A Hybrid Deep Computation Model for Feature Learning on Aero-engine Data – Applications to Fault Detection

Zhikui Chen[a], Xu Yuan[a], Mingyang Sun[a], Jing Gao[a,*], Peng Li[a]

[a]*the School of Software Technology, Dalian University of Technology and the Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian, China.*

Abstract

Recently, the security of aircraft has attracted much attention with some air crashes occurring. The gas-path faults, as the most common faults of the aircraft, pose a vast challenge on the security of aircraft, due to the complexity of the aero-engine structure. In this paper, a hybrid deep computation model is proposed to effectively detect the gas-path faults based on the performance data. In detail, to capture the local spatial features of the gas-path performance data, an un-fully-connected convolutional neural network of one-dimensional kernels is employed. Furthermore, to model the temporal patterns hidden in the gas-path faults, a recurrent computation architecture is introduced. Finally, extensive experiments are conducted on the real aero-engine data. The results show that the proposed model can outperform the compared models.

Keywords: `Feature learning`, `Deep computation`, `Gas-path fault detection`

1. Introduction

With the development of technology, the living custom of people is changing greatly. Aircraft are becoming the prevalent vehicle. As a result, the security of aircraft attracts much attention. Aero-engine is the power source and its

*Corresponding author: Jing Gao.
Email address: `gaojing@dlut.edu.cn` (Jing Gao)

working state has a great effect on the security of aircraft. A lot of terrible accidents caused by the fault of aero-engine have happened. For instance, a Southwest Airlines flight from New York to Dallas crashed due to the engine failure [1]. The aero-engine confronts many unexpected faults, such as mechanical breakdown and gas-path fault, since it always works in an environment of high temperature, high pressure, and high vibration. Among all the faults of the aero-engine, the gas-path fault is the most common fault. The effective detection of gas-path fault can prevent a great number of aircraft disasters. Thus, the detection of the gas-path fault is a vital research.

In recent years, many detection methods have been proposed to detect the gas-path faults. For example, Xue et al. proposed a fusion method based on Kalman filters to detect the gas-path faults [2]. Sun et al. designed a set of rules that can combine the conflicting evidences based on rough set and Dempster Shafer theories [3]. The above mathematic-model-based methods cannot model the intrinsic faults patterns well, due to the complexity of the aero-engine structure. To solve this problem, some researchers devised the data-driven methods to detect gas-path faults. For example, Yu et al. used the least squares support vector machine to classify the aero-engine faults [4]. Zhao et al. designed a neural network to learn features of the diagnosis data [5]. Those data-driven methods can improve the detection accuracy of gas-path faults. However, those models are shallow models that cannot capture intrinsic fault patterns of more advanced modern aero-engines. Thus, the accurate detection of gas-path faults requires novel methods.

Deep learning, as a novel method, can well perceive high-order state features of the big data and produce effective non-linear representations [6] [7]. It has been widely used in image classification [8] and speech recognition [9]. Recently, inspired by outperforming results of deep learning in other domains, the application of deep learning to the gas-path fault detection of aero-engines has attracted much attention.

In this paper, a hybrid deep computation model is proposed for the detection of aero-engine gas-path faults. In details, an un-fully-connected convolutional neural network of one-dimensional kernels is employed to capture the local features of the gas-path performance data, which can take advantage of the space topologies hidden in the gas-path data. Furthermore, a recurrent computation architecture is employed to mine the temporal dependencies of fault patterns in the gas-path data. Then, based on the two above computing modules, a convolutional and recurrent architecture is proposed to capture the instinct patterns of the gas-path faults. Finally, to evaluate the hybrid deep computation model, extensive experiments are conducted on the real aero-engine data. The results show that the proposed model can outperform the comparison models. Thus, the contributions of this work can be summarized as follows:

- To improve the safety of aircraft, a hybrid deep computation model is designed, which can effectively detect gas-path faults for aero-engines.
- To capture instinct patterns of gas-path faults, convolutional and recurrent computing modules are designed, which can learn the spatial and temporal features of gas-path data. Furthermore, the corresponding back-propagation rules are introduced to train the proposed modules.
- To evaluate the proposed model, extensive experiments are conducted on the real aero-engine data, which illustrate the effectiveness of the hybrid deep computation model.

The rest of paper is organized as follows. In Section 2, a brief review of some related works is introduced. The details of proposed model are given in Section 3 and the hybrid error back propagation is introduced in Section 4. In Section 5, the detailed experimental setting and results are provided. Finally, Section 6 concludes this paper.

2. Related Work

In this paper, a convolutional and recurrent architecture is proposed to detect the patterns of faults in gas-path data sequences. Thus, the related works about fault detection, convolutional neural networks and recurrent neural networks are introduced in this section.

2.1. Fault Detection

In the past, a lot of methods were proposed for the detection of aero-engine faults. For example, Xue et al. [2] proposed an aero-engine fault detection scheme based on the information fusion by designing a set of Kalman filters to detect the characteristics of sensors and actuators. The Kalman-based fault detection algorithm is limited by the slow response speed, low multi-fault accuracy and off-design point fault detection. To solve these problems, Pan et al. [13] presented a fault detection method based on the improved Broyden algorithm. At the same time, some machine-learning-based methods were proposed, which achieved outperforming results. The machine-learning-based methods are easy to be extended to the fault detection of various kinds of aero-engines. For instance, Lu et al. [10] proposed a multi-output least square support vector regression model based on the particle swarm optimization to construct a data-driven aero-engine component fault detection model. Xu et al. [11] introduced a hierarchical SVM-based multi-classification model for fault detection of aero-engines in large-scale dataset, in which the relative boundary vector is used to filter out redundant data samples. To further improve the performance of SVM on the aero-engine fault detection, Huang et al. [12] employed the genetic algorithm and the idea of simulated annealing to optimize the parameters of SVM. To a certain extent, those machine-learning-based methods improve the aero-engine fault accuracy. However, due to their shallow modeling ability, they cannot fully extract the deep features of gas-path data, especially in a non-linear way. Thus, in this paper, a hybrid deep computation model is proposed for the detection of aero-engine gas-path faults.

2.2. Convolutional neural networks

Convolutional neural networks (CNNs) which can well capture the deep spatial features of data are very suitable to model non-linear systems. Over the past few years, many representative models have been introduced. For example, Krizhevsky et al. [14] proposed the AlexNet based on the ReLU activation function and the dropout fully-connected layers, which can ensure the fast convergence to the deep intrinsic features and reduce overfitting of the model. To further extract the deeper features of data, Szegedy et al. [15] then presented GoogLeNet, which can greatly increase the network depth without increasing the number of parameters. Meanwhile, Simonyan et al. [16] developed the VGGNet model by replacing the convolution layer of large filters with several continuous convolution layers of smaller filters. However, as the network depth increases, the training of deep models poses a vast challenge. To solve that challenge, He et al. [17] proposed the ResNet model, in which a residual structure is introduced to effectively back-propagate the network loss, greatly increasing the network depth. To further enhance the performance of the network representation, Hu et al. [18] developed a squeeze-and-excitation network, in which the modulation weights for feature maps of each channel are adaptively calculated to recalibrate channel-wise features. Recently, some researchers have focused on exploring inherent features in raw signals. Jiang et al. [19] designed a new convolutional neural architecture for fault detection based on a multi-scale coarse-grained layer where the consecutive coarse-grained signals are constructed from raw current signals. Those models can well extract the spatial features of data. However, those CNN-based models cannot well learn the time-dependent characteristic. In this paper, a convolutional and recurrent model is introduced to extract the temporal dependency between spatial features in aero-engine gas-path signals.

2.3. Recurrent neural networks

Recurrent neural networks (RNNs) are powerful methods of modeling the time-domain data, which are suitable to learn the time-dependent relationship. Many representative recurrent computing architectures were proposed. For example, Hochreiter et al. [20] proposed long short-term memory (LSTM) network to deal with the long-term dependency. Cho et al. [21] proposed gated recur-

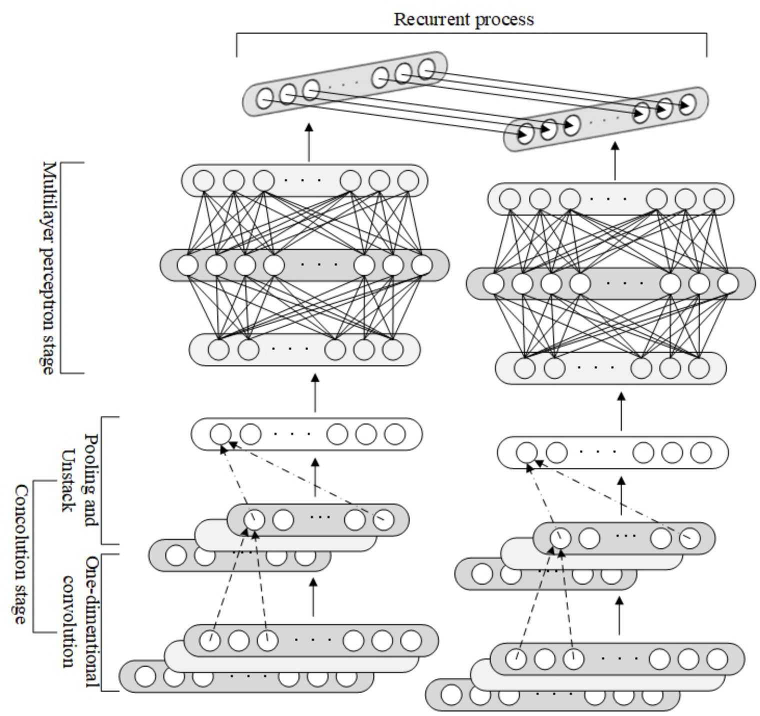

Figure 1: The hybrid deep computation model used for the gas-path fault detection of aero-engines.

rent unit (GRU) network to simplify the complex internal structure of LSTM. Recurrent structures have been applied in fault detection. For instance, Bruin et al. [22] used RNN to detect and identify faults in railway track circuits based on the measurement signals. Yoshikawa et al. [23] conducted an empirical study on the application of LSTM network in the fault detection for aileron failures of aircraft. RNNs are a promising method for the fault detection, due to the extraction of the temporal dependencies. However, they ignore the spatial dependencies of local features in data. In this paper, a convolutional and recurrent model is introduced to extract the spatial and temporal features of gas-path signals.

3. The Hybrid Deep Computation Model

In this section, a hybrid deep computation (HDC) model is proposed to learn intrinsic features of the aero-engine data for gas-path fault detection. In particular, each gas-path data collected continuously from performance parameters of aero-engines is first split into k time steps as the input. Then, the spatial and temporal features are modeled in a layer-wise manner. The hybrid deep computation is composed of a one-dimension convolutional network, a multi-layer perceptron and a recurrent network, as shown in Figure 1. In the hybrid deep computation network, the one-dimension convolutional network perceives local topology features of gas-path data. The multi-layer perceptron is used to fit the complex relationships of those local features in the high-level space. The recurrent network models the temporal dependencies between local features.

3.1. The Local Spatial Features Extraction

The gas-path parameter data of aero-engines is a kind of continuous and sequential data. In the data sequence, adjacent values are corelated, and the values which are far away from each other are respectively independent. In other words, there are some local spatial topologies in the gas-path parameter data, which is verified by the following ablation experiments (Section 5, the experiments of RNN) . To capture spatial topologies in the gas-path parameter data of aero-engines, the one-dimension convolutional network is employed, which learns local features of data [27].

In detail, each gas-path parameter sequential data X is first split into k time steps. Each time step x_i is embedded into d-dimension vector. Each sequential data is modeled as a data block $X \in R^{k \times d}$. Afterwards, the one-dimension convolution is used to learn local features of the gas-path parameter data by the following form:

$$C = f\left(W * X + b\right), \tag{1}$$

where C represents the feature map value, W is the kernel, $*$ denotes the one-dimension convolutional operation, and f is the non-linear activation function, such as hyperbolic tangent, sigmoid and ReLU. In this paper, ReLU is chosen as the activation function, since it can well disentangle the complex correlations and is more robust to the unknown noise. More specially, for each position j in the data block, a window vector v_j with l consecutive time steps can be obtained, denoted as $v_j = [x_j, x_{j+1}, \cdots, x_{j+l-1}]$. The j-th feature in the feature map C is computed as follows:

$$C^j = f\left(W \otimes v_j + b\right), \tag{2}$$

where $\otimes$ is the inner product between the convolutional kernel and the j-th window vector.

As shown in Figure 1, in convolutional network, one-dimension filters of different length are used to fully perceive the local features of the gas-path parameter data. After that, the max pooling layer is added to filter the salient features that should be retained in the high-level feature vector, since some values in the gas-path sequence do not have significant influence on the recognition of aero-engine faults.

3.2. The High-level Features Extraction

The linear combination of the local spatial features hidden in the gas-path fault data cannot well represent the various patterns of the gas-path faults, since they are the basic elements shared by all the gas-path faults. Also, the linear combination cannot well capture the correlations between features due to the complexity of the aero-engine structure. Thus, the multilayer perceptron (MLP) is introduced to combine local spatial features [25], fitting the high-level features, as shown in Figure 1. The MLP mapping is expressed as follows:

$$O = f\left(W \cdot C + b\right), \tag{3}$$

where O denotes the high-level feature of the gas-path fault data, and C is the local salient feature.

3.3. The Long-Term Dependency Features Extraction

The aero-engine gas-path fault is the accumulated result that is produced by the inter-reaction of components of aero-engines. The different order of abnormal features in the sequence can give rise to different faults. That is, those aero-engine gas-path features are of high temporal dependency. For example, if the fan cowl abnormal pattern is above the abnormal temperature pattern of low-pressure turbine in the data sequence, this abnormal data sequence is defined as the fan fairing failure, otherwise as the fan wear seal failure. Thus, the long short-term memory (LSTM) network is used to detect the temporal characteristic of gas-path data in this paper.

LSTM models the temporal features of the aero-engine gas-path fault data via:

$$\begin{aligned} f_t &= \sigma\left(W_f X_t + b_f\right) \\ i_t &= \sigma\left(W_i X_t + b_i\right) \\ {c'}_t &= \tanh\left(W_c X_t + b_c\right) \\ c_t &= f_t \odot c_{t-1} + i_t \odot {c'}_t \\ o_t &= \sigma\left(W_o X_t + b_o\right) \\ h_t &= o_t \odot \tanh\left(c_t\right) \end{aligned}, \tag{4}$$

where X_t is the high-level feature of the t-th data block, h_{t-1} denotes the previous hidden output, W is the transfer parameter, σ is the sigmoid function, *tanh* is the hyperbolic tangent function, and the operator $\odot$ represents the Hadamard product [20].

In the LSTM network, the temporal dependency of the gas-path data is captured by the special structure of memory cell, as shown in Figure 2. Specifically, a forget gate unit f controls the amount of information of the previous sub-local features in the sequence, which can be transferred into next state. The input unit is responsible for receiving the local feature of the current time step. The output unit produces the spatial and temporal features of the fault detection in the aero-engine gas-path data.

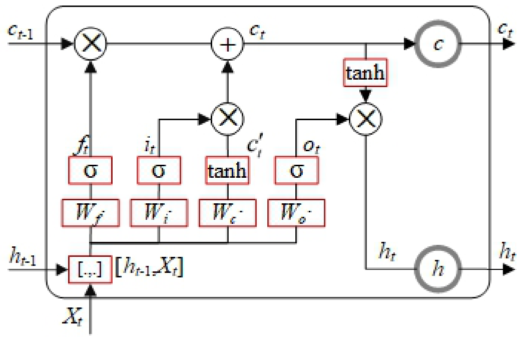

Figure 2: LSTM recurrent unit internal structure.

4. The HDC Learning For Fault Detection

As shown in Figure 1, each gas-path parameter sequential data modeled as a data block is fed into HDC. Then, the output feature is input into a softmax layer to recognize the gas-path parameter sequential data. The HDC learning is based on the back-propagation algorithm optimized by minimizing the cross-entropy loss. It is composed of four stages, i.e., the output layer learning, the LSTM network learning, the MLP network learning, and the CNN network learning.

The Output Layer Learning. Given a training sample x_i and its true probability distribution $y_i \in R^{|K|}$, where K is the number of fault patterns, and y_i^j=1 if the sample belongs to class j, otherwise $y_i^j = 0$ for each label $j \in \{1, 2, \cdots, K\}$. Let $\tilde{y}_i \in R^{|K|}$ be the predicted probability distribution and $\tilde{y}_i^j \in [0, 1]$ for each label $j \in \{1, 2, \cdots, K\}$. The cross-entropy function with the weight penalty is used to train parameters as follows:

$$L = -\frac{1}{|N|} \sum_{i}^{N} \sum_{j}^{K} y_i^j \ln \tilde{y}_i^j + \frac{\lambda}{2} \|\theta\|^2 \tag{5}$$

where $|N|$ is the number of training data, $\lambda > 0$ is a hyper-parameter of L2-norm regularization term and θ are the trainable parameters.

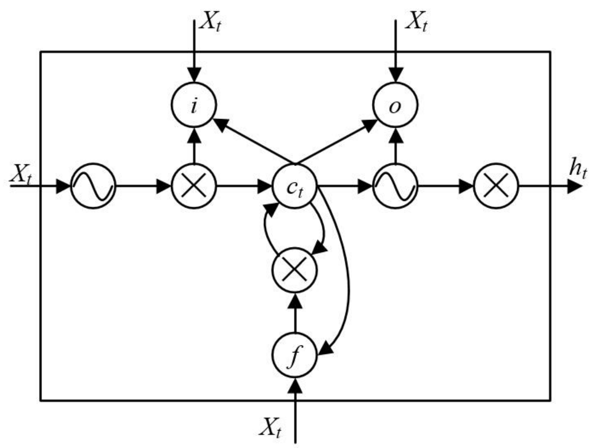

Figure 3: The unrolled structure of the LSTM recurrent unit.

Thus, the back-propagation of the loss in the softmax output layer is computed as follows:

$$\delta_j = \tilde{y}_i^j \left(1 - \frac{y_i^j}{\tilde{y}_i^j}\right) = \tilde{y}_i^j - y_i^j. \tag{6}$$

The detail derivation of the softmax output layer is in [24].

The LSTM Network Learning. In the LSTM network learning, the previous local features can influence the subsequent features. Thus, the loss can back propagate through time. For example, as shown in the Figure 3, the activation of the cell c_t is used to compute h_t and the next cell state c_{t+1}. The c_t receives gradients from h_t and the cell state of the next time step c_{t+1}. The back-propagation of the loss in the LSTM network is expressed as follows:

$$\delta_p^T = \prod_{j=p}^{t-1} \left(\delta_{o,j}^T W_{oh} + \delta_{f,j}^T W_{fh} + \delta_{i,j}^T W_{ih} + \delta_{c',j}^T W_{ch}\right). \tag{7}$$

The detail derivation of the LSTM network is shown in [20]

The MLP network Learning. In the MLP layer, the loss is transferred by the mini-batch stochastic gradient descent algorithm. For instance, let l be the number of layers, n^l be the number of neurons in the l-th layer, $f_l(\cdot)$ be the activation function for l-th layer, $W_M^{(l)} \in R^{n^l \times n^{l-1}}$ represent the weight matrix from layer *l-1* to l, $b_M^{(l)} \in R^{n^l}$ represent the bias and $z_M^{(l)} \in R^{n^l}$ denote the state of neurons in the *l-th* layer. The MLP propagates loss via the following mapping:

$$\delta_t^{(l)} = \frac{\delta_t^M}{\partial z_M^{(l)}} = f_l{}' \left(z_M^{(l)} \right) \odot \left(\left(W_M^{(l+1)} \right)^T \cdot \delta_t^{(l+1)} \right) \tag{8}$$

The detail derivation of the MLP network is shown in [25]

The CNN network Learning. The CNN loss E_c is propagated through the pooling and convolutional layers. Specifically, for the pooling layer l, the back-propagation computation of the loss to the upper convolutional layer is as follows:

$$\delta^{l-1} = \frac{\partial E_c}{\partial z_c^{l-1}} = \frac{\partial E_c}{\partial a_c^{l-1}} \frac{\partial a_c^{l-1}}{\partial z_c^{l-1}} = up\left(\delta^l\right) \odot f'\left(z_c^{l-1}\right), \tag{9}$$

where a_c denotes the result processed by the activation function and z_c is the weighted sum, $up\left(\cdot\right)$ represents an upsampling function, which completes the operation of pooling loss matrix expansion and loss redistribution.

For the convolutional layer l, the loss is propagated back to the upper pooling layer as follows:

$$\delta^{l-1} = \delta^l \frac{\partial z_c^l}{\partial a_c^{l-1}} \frac{\partial a_c^{l-1}}{\partial z_c^{l-1}} = \delta^l * rot180\left(W_c^l\right) f'\left(z_c^{l-1}\right), \tag{10}$$

where * is a similar operation to convolution, $rot180\left(W_c^l\right)$ rotates the filter W_c^l by 180 degrees. The detail derivation of the CNN network is shown in [26] and [27].

5. Experiments

In this section, extensive experiments on real aero-engine data are conducted to evaluate the HDC model. All the experiments are carried out on the server with 64-GB memory and 10-core, 20-thread, Inter Xeon E7-4800CPU.

5.1. The Dataset Description

The experiment dataset of desensitized data is provided by AVIC Shenyang Engine Design Institute. It includes five fault patterns, i.e., fan fairing failure

Table 1: The description of gas-path data

Fault patterns	Number of sequences (A)	Number of sequences (B)	Number of indicators
F_1	32180	25447	23
F_2	28555	25354	23
F_3	28207	25687	23
F_4	28266	25438	23
F_5	28272	25734	23

(F_1), supercharger failure(F_2), combustion chamber failure (F_3), fan wear seal failure (F_4) and low-pressure turbine failure (F_5). Each parameter sequence contains 23 performance indicators such as low-pressure rotor speed (N1), high-pressure rotor speed (N2), throttle lever angle (TLA), atmospheric temperature (T0) and pressure (P0), etc. The detailed description about the number of gas-path data is as shown in Table 1.

In the experiment, all the gas-path data are split into blocks in which each block contains 32 sequences. 300 data blocks in each fault pattern are randomly selected as the test data, while the rest are used to train the models.

5.2. The Experimental Settings

In the one-dimension convolutional network, 32 filters of various lengths (1, 3 and 5) are used to learn the spatial features of gas-path data with the stride of all filters set to be 1. Then, a max-pooling layer is used to extract the salient features, and it outputs a spatial feature vector of 96 dimensions. In the multi-layer perceptron network, a 128-dimension layer and a 256-dimension layer are used to model high-level features of gas-path data sequences, followed by a 128-dimension output layer. Then, each local high-level feature is fed into the LSTM network at a time step, in which the cell state dimension and hidden output dimension are both set to be 150. The output of the LSTM layer is dropped out with a probability of 0.5. Additionally, the L2-regularization with a factor of 0.001 is set to the weights in the softmax layer. After that, all of weights are randomly initialized by the Xavier method with all of the biases set to 0. The initial hidden state $h_0 \in R^{d'}$ and cell state $c_0 \in R^{d'}$ are set to 0.

To verify the effectiveness of the proposed model, several common approaches for the fault detection task are used as the comparison approaches. These methods are introduced as follows:

- **SVM**: The support vector machine is a classic supervised discriminative model based on the margin maximization strategy. It has been widely used in all kinds of classification tasks including fault detection in aero-engine. In the experiment, the local high-order feature vectors are input into SVM as the training data, since SVM cannot well model the complex relationship in the raw gas-path data. In the experiment, it is used to verify the HDC performance.
- **ABC-BP**: The back propagation neural network (BPNN) is a multi-layer feed-forward neural network trained by the back-propagation algorithm. Artificial Bee Colony (ABC) algorithm is used to optimize the parameter initialization of the back-propagation neural network. Similarly, the local high-order feature vectors are also used as the training data. In the experiment, it is used to verify the HDC performance.
- **SVM+BP+D-S**: The authors in [28] proposed a fusion diagnosis method for gas-path faults of aero-engines based on the D-S evidence theory. The diagnosis results of BPNN and SVM are used as evidence for decision fusion. In the experiment, it is used to verify the HDC performance.
- **CNN**: The one-dimension convolutional network with the softmax layer is used to verify the effectiveness of the temporal dependency between local spatial features, which is trained by an end-to-end manner.
- **RNN**: The LSTM-based recurrent network with the softmax layer is used to verify the effectiveness of spatial features between gas-path data, which is trained by an end-to-end manner.

5.3. Experimental Results of Recall

Figure 4 shows the results produced by those models in terms of recall. HDC produces the best results for all the kinds of fault patterns. In particular, HDC achieves the recall that is higher than 80%. SVM and ABC-BP produce the worst recall, since they are only based on the local high-level spatial features of gas-path performance parameter data. Also, those two methods cannot well fit the complex temporal dependencies between local high-level spatial features. D-S produces the similar results with SVM and ABC-BP. That is because D-S merges the results of SVM and ABC-BP – if conflicting gas-path sequences are classified into correct categories, the recall increases; otherwise decreases. CNN and RNN produce the higher recall than SVM, ABC-BP and D-S, since CNN and RNN are the end-to-end deep models that can model part of spatial and temporal features of gas-path fault patterns.

The aero-engine works in an environment of high pressure and high temperature, resulting that the gas-path data contains much noise. Thus, the robustness of models to the noise is very crucial in the fault detection of aero-engines. To further evaluate the robustness of the proposed model, some experiments are conducted on the gas-path data in which artificial noise that may be introduced during data acquisition and transmission is added. The results are shown in Figure 5. Similar observations can be obtained from those results. Particularly, the detection accuracy of all the models decreases. HDC still produces the best results for all the kinds of fault patterns, indicating that HDC is of high robustness.

5.4. Experimental Results of Kappa coefficient

To fully evaluate the performance of the proposed model, the Kappa coefficient is also used to measure classification accuracy of the fault detection. The Kappa coefficient that is widely used to assess the multi-class classification is computed based on the confusion matrix as follows:

$$k = \frac{p_0 - p_e}{1 - p_e}, \tag{11}$$

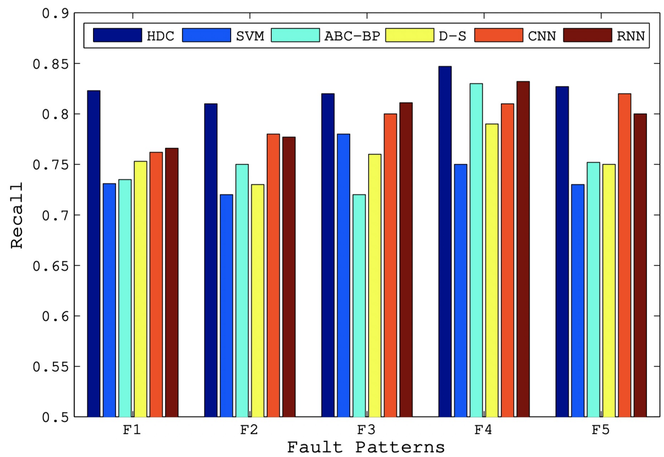

Figure 4: The performance comparison of different approaches.

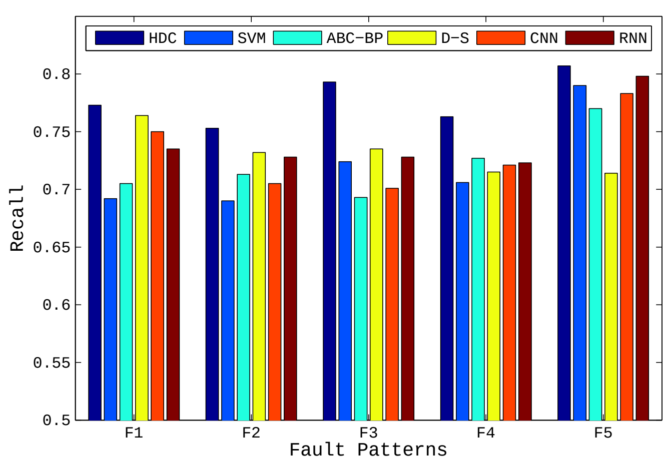

Figure 5: The performance comparison of different approaches with artificial noise.

Table 2: The confusion matrix of HDC

Predicted \ Actual	F_1	F_2	F_3	F_4	F_5	Total_Pre
F_1	247	21	22	2	5	297
F_2	2	243	2	20	9	276
F_3	1	23	246	0	23	293
F_4	26	6	11	254	15	312
F_5	24	7	19	24	248	322
Total_Act	300	300	300	300	300	1500

Table 3: The confusion matrix of HDC with artificial noise

Predicted \ Actual	F_1	F_2	F_3	F_4	F_5	Total_Pre
F_1	232	25	17	35	3	312
F_2	13	226	4	8	23	274
F_3	14	43	238	6	2	303
F_4	23	6	5	229	30	293
F_5	18	0	36	22	242	318
Total_Act	300	300	300	300	300	1500

where p_0 represents the total classification accuracy, p_e is obtained by:

$$p_e = \frac{a_1 \times b_1 + a_1 \times b_1 + \cdots + a_K \times b_K}{N \times N}, \tag{12}$$

where a_i denotes the number of actual samples of class i, b_i denotes the number of predicted samples of class i, K is the number of categories and N is the number of test samples. A high Kappa value indicates a good fault detection model.

Table 2, Table 3 and Figure 6 show the results on the Kappa coefficient. In particular, Table 2 and Table 3 denote the confusion matrices on real data and noisy data, respectively. Figure 6 demonstrates the Kappa coefficients of all the models.

According to the results in the confusion matrices, the HDC Kappa coefficients are 0.7817 and 0.7225 on real data and noisy data, respectively.

As shown in Figure 6, the Kappa coefficients of the proposed model are

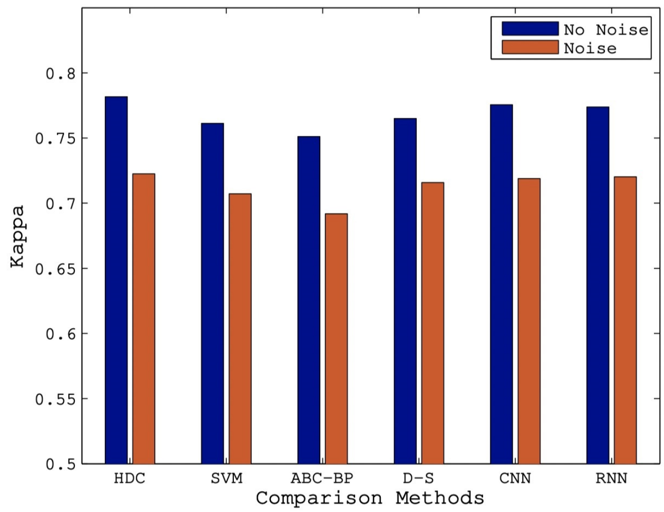

Figure 6: The Kappa coefficients of different methods.

higher than those of the compared models, indicating that HDC can achieve higher stability and reliability.

6. Conclusion

In this paper, a hybrid deep computation (HDC) model is proposed for gas-path fault detection of aero-engines. HDC is composed of a convolution network, a multi-layer perceptron network and a recurrent network. It learns the spatial and temporal dependencies hidden in the gas-path data of aero-engines. Extensive experimental results show that HDC outperforms other common methods in the fault detection task for aero-engines and is more robust to the noise.

7. Acknowledgements

This work was supported in part by the Key Laboratory of Vibration and Control of Aero-Propulsion System Ministry of Education, Northeastern University No. VCAME201705, the National Natural Science Foundation of China under Grants No. 61602083 and 61672123, the Fundamental Research Funds for the Central Universities under Grant No. DUT2017TB02, the Dalian University of Technology Fundamental Research Fund under Grant No. DUT15RC(3)100, and the Doctoral Scientific Research Foundation of Liaoning Province No. 20170520425.

[1] https://en.wikipedia.org/wiki/Southwest_Airlines_Flight_1380, 2018 (accessed 25 August 2019).

[2] W. Xue, T. Wang, Application of a bank of Kalman filters for aircraft engine sensor actuator fault diagnosis, Computer Measurement and Control 23(7)(2015) 2285-2287.

[3] W. Sun, W. Li, W. Li, Avionic devices fault diagnosis based on fusion method of rough set and D-S theory, Journal of Beijing University of Aeronautics and Astronautics 41(10)(2015) 1902-1909.

[4] G. Yu, J. Huang, Aero-engine fault diagnosis based on support vector machine and Kalman filter, Aeroengine 38(1)(2012) 47-50.

[5] S. Zhao, X. Huang, Fault diagnosis for aeroengine gas path components based on neural network multisensory data fusion, Journal of Aerospace Power 23(1)(2008) 163-168.

[6] Q. Zhang, L. T. Yang, Z. Chen, P. Li, A survey on deep learning for big data, Information Fusion 42(2018) 146-157.

[7] J. Gao, P. Li, Z. Chen, A canonical polyadic deep convolutional computation model for big data feature learning in Internet of Things, Future Generation Computer Systems, 99(2019) 508-516.

[8] P. Li, Z. Chen, L. T. Yang, J. Gao, Q. Zhang, M. J. Deen, An Improved Stacked Auto-Encoder for Network Traffic Flow Classification, IEEE Network 32(6)(2018) 22-27.

[9] P. Li, Z. Chen, L. T. Yang, J. Gao, Q. Zhang, M. J. Deen, An Incremental Deep Convolutional Computation Model for Feature Learning on Industrial

Big Data, IEEE Transaction on Industrial Informatics 15(3)(2018) 1341-1349.

[10] F. Lu, J. Huang, X. Qiu, Application of multi-outputs LSSVR by PSO to the aero-engine model, Journal of Systems Engineering and Electronics 20(5)(2009) 1153-1158.

[11] Q. Xu, J. Shi, S. Geng, Aero-engine fault diagnosis by a new fast multi-class support vector algorithm, Journal of Propulsion Technology 33(6)(2012) 961-967.

[12] Q. Huang, G. Zhang, T. Zhang, J. Wang, A kind of approach for aero engine gas path fault diagnosis, in: Proceedings of 2017 IEEE International Conference on Prognostics and Health Management, Dallas, TX, USA, 2017, pp. 55-60.

[13] Y. Pan, Q. Li, Y. Wang, Aero-Engine Gas Path Fault Diagnosis based on Broyden Algorithm, Journal of Propulsion Technology 38(1)(2017) 191-198.

[14] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet Classification with Deep Convolutional Neural Networks, in: Proceedings of Advances in Neural Information Processing Systems 25, Lake Tahoe, Nevada, United States, 2012, pp. 1106-1114.

[15] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 2015, pp. 1-9.

[16] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, in: Proceedings of 3rd International Conference on Learning Representations, San Diego, CA, USA, 2015. arXiv preprint arXiv:1409.1556v6.

[17] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, in: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 770-778.

[18] J. Hu, L. Shen, G. Sun, Squeeze-and-Excitation Networks, in: Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 7132-7141.

[19] G. Jiang, H. He, J. Yan, P. Xie, Multiscale Convolutional Neural Networks for Fault Diagnosis of Wind Turbine Gearbox, IEEE Transactions on Industrial Electronics 66(4)(2019) 3196-3207.

[20] S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, Neural Computation 9(8)(1997) 1735-1780.

[21] K. Cho, B. Merrienboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, Y. Bengio, Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 2014, pp. 1724-1734.

[22] T. Bruin, K. Verbert, R. Babuska, Railway Track Circuit Fault Diagnosis Using Recurrent Neural Networks, IEEE Transaction on Neural Network Learning Systems 28(3)(2017) 523-533.

[23] N. Yoshikawa, N. Belkhir, S. Suzuki, Recurrent neural network-based fault detector for aileron failures of aircraft, in: Proceedings of 11th Asian Control Conference, Gold Coast, Australia, 2017, pp. 1476-1481.

[24] I. M. Elfadel, J. L. Wyatt Jr., The "softmax" nonlinearity: Derivation using statistical mechanics and useful properties as a multiterminal analog circuit element, in: Proceedings of Advances in Neural Information Processing Systems 6, Denver, Colorado, USA, 1993, pp. 882-887.

[25] D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning internal representations by error propagation, in: D. E. Rllmelhart and J. L. McCelland

(Eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol 1: Foundations, MIT Press, Cambridge, MA, 1986.

[26] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86(11)(1998) 2278-2324.

[27] Y. Kim, Convolutional neural networks for sentence classification, in: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 2014, pp. 1746-1751.

[28] X. Li, W Qu, Aero-engine fault diagnosis based on D-S evidence theory, in: Proceedings of 2018 37th Chinese Control Conference, Wuhan, China, 2018, pp. 963-967.