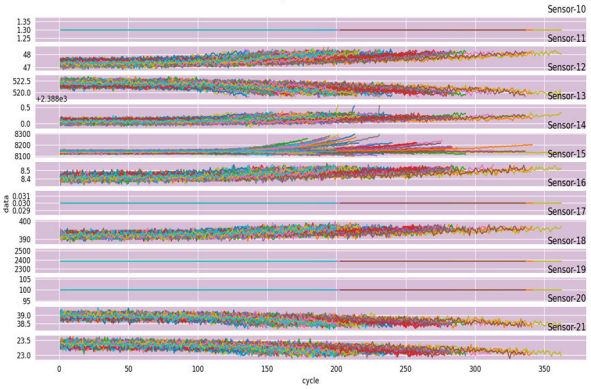

FIGURE 11. Sensor data for all engines (data from sensor 10 to sensor21).

the vibration signal to a specific equal interval [32].

$$x_i' = \frac{x_i - min x_i}{max x_i - min x_i} \tag{10}$$

In [6], [9], and [10], a piece-wise linear RUL target function is proposed, as shown in **FIGURE 12**, which limits the maximum RUL to a constant value and then starts to linearly degrade after a certain degree of usage. Different datasets have different maximum values. Here, the dataset of FD001 is taken as an example. In order to better simulate the change of the remaining useful life along time, we process the turbofan target data in the same way as [6], [9], and [10].

C. TRAINING PARAMETERS

1) THE CNN PATH

As shown in **FIGURE 6,** each CNN layer uses the ReLU activation function for activation and one-dimensional maximum pooling for pooling. The previous CNN layer has a filter of size 2×1, the number of channels produced by the convolution is 64, and the maximum pooling layer has a filter of size 2×1. The filter size of the latter CNN layer and the maximum pooling layer filter are the same as the previous CNN layer, but the number of channels produced by the convolution is 128. After two convolutional layers, we input it to the CBAM layer to let the convolutional network focus on important spatial features. CBAM is a lightweight attention module that can be well embedded into the convolution

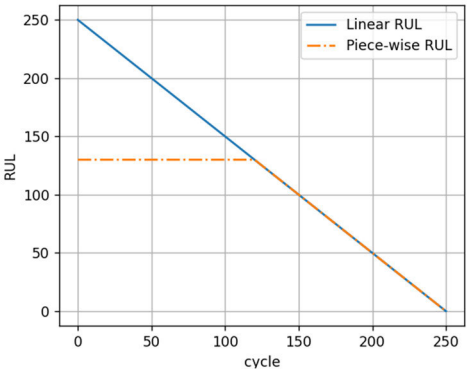

FIGURE 12. Piece-wise RUL of C-MAPSS RUL_FD001 dataset.

network, which we mentioned above. Finally, the output of the CNN path is obtained through two linear layers.

2) THE LSTM PATH

As shown in **FIGURE 8,** the data is first fed into an LSTM layer with a hidden layer size of 50 and a layer count of 1. Then the output of our LSTM is input to the attention machanism layer to let the network pay attention to important time serie temporal features. Finally, the output of the LSTM path is obtained through two linear layers and one dropout layer. The dropout rate of dropout layer is 0.2.

3) MODEL TRAINING PROCESS

For the data that has been preprocessed, we use the sliding window method to collect sample data and then pass the collected sample data into the network model for training. The batch size of the training set is 100, and the batch size of the test set is 64. The time window has a size of 30 [21] and a step size of 1. The Adam algorithm is used for optimization in the training process and its learning rate is 0.001.

4) EVALUATION CRITERIA

To better compare the performance of the proposed model on testing data, we need some objective measure of performance. In this experiment, we use two ways: Scoring Function and Root Mean Square Error (RMSE) [27], [47]. **FIGURE 13** shows the difference between these two functions.

a: SCORING FUNCTION

This scoring function was proposed by Saxena et al. [33] and it can be discribe as follow:

$$S = \begin{cases} \sum_{i=1}^{N} \left(e^{-\frac{h_i}{13}} - 1\right) & for\ h_i < 0 \\ \sum_{i=1}^{N} \left(e^{\frac{h_i}{10}} - 1\right) & for\ h_i \geq 0 \end{cases} \tag{11}$$

where S is the computed score, N is the number of engines in testing data, and $h_i = estimated\ RUL_i - true\ RUL_i$.

The scoring function is more inclined to early prediction, i.e., the estimated RUL value is smaller than the actual RUL value [16], and later prediction may cause equipment to be unmaintained due to too late prediction. This is in line with

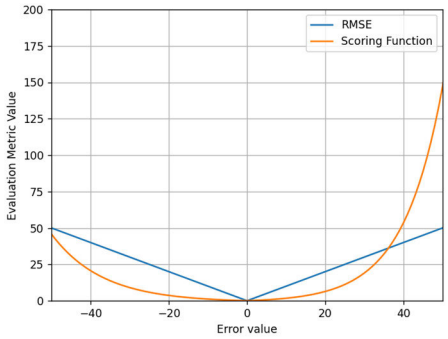

FIGURE 13. The difference between RMSE and scoring function.

TABLE 3. The test RMSE, score and test time for each dataset.

	FD001	FD002	FD003	FD004
RMSE	14.35	26.50	14.43	28.04
score	309.32	42579.42	381.41	12179.41
test time	0.69s	0.61s	0.62s	0.72s

the characteristics of the aerospace industry. The safe operation and maintenance of equipment is extremely important to this industry, and a small failure may cause immeasurable losses. The score function is also the scoring function used in the PHM08 challenge [16].

b: RMSE

RMSE is to give equal weight to early and late predictions [29]. This is not the same as the scoring function. Combining the scoring function and RMSE can more comprehensively evaluate the performance of the model. The formulation of RMSE is as follows:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} h_i^2} \tag{12}$$

D. RESULTS

1) VISUALIZATION OF PREDICTION RESULTS

Based on the above theory and experimental process, we tested the CALAP model. The test set contains FD001-FD004. We train each data set ten times, take the network model with the smallest RMSE each time for efficiency testing, and obtain the average test RMSE, test score and test time of the model, as shown in TABLE 3. **FIGURE 14** shows the prediction effect of the model on the testing data.We can see that for FD001 and FD003, the model has a good prediction effect, and for FD002 and FD004, the prediction effect is relatively reduced. This is influenced by the characteristics of different datasets, which we will analyze in detail in the "COMPARE TO THE STATUS-OF-THE-ART" section.

2) ABLATION STUDY

In order to show the fusion effect of the proposed model more intuitively, we conduct ablation experiments on the proposed model. By disassembling and combining the model, we can get the following five parts:

TABLE 4. The results of the ablation experiment.

Five parts	FD001		FD004	
	RMSE	Score	RMSE	Score
One-dimensional CNN	15.22	371.02	29.61	22152.92
One-dimensional CNN + CBAM	14.62	333.14	29.50	20051.64
LSTM	14.73	328.71	28.83	12141.29
LSTM + AttentionMechanism	14.63	317.12	28.64	10576.23
Proposed model	**14.35**	**309.32**	**28.04**	**12129.41**

1. One-dimensional CNN
2. One-dimensional CNN + CBAM
3. LSTM
4. LSTM + Attention Mechanism
5. Proposed model

Here we show two representative sub-datasets: FD001 and FD004 as examples. The specific results are shown in TABLE 4 and **FIGURE 15**. In addition, we take the FD001 data set as an example, use t-sne technology to visualize the hidden feature extraction of 30 samples in several different network layers, and provide a visualization of one of the sample data, as shown in **FIGURE 16**. It can be seen from the figure that the features of the original sample after dimensionality reduction are messy, and from the visualized data, it can be seen that the importance of each feature is not clearly distinguished, and the importance of the same feature at different time steps is not significantly different. However, after the convolutional neural network and the long-short memory network, the hidden features extracted by them show a certain regular distribution after dimensionality reduction. It can be seen from the visualized data that some features start to show a large proportion, and the proportion of the same feature at different time steps is also quite different. Then the attention technology was introduced. We embed CBAM in the CNN path and embed the attention mechanism in the LSTM path, allowing the network to focus on important factors, so that the important points of different features can be distinguished. From the visualized data, it can be seen that certain features passing through the CBAM layer and the attention mechanism layer have a prominent proportion, and the proportion of the same feature at different time steps is also quite different. This shows to a certain extent that the introduction of attention technology enables the network to focus on key features, making the proportion of features that can affect the remaining useful life of the turbofan engine larger, thereby improving the prediction effect of the remaining useful life. At the same time, the effectiveness and superiority of the model are also verified to a certain extent.

3) COMPARE TO THE STATUS-OF-THE-ART

TABLE 5 shows the experimental results of the proposed method on the test data and compares it with the state-of-the-art methods. All methods outperform FD002 and FD004 on FD001 and FD003. This is because FD001 and FD003 are relatively simple, and there are relatively few operating conditions and fault types, which were mentioned in the

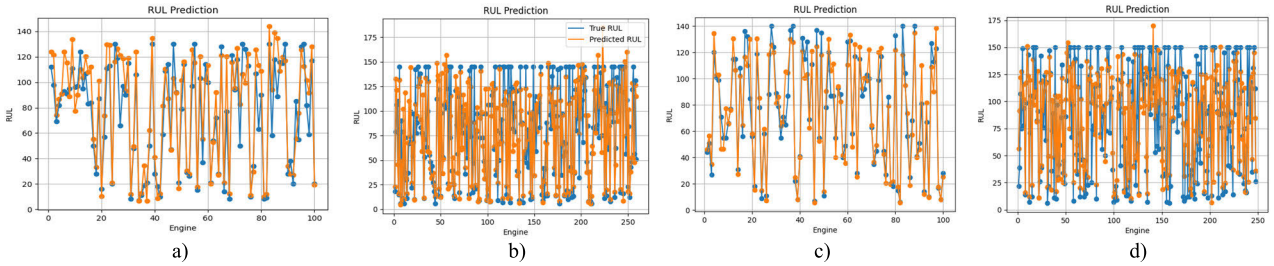

FIGURE 14. Prediction performance on each dataset a) FD001 b) FD002 c) FD003 d) FD004.

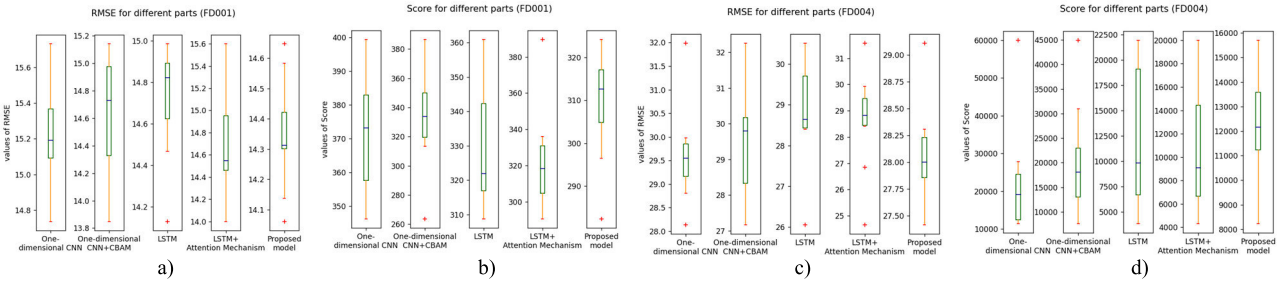

FIGURE 15. The results of the ablation experiment a) RMSE results (FD001) b) Score results (FD001) c) RMSE results (FD004) d) Score results (FD004).

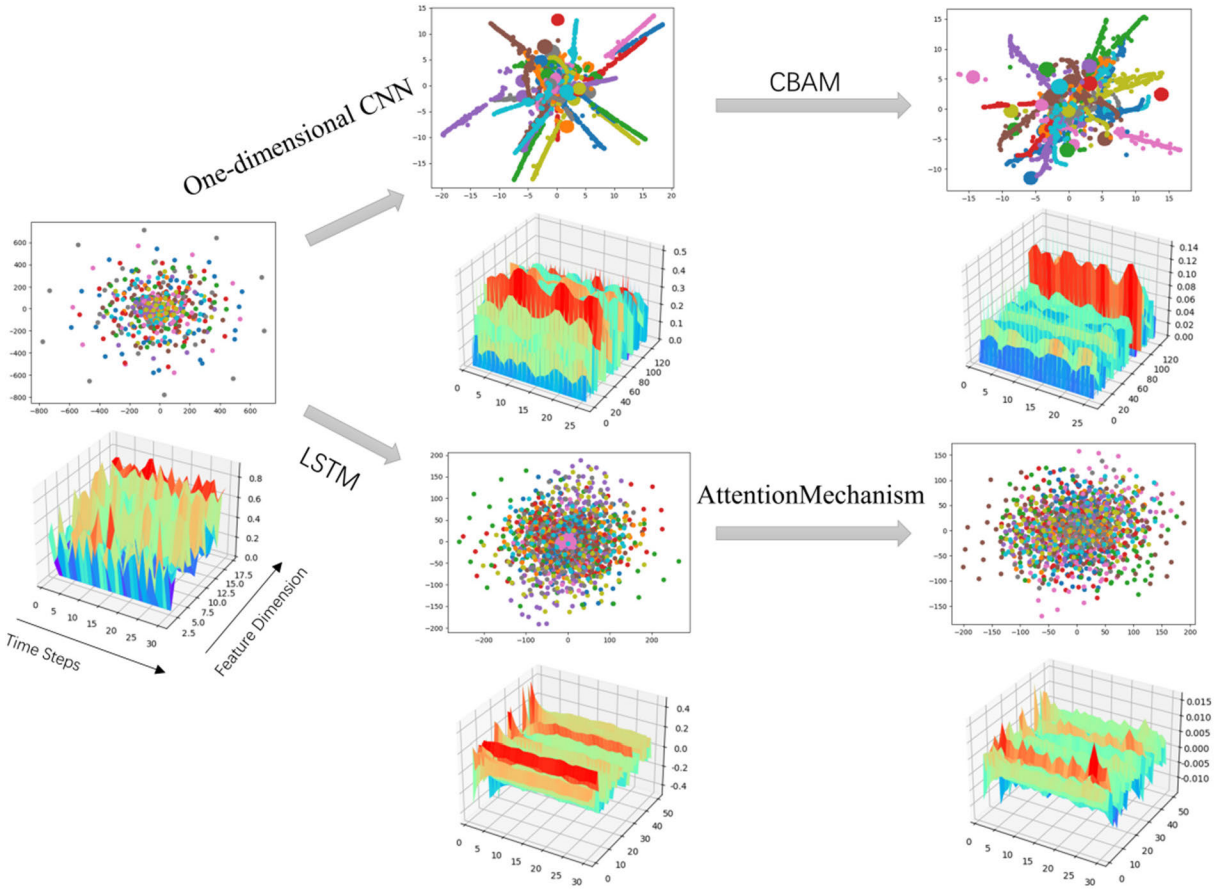

FIGURE 16. Network visualization with t-sne and data visualization.

"DATA INTRODUCTION" section. In addition, the number of engines tested by FD002 is 259, and the number of engines tested by FD004 is 248, both much higher than FD001 and FD002. Therefore, the sum of the scores of all the engines of FD002 and FD004 are of different orders of magnitude. Moreover, the performance of the proposed model has a clear

TABLE 5. Compare to the state-of-the-art methods on testing data.

		CNN [6]	RF [16]	GB [16]	DCNN-BiLSTM [49]	C-L-Hybrid [48]
FD001	RMSE	18.45	17.91	15.67	15.98	16.13
	Score	1286.70	479.75	474.01	532.16	303
FD002	RMSE	30.30	29.59	29.09	N/A	N/A
	Score	13570	70456.86	87280.06	N/A	N/A
FD003	RMSE	19.82	20.27	16.84	15.63	17.12
	Score	1596.20	711.13	576.72	365.41	1420
FD004	RMSE	29.16	31.12	29.01	N/A	N/A
	Score	7886.40	46567.63	17817.92	N/A	N/A

		LSTM [10]	DBN [16]	MODBNE [16]	Attention-Based [21]	CALAP (Proposed)
FD001	RMSE	16.14	15.21	15.04	14.54	14.35
	Score	338	417.59	334.23	322.44	309.32
FD002	RMSE	24.49	27.12	25.05	N/A	26.50
	Score	4450	9031.64	5585.34	N/A	42579.42
FD003	RMSE	16.18	14.71	12.51	N/A	14.43
	Score	852	442.43	421.91	N/A	381.41
FD004	RMSE	28.17	29.88	28.66	27.08	28.04
	Score	5550	7954.51	6557.62	5649.14	12179.41

advantage over LSTM on simple datasets. But on complex datasets, the performance of LSTM model is better than the proposed model. At the same time, through the experimental data, it can be observed that the CALAP model is not capable of early prediction of the remaining life of the engine operating under the six operating conditions, and the score is relatively large. Therefore, the proposed model is more suitable for prediction tasks that are not so demanding for early prediction.

V. CONCLUSION AND RESEARCH OUTLOOK

In this paper, we propose a data-driven deep learning fusion model—CALAP. First, we preprocess the data, use the time window to collect sample data, and use parallel methods to enable CNN and LSTM to extract spatial features and temporal features respectively. At the same time, the two combine the corresponding attention mechanism to learn important features. Finally, The proposed model is compared with existing methods, the relative superiority of the proposed method is confirmed under the scoring function and mean square error (RMSE) evaluation criteria.

Judging from the experimental results, there is still room for improvement in the prediction accuracy of turbofan engines. In future work, we will try to further optimize the model to improve the prediction accuracy and make it produce better results on complex datasets. The dataset used in this article is a classic dataset in the field of RUL prediction, but the year of the dataset is relatively old. We will try to use the latest dataset for research in subsequent research. We also plan to develop a RUL prediction model under complex working conditions based on transfer learning. Moreover, the model can not only be used in the field of RUL prediction, but also transferable to other suitable tasks.

REFERENCES

[1] Y. Hu, S. Liu, H. Lu, and H. Zhang, "Remaining useful life model and assessment of mechanical products: A brief review and a note on the state space model method," *Chin. J. Mech. Eng.*, vol. 32, no. 1, p. 15, Dec. 2019, doi: 10.1186/s10033-019-0317-y.

[2] H.-C. Trinh and Y.-K. Kwon, "A data-independent genetic algorithm framework for fault-type classification and remaining useful life prediction," *Appl. Sci.*, vol. 10, no. 1, p. 368, Jan. 2020, doi: 10.3390/app10010368.

[3] N. Gebraeel, M. Lawley, R. Liu, and V. Parmeshwaran, "Residual life predictions from vibration-based degradation signals: A neural network approach," *IEEE Trans. Ind. Electron.*, vol. 51, no. 3, pp. 694–700, Jun. 2004, doi: 10.1109/TIE.2004.824875.

[4] Z. Tian, "An artificial neural network method for resmaining useful life prediction of equipment subject to condition monitoring," *J. Intell. Manuf.*, vol. 23, no. 2, pp. 227–237, Apr. 2012, doi: 10.1007/s10845-009-0356-9.

[5] P. Lim, C. K. Goh, and K. C. Tan, "A time window neural network based framework for remaining useful life estimation," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Vancouver, BC, Canada, Jul. 2016, pp. 1746–1753, doi: 10.1109/IJCNN.2016.7727410.

[6] G. S. Babu, P. Zhao, and X.-L. Li, "Deep convolutional neural network based regression approach for estimation of remaining useful life," in *Proc. Int. Conf. Database Syst. Adv. Appl.*, 2016, pp. 214–228, doi: 10.1007/978-3-319-32025-0_14.

[7] L. Ren, Y. Sun, H. Wang, and L. Zhang, "Prediction of bearing remaining useful life with deep convolution neural network," *IEEE Access*, vol. 6, pp. 13041–13049, 2018, doi: 10.1109/ACCESS.2018.2804930.

[8] Y. Yoo and J.-G. Baek, "A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network," *Appl. Sci.*, vol. 8, no. 7, p. 1102, Jul. 2018, doi: 10.3390/app8071102.

[9] F. O. Heimes, "Recurrent neural networks for remaining useful life estimation," in *Proc. Int. Conf. Prognostics Health Manage.*, Denver, CO, USA, Oct. 2008, pp. 1–6, doi: 10.1109/PHM.2008.4711422.

[10] S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, "Long short-term memory network for remaining useful life estimation," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Dallas, TX, USA, Jun. 2017, pp. 88–95, doi: 10.1109/ICPHM.2017.7998311.

[11] Y. Wu, M. Yuan, S. Dong, L. Lin, and Y. Liu, "Remaining useful life estimation of engineered systems using vanilla LSTM neural networks," *Neurocomputing*, vol. 275, pp. 167–179, Jan. 2018, doi: 10.1016/j.neucom.2017.05.063.

[12] 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, doi: 10.1109/ACCESS.2019.2933676.

[13] J. Ma, H. Su, W.-L. Zhao, and B. Liu, "Predicting the remaining useful life of an aircraft engine using a stacked sparse autoencoder with multilayer self-learning," *Complexity*, vol. 2018, pp. 1–13, Jul. 2018, doi: 10.1155/2018/3813029.

[14] C. Sun, M. Ma, Z. Zhao, S. Tian, R. Yan, and X. Chen, "Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2416–2425, Apr. 2019, doi: 10.1109/TII.2018.2881543.

[15] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, Jul. 2006, doi: 10.1162/neco.2006.18.7.1527.

[16] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, "Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2306–2318, Oct. 2017, doi: 10.1109/TNNLS.2016.2582798.

[17] H. Li, Z. Tian, H. Yu, and B. Xu, "Fault prognosis of hydraulic pump based on bispectrum entropy and deep belief network," *Meas. Sci. Rev.*, vol. 19, no. 5, pp. 195–203, Oct. 2019, doi: 10.2478/msr-2019-0025.

[18] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[19] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Computer Vision—ECCV 2018* (Lecture Notes in Computer Science), vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, doi: 10.1007/978-3-030-01234-2_1.

[20] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Proc. Int. Conf. Prognostics Health Manage.*, Denver, CO, USA, Oct. 2008, pp. 1–9, doi: 10.1109/PHM.2008.4711414.

[21] Z. Chen, M. Wu, R. Zhao, F. Guretno, R. Yan, and X. Li, "Machine remaining useful life prediction via an attention-based deep learning approach," *IEEE Trans. Ind. Electron.*, vol. 68, no. 3, pp. 2521–2531, Mar. 2021, doi: 10.1109/TIE.2020.2972443.

[22] Z. Gu, A. Xiong, J. Wang, K. Fu, L. Wen, and H. Yang, "Research on prediction of remaining useful life of underwater turntable based on slow feature analysis," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Guilin, China, Apr. 2022, pp. 1555–1560, doi: 10.1109/ICMA54519.2022.9855959.

[23] J. Chen, D. Chen, and G. Liu, "Using temporal convolution network for remaining useful lifetime prediction," *Eng. Rep.*, vol. 3, no. 3, Mar. 2021, doi: 10.1002/eng2.12305.

[24] V. Mnih, N. Heess, A. Graves, and K. Kavukcuoglu, "Recurrent models of visual attention," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, vol. 2, Cambridge, MA, USA: MIT Press, 2014, pp. 1–12.

[25] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*.

[26] A. Vaswani, N. M. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates, 2017, pp. 6000–6010.

[27] 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, doi: 10.1016/j.ymssp.2017.11.016.

[28] A. Saxena and K. Goebel, "Turbofan engine degradation simulation data set," NASA Prognostics Data Repository, NASA Ames Res. Center, Moffett Field, CA, USA, 2008. [Online]. Available: https://www.nasa.gov/intelligent-systems-division/discovery-and-systems-health/pcoe/pcoe-data-set-repository/

[29] C. Zhao, X. Huang, Y. Li, and M. Y. Iqbal, "A double-channel hybrid deep neural network based on CNN and BiLSTM for remaining useful life prediction," *Sensors*, vol. 20, no. 24, p. 7109, Dec. 2020, doi: 10.3390/s20247109.

[30] X. Liu, L. Liu, D. Liu, L. Wang, Q. Guo, and X. Peng, "A hybrid method of remaining useful life prediction for aircraft auxiliary power unit," *IEEE Sensors J.*, vol. 20, no. 14, pp. 7848–7858, Jul. 2020, doi: 10.1109/JSEN.2020.2979797.

[31] Y. Duan, H. Li, M. He, and D. Zhao, "A BiGRU autoencoder remaining useful life prediction scheme with attention mechanism and skip connection," *IEEE Sensors J.*, vol. 21, no. 9, pp. 10905–10914, May 2021, doi: 10.1109/JSEN.2021.3060395.

[32] B. Fu, W. Yuan, X. Cui, T. Yu, X. Zhao, and C. Li, "Correlation analysis and augmentation of samples for a bidirectional gate recurrent unit network for the remaining useful life prediction of bearings," *IEEE Sensors J.*, vol. 21, no. 6, pp. 7989–8001, Mar. 2021, doi: 10.1109/JSEN.2020.3046653.

[33] Z. Zhao, B. Liang, X. Wang, and W. Lu, "Remaining useful life prediction of aircraft engine based on degradation pattern learning," *Rel. Eng. Syst. Saf.*, vol. 164, pp. 74–83, Aug. 2017, doi: 10.1016/j.ress.2017.02.007.

[34] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Multi-layer domain adaptation method for rolling bearing fault diagnosis," *Signal Process.*, vol. 157, pp. 180–197, Apr. 2019, doi: 10.1016/j.sigpro.2018.12.005.

[35] X. Li, W. Zhang, and Q. Ding, "Understanding and improving deep learning-based rolling bearing fault diagnosis with attention mechanism," *Signal Process.*, vol. 161, pp. 136–154, Aug. 2019, doi: 10.1016/j.sigpro.2019.03.019.

[36] Y. Ding, M. Jia, Q. Miao, and P. Huang, "Remaining useful life estimation using deep metric transfer learning for kernel regression," *Rel. Eng. Syst. Saf.*, vol. 212, Aug. 2021, Art. no. 107583, doi: 10.1016/j.ress.2021.107583.

[37] Y. Deng, D. Shichang, J. Shiyao, Z. Chen, and X. Zhiyuan, "Prognostic study of ball screws by ensemble data-driven particle filters," *J. Manuf. Syst.*, vol. 56, pp. 359–372, Jul. 2020, doi: 10.1016/j.jmsy.2020.06.009.

[38] C.-G. Huang, H.-Z. Huang, Y.-F. Li, and W. Peng, "A novel deep convolutional neural network-bootstrap integrated method for RUL prediction of rolling bearing," *J. Manuf. Syst.*, vol. 61, pp. 757–772, Oct. 2021, doi: 10.1016/j.jmsy.2021.03.012.

[39] H. Li and C. Wang, "Combining first prediction time identification and time-series feature window for remaining useful life prediction of rolling bearings with limited data," *Proc. Inst. Mech. Eng., O, J. Risk Rel.*, vol. 2023, Jan. 2023, Art. no. 1748006, doi: 10.1177/1748006X221147441.

[40] R. Zhu, W. Peng, D. Wang, and C.-G. Huang, "Bayesian transfer learning with active querying for intelligent cross-machine fault prognosis under limited data," *Mech. Syst. Signal Process.*, vol. 183, Jan. 2023, Art. no. 109628, doi: 10.1016/j.ymssp.2022.109628.

[41] X. Zhou, N. Zhai, S. Li, and H. Shi, "Time series prediction method of industrial process with limited data based on transfer learning," *IEEE Trans. Ind. Informat.*, vol. 19, no. 5, pp. 6872–6882, May 2023, doi: 10.1109/TII.2022.3191980.

[42] Y. Le Cun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a back-propagation network," in *Proc. 2nd Int. Conf. Neural Inf. Process. Syst. (NIPS)*. Cambridge, MA, USA: MIT Press, 1989, pp. 396–404. [Online]. Available: https://dl.acm.org/doi/10.5555/2969830.2969879

[43] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[44] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[45] F.-K. Wang, Z. E. Amogne, and J.-H. Chou, "A hybrid method for remaining useful life prediction of proton exchange membrane fuel cell stack," *IEEE Access*, vol. 9, pp. 40486–40495, 2021, doi: 10.1109/ACCESS.2021.3064684.

[46] J.-Y. Wu, M. Wu, Z. Chen, X.-L. Li, and R. Yan, "Degradation-aware remaining useful life prediction with LSTM autoencoder," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–10, 2021, doi: 10.1109/TIM.2021.3055788.

[47] C. Ferreira and G. Gonçalves, "Remaining useful life prediction and challenges: A literature review on the use of machine learning methods," *J. Manuf. Syst.*, vol. 63, pp. 550–562, Apr. 2022, doi: 10.1016/j.jmsy.2022.05.010.

[48] Z. Kong, Y. Cui, Z. Xia, and H. Lv, "Convolution and long short-term memory hybrid deep neural networks for remaining useful life prognostics," *Appl. Sci.*, vol. 9, no. 19, p. 4156, Oct. 2019, doi: 10.3390/app9194156.
[49] W. Chen, W. Chen, H. Liu, Y. Wang, C. Bi, and Y. Gu, "A RUL prediction method of small sample equipment based on DCNN-BiLSTM and domain adaptation," *Mathematics*, vol. 10, no. 7, p. 1022, Mar. 2022, doi: 10.3390/math10071022.
[50] L. Ren, Y. Sun, J. Cui, and L. Zhang, "Bearing remaining useful life prediction based on deep autoencoder and deep neural networks," *J. Manuf. Syst.*, vol. 48, pp. 71–77, Jul. 2018, doi: 10.1016/j.jmsy.2018.04.008.
[51] J. Couture and X. Lin, "Novel image-based rapid RUL prediction for Li-ion batteries using a capsule network and transfer learning," *IEEE Trans. Transport. Electrific.*, vol. 9, no. 1, pp. 958–967, Mar. 2023, doi: 10.1109/TTE.2022.3173918.
[52] W. Mao, K. Liu, Y. Zhang, X. Liang, and Z. Wang, "Self-supervised deep tensor domain-adversarial regression adaptation for online remaining useful life prediction across machines," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–16, 2023, doi: 10.1109/TIM.2023.3265109.
[53] Y. Jiang, T. Xia, D. Wang, X. Fang, and L. Xi, "Adversarial regressive domain adaptation approach for infrared thermography-based unsupervised remaining useful life prediction," *IEEE Trans. Ind. Informat.*, vol. 18, no. 10, pp. 7219–7229, Oct. 2022, doi: 10.1109/TII.2022.3154789.
[54] Y. Ding, M. Jia, J. Zhuang, and P. Ding, "Deep imbalanced regression using cost-sensitive learning and deep feature transfer for bearing remaining useful life estimation," *Appl. Soft Comput.*, vol. 127, Sep. 2022, Art. no. 109271, doi: 10.1016/j.asoc.2022.109271.
[55] Q. Qian, Y. Wang, T. Zhang, and Y. Qin, "Maximum mean square discrepancy: A new discrepancy representation metric for mechanical fault transfer diagnosis," *Knowl.-Based Syst.*, vol. 276, Sep. 2023, Art. no. 110748, doi: 10.1016/j.knosys.2023.110748.
[56] Q. Qian, J. Zhou, and Y. Qin, "Relationship transfer domain generalization network for rotating machinery fault diagnosis under different working conditions," *IEEE Trans. Ind. Informat.*, vol. 19, no. 9, pp. 9898–9908, Sep. 2023, doi: 10.1109/TII.2022.3232842.
[57] Q. Qian, Y. Qin, J. Luo, Y. Wang, and F. Wu, "Deep discriminative transfer learning network for cross-machine fault diagnosis," *Mech. Syst. Signal Process.*, vol. 186, Mar. 2023, Art. no. 109884, doi: 10.1016/j.ymssp.2022.109884.
[58] Q. Qian, Y. Qin, J. Luo, and D. Xiao, "Cross-machine transfer fault diagnosis by ensemble weighting subdomain adaptation network," *IEEE Trans. Ind. Electron.*, vol. 70, no. 12, pp. 12773–12783, Dec. 2023, doi: 10.1109/TIE.2023.3234142.
[59] M. Wei, M. Ye, Q. Wang, and J. P. Twajamahoro, "Remaining useful life prediction of lithium-ion batteries based on stacked autoencoder and Gaussian mixture regression," *J. Energy Storage*, vol. 47, Mar. 2022, Art. no. 103558, doi: 10.1016/j.est.2021.103558.
[60] Q. Ye and C. Liu, "An unsupervised deep feature learning model based on parallel convolutional autoencoder for intelligent fault diagnosis of main reducer," *Comput. Intell. Neurosci.*, vol. 2021, pp. 1–12, Sep. 2021, doi: 10.1155/2021/8922656.
[61] W. Kong and H. Li, "Remaining useful life prediction of rolling bearing under limited data based on adaptive time-series feature window and multi-step ahead strategy," *Appl. Soft Comput.*, vol. 129, Nov. 2022, Art. no. 109630, doi: 10.1016/j.asoc.2022.109630.

MINGYAN WU (Student Member, IEEE) is currently pursuing the bachelor's degree with the School of Computer Science, Yangtze University.

She is in charge of the University Student Innovation and Entrepreneurship Training Program. Her research interests include prognostics and health management, anomaly sensor data detection, computer vision, and natural language processing. She received the Provincial Prize in the National College Students Mathematical Modeling Competition and the Provincial Prize in the China Computer Design Competition.

QING YE received the Ph.D. degree in computer science and technology from the Wuhan University of Technology.

She is currently an Associate Professor with the School of Computer Science, Yangtze University. The projects hosted and participated in the past five years include, the National Natural Science Foundation Project, the General Project of Natural Science Foundation of Hubei Provincial Department of Science and Technology, the Major Project of Technology Innovation Special Project of Hubei Provincial Department of Science and Technology, and the Young Talent Project of Science and Technology Research Program of Hubei Provincial Department of Education. Her research directions are deep learning, intelligent status monitoring, and equipment health status management.

JIANXIN MU (Student Member, IEEE) is currently pursuing the bachelor's degree with the School of Mechanical Engineering, Yangtze University.

During the period of school, he has participated in the 15th Higher Education Cup National College Students Advanced Surveying and Mapping Technology and Product Information Modeling Innovation Competition, and won multiple awards. His current research interests include environmentally friendly mechanical design and mechanical structure strength fatigue analysis.

ZUYU FU (Student Member, IEEE) is currently pursuing the bachelor's degree with the School of Mechanical Engineering, Yangtze University.

He won two national inspirational scholarships and a utility model patent as a first inventor. His research direction are mechanical structure strength fatigue analysis and finite element simulation analysis. He has participated in energy conservation and emission reduction competitions, mathematical modeling competitions, and mechanical innovation design competitions, and won provincial awards or above. He was selected into the Innovative Talent Training Program of Yangtze University.

YILIN HAN is currently pursuing the bachelor's degree with the School of Electronic Information, Yangtze University.

He achieved excellent results during the bachelor's studies. He attempts to understand the most fundamental principles of machine algorithms and wants to innovate. He has participated in the College Student's Innovative Entrepreneurial Training Plan Program and the National College Student Mathematical Modeling Competition, during his university years, and won the third prize in the province in the National College Student Mathematical Modeling Competition. He is also preparing to take the postgraduate entrance exam in the field of 5G communication engineering technology. His current research interests include machine learning data analytics and embedded software development.