# Data and Prior Knowledge Driven FMEA Severity Assessment via Deep Attention Networks

Hui Huang[1], Jieyu Chen[2], Baihong Zhong[3]
*1)China Electronic Product Reliability and Environmental Testing Research Institute (CEPREI); Key Laboratory of Industrial Equipment Quality Big Data, MITT, Guangzhou 510610, China, 183165263@qq.com.*
*2)China Electronic Product Reliability and Environmental Testing Research Institute (CEPREI); Key Laboratory of Industrial Equipment Quality Big Data, MITT,Guangzhou 510610, China, ceprei_chenjy@163.com (Corresponding author).*
*3)China Electronic Product Reliability and Environmental Testing Research Institute (CEPREI); Key Laboratory of Industrial Equipment Quality Big Data, MITT, Guangzhou 510610, China, China, zhongbh@163.com (Corresponding author).*

**Abstract:** Traditional Failure Mode and Effects Analysis (FMEA) suffers from subjectivity and static assumptions in severity evaluation, especially when applied to complex, data-rich industrial systems. To overcome these limitations, this paper proposes a new dynamic FMEA severity quantification hybrid framework that integrates real-time equipment status monitoring data with domain prior knowledge through deep learning. A deep attention network is proposed to capture long-term dependencies in multivariate sensor streams, while adaptively processing key health status features through attention mechanisms. Engineering prior knowledge is utilized to establish fault severity degrees, integrating real-time equipment condition monitoring data to quantify FMEA severity, thereby enabling a dynamic quantitative representation of FMEA severity. Experimental verification on real industrial datasets shows that the proposed method can achieve dynamic FMEA severity assessment of equipment. This work bridges the gap between predictive health monitoring and equipment risk assessment, providing a practical solution for reliability centered maintenance in data rich industrial environments.

**Keywords:** Failure Mode and Effects Analysis (FMEA); Severity assessment; Attention mechanism; Reliability engineering; Predictive maintenance.

## 1. Introduction

The running reliability of the equipment is decisive for the system safety, production efficiency and economic benefits[1]. The occurrence of equipment failure can readily lead to shutdown and production disruption, thereby causing safety incidents, environmental pollution, significant economic losses, and even posing a threat to personnel safety[2-3]. For instance, within critical infrastructure sectors like power equipment, equipment failures exhibit characteristics of contagion, abruptness, and catastrophic impact, and minor malfunctions may trigger a chain reaction, culminating in severe repercussions such as regional power grid collapse and interruption of energy supply[4]. Consequently, precisely assessing the severity of equipment failures and proactively identifying high-risk failure patterns constitute the essential prerequisite for ensuring the safe and stable operation of equipment.

As a classic reliability analysis and risk assessment method, failure mode and effects analysis (FMEA) systematically identifies potential failure modes in equipment, analyzes the causes and impacts of these failures, quantitatively evaluates the severity, occurrence probability, and detectability of failures, and calculates the risk priority number (RPN)[5-6], which provides a basis for failure prevention and maintenance decision-making. FMEA has been extensively applied across all stages of the equipment life-cycle, emerging as a crucial method for enhancing equipment reliability and safety. However, traditional FMEA severity assessments heavily rely on expert experience and qualitative judgment, leading to subjective and inconsistent results. Furthermore, traditional FMEA, being based on a static, experience-driven approach, struggles to accommodate the operational characteristics of dynamic systems such as power equipment. It fails to track the real-time evolution of equipment health status, resulting in delayed severity assessments and inadequate targeting of preventive maintenance, thereby falling short of meeting the real-time risk management and control requirements of dynamic systems.

With the rapid advancement of sensor technology and big data analytics, the comprehensive collection and deep mining of equipment operational data have become feasible. The data-driven approach offers a novel avenue to overcome the limitations inherent in traditional FMEA, and it has been successfully implemented in various scenarios, including fault diagnosis, health assessment, and life prediction[7-9]. By extracting latent fault feature and health status information from the operational data, dynamic perception and quantitative evaluation of equipment conditions can be achieved. By integrating the data-driven approach with FMEA theory, a data-driven dynamic severity assessment method for FMEA can be established, which compensates for the deficiencies of traditional FMEA methods in dynamic system applications.

This paper introduces a FMEA severity assessment method grounded in a deep attention network, designed to surmount the constraints of traditional approaches that rely heavily on experience and static evaluation, and to offer a novel technical pathway for quantifying equipment risks within dynamic systems. This method employs the deep attention network to thoroughly analyze the time-series signals generated during equipment operation, precisely identify the degradation patterns of health status, and automatically highlight critical features closely associated with the progression of health status degradation through the attention mechanism, thereby enhancing the precision and efficacy of feature extraction. Building upon this, the FMEA severity degrees are classified by integrating domain-specific prior knowledge, establishing a data-driven FMEA severity assessment framework that ultimately enables dynamic and accurate evaluation of FMEA severity degrees.

The rest of this article is as follows. The second section introduces the FMEA severity assessment process and highlights the limitations inherent in traditional methods. The third section elucidates the principles underlying the proposed method, followed by experimental analysis and discussion in the fourth section. Finally, the fifth section summarizes the contributions of this paper.

## 2. FMEA Severity Assessment

Severity assessment constitutes a pivotal element of Failure Mode and Effects Analysis (FMEA), serving to quantify the extent of impact a particular failure mode exerts on system functionality, safety, economic aspects, and other relevant dimensions. The fundamental procedure typically encompasses the identification of potential failure modes, an analysis of the ultimate repercussions of the failure on the system and its downstream processes, and the classification of these consequences according to predetermined evaluation criteria (e.g., a scale ranging from 1 to 10)[10]. Traditional evaluation methodologies predominantly hinge on expert experience, industry benchmarks, and historical fault cases to make qualitative assessments. By employing expert scoring or integrating techniques such as the Analytic Hierarchy Process (AHP) and fuzzy comprehensive evaluation, subjective judgments are converted into severity scores, and these scores are subsequently utilized to compute the Risk Priority Number (RPN) and ascertain the priority of risks.

However, traditional FMEA severity assessment methods exhibit notable deficiencies when applied to dynamic systems. Given that traditional FMEA typically relies on static experience and offline analysis, it fails to integrate multi-source monitoring data in real-time during equipment operation, thereby hindering the dynamic tracking of equipment health status evolution trends. Consequently, severity assessment results lag behind actual changes in operating conditions. This static evaluation approach is ineffective in identifying early fault indicators, rendering preventive maintenance strategies lacking in specificity and timeliness, and struggling to meet the demands of complex dynamic systems for real-time risk alerts and precise maintenance. Therefore, there is an urgent need to develop a data-driven FMEA severity assessment method that can dynamically and objectively quantify fault severity by extracting health features from equipment operation data, thereby enhancing system operational reliability and the intelligence level of maintenance decision-making.

## 3. Severity Assessment of FMEA Based on Deep Attention Network

To address the limitations of traditional FMEA methods in dynamic system applications, this article introduces a FMEA severity assessment approach based on the Deep Temporal Attention Network (DTAN), as illustrated in Figure 1. This method employs DTAN to extract health state degradation information from time-series signals and integrates domain prior knowledge to classify FMEA severity degrees, thereby enabling dynamic and precise quantitative evaluation of equipment fault severity.

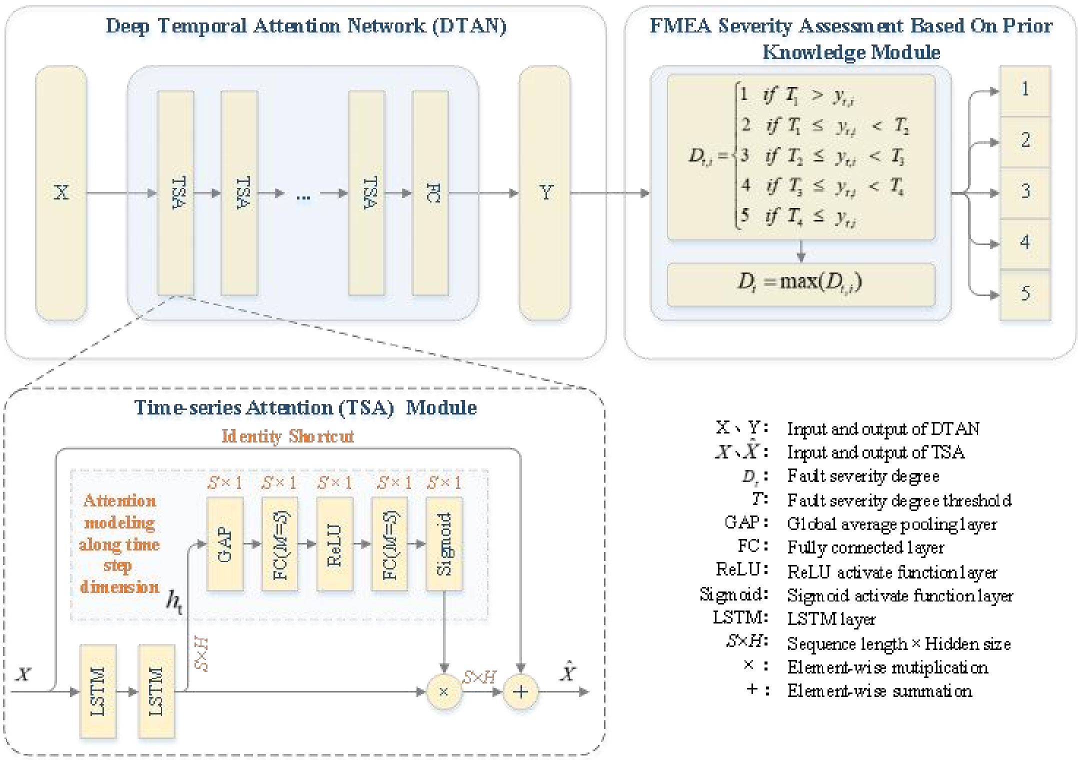


*Figure 1.* FMEA severity assessment framework based on deep attention mechanism

DTAN utilizes multi-dimensional temporal signals as input and employs multiple stacked Time-series Attention (TSA) modules to model the long-term dependencies of these signals across both time steps and sensors. By adaptively applying attention weighting to different time steps, the TSA module can effectively capture key patterns that are highly correlated with the degradation evolution of equipment health status, thereby generating a more robust health status representation for subsequent severity assessment. The FMEA severity assessment based on prior knowledge module, aims to organically combine data-driven health status representations with prior rules from the engineering domain, which maps the output of DTAN to severity degrees determined by prior knowledge of both alarm thresholds and statistical analysis thresholds, achieving an objective and refined quantification of FMEA severity degrees. The remainder of this section will provide a detailed introduction to the implementation details of DTAN and the FMEA severity degree setting strategy that incorporates prior knowledge.

### 3.1. CONSTRUCTION OF DEEP TEMPORAL ATTENTION NETWORK

DTAN primarily consists of an input layer, multiple TSA modules, and a fully connected (FC) output layer. Its core lies in leveraging multiple TSA modules to perform cross-time-step and cross-signal-source correlation modeling on the input data, thereby enabling precise mining of health status. Among these components, the TSA module is mainly composed of two long short-term memory (LSTM) layers, one global average pooling (GAP) layer, two FC layers, one rectified linear unit activation function (ReLU) layer, one Sigmoid layer, and an identity shortcut. It utilizes LSTM to capture the long-term dependencies of multidimensional time-series signals and achieve cross-signal-source correlation modeling. Furthermore, by conducting attention modeling on the output features of the two LSTM layers in the time-step dimension, the deep model is encouraged to focus more on the critical information embedded in different time steps, thereby mining information that is strongly correlated with the equipment's health status.

The TSA module selectively memorizes and updates multi-dimensional time-series signals by utilizing the gating mechanism (forget gate, input gate, output gate) of LSTM, which can effectively extract dynamic features from each signal dimension while maintaining long-term temporal dependencies through the cell state, laying a foundation for temporal representation in cross-variable (cross-signal source) association modeling. The LSTM cell structure is shown in Figure 2, and the specific implementation process is described as follows.

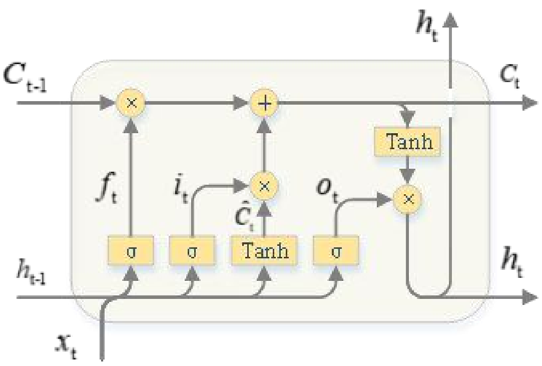


*Figure 2*. Schematic diagram of LSTM cell structure

The forget gate $f_t$ is used to determine the retention and discarding of historical information $C_{t-1}$; the input gate $i_t$ is used to filter useful information from the current input $x_t$ and incorporate it into the cell state $\hat{C}_t$; the output gate $o_t$ is used to control the output of useful information from the current cell state to subsequent cells[11-12]. The working principle of an LSTM cell can be described by the following formulas.

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\hat{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3}$$

$$C_t = f_t \times C_{t-1} + i_t \times \hat{C}_t \tag{4}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t \times \tanh(C_t) \tag{6}$$

where $\sigma(\cdot)$ is the Sigmoid activation function, with a range of (0,1); $h_{t-1}$ and $C_{t-1}$ are respectively the output and the cell state of the LSTM cell at the ($t$-1)-th time step; $h_t$, $\hat{C}_t$, and $C_t$ are respectively the output, the candidate cell state, and the final cell state of the LSTM cell at the $t$th time step; $x_t$ is the input information at the $t$-th time step; $W_f$, $W_i$, $W_C$, $W_o$, and $b_f$, $b_i$, $b_C$, $b_o$ are respectively the weight parameters and bias terms of the forget gate $f_t$, input gate $i_t$, candidate state $\hat{C}_t$, and output gate $o_t$; $\tanh(\cdot)$ is the tanh activation function, with a range of (-1,1).

First, the TSA module uses two LSTM layers to model the input $X$, obtaining the output feature vector $h_t$, as shown in the following formula.

$$h_t = \mathrm{LSTM}(X) \tag{7}$$

Then, to obtain the correlation between feature vectors at different time steps, GAP operations are performed on the LSTM output $h_t$ along the time step dimension, transforming the information carried by each time step into the corresponding descriptive vector, with the calculation process shown as follows.

$$g_s = \frac{1}{H} \sum_{i=1}^{H} h_t(i) \tag{8}$$

where $g$ is the description vector obtained on the time step dimension through GAP operation; $g_s$ is the $s$-th feature of $g$, and $s$=1,2,..., $S$, where $S$ is the number of time steps; $t$ is the time step index number of $h_t$, and $H$ is the output feature dimension of LSTM cell.

Subsequently, to effectively capture the important information implied between different time steps, two FC layers with ReLU activation functions are employed to explicitly model the dependencies between different time steps, with the calculation process shown as follows.

$$f = W_2 \cdot \delta(W_1 \cdot g + b_1) + b_2 \tag{9}$$

where $f$ is the feature vector outputted by two FC layers; $W_1$ and $W_2$ , $b_1$ and $b_2$ are the weight parameters and bias terms of the first and second FC layers, respectively; $\delta(\cdot)$ is the ReLU activation function.

After that, the Sigmoid function is used to convert $f$ to (0,1) to obtain the attention weights $\gamma$ corresponding to the feature vector $h_t$ at each time step, as shown in the following equation.

$$\gamma = \frac{1}{1+\exp(-f)} \tag{10}$$

On this basis, the feature vector $h_t$ is readjusted by attention weights $\gamma$ to highlight the important timing information in the feature vector $h_t$, as shown in the following equation.

$$\hat{h}_t = h_t \cdot \gamma \tag{11}$$

Finally, for each TSA module, its output result can be calculated using the following equation.

$$\hat{X} = X + \hat{h}_t \tag{12}$$

It is worth noting that the number $M$ of neurons in the FC layer, is set to $S$ (i.e., $M = S$, where $S$ is the number of time steps in the feature vector $h_t$ ), which is done to ensure that the number of attention weights obtained is equal to the time steps length of the feature vector $h_t$. This allows for flexible setting of weights at each time step to extract important features related to equipment health status. Overall, the TSA module introduces an attention mechanism along the time step dimension, guiding deep models to dynamically focus on the most critical moments in the sequence relevant to the current task, which enables more effective mining of long-term dependencies and key temporal features, enhancing the deep model's ability to model complex temporal patterns.

### 3.2. FMEA SEVERITY ASSESSMENT BASED ON PRIOR KNOWLEDGE

The core idea of the FMEA severity assessment based on prior knowledge module lies in the combination of data-driven statistical characteristics and engineering domain prior rules. The prior knowledge utilized in this article mainly includes two types: one is the out-of-limit alarm threshold defined in the condition monitoring system (reflecting the engineering safety boundary), and the other is the number of FMEA severity degrees determined based on engineering application requirements. Typically, out-of-limit alarm thresholds are provided by design documents, maintenance manuals, etc., and in engineering applications, they often need to be adjusted based on actual conditions. The number of FMEA severity degrees is determined by actual engineering requirements, and this article divides them into 5 degrees: 1, 2, 3, 4, and 5. Among them, degree 1 indicates that the equipment is in a healthy and fault-free state, while degree 5 indicates that the equipment is in a severe fault state. For high-reliability equipment, when its monitoring signal amplitude is out of limit,, it is highly likely to trigger severe accidents. Therefore, this article defines the severity degree corresponding to the out-of-limit alarm state of the equipment monitoring signal as degree 5. On the basis of determining the out-of-limit alarm thresholds and the number of FMEA severity degrees, it is still necessary to determine the severity degree thresholds to evaluate the severity degree in which the equipment is.

The equipment monitoring signals involved in this article have an alarm upper limit, and the severity degree threshold is determined by box plot analysis based on actual engineering experience. Specifically, first, box plot analysis is employed to statistically model multi-dimensional time series signals and extract their quartile statistics. Then, the third quartile is used as the initial severity degree threshold for FMEA. Subsequently, the interval between this initial severity degree threshold and the alarm upper limit is evenly divided to generate three equally spaced subintervals, thereby forming five fault severity degrees ranging from minor to severe, which ultimately achieves the quantitative classification and refined discrimination of fault severity. The FMEA severity degree classification is shown in the following formula.

$$D_{t,i} = \begin{cases} 1 & if\ T_1 > y_{t,i} \\ 2 & if\ T_1 \le y_{t,i} < T_2 \\ 3 & if\ T_2 \le y_{t,i} < T_3 \\ 4 & if\ T_3 \le y_{t,i} < T_4 \\ 5 & if\ T_4 \le y_{t,i} \end{cases} \tag{13}$$

where $D_{t,i}$ is the fault severity degree corresponding to $y_{t,i}$ at *t*-th moment; $y_{t,i}$ is the *t*-th output value of Y (i.e., the output of DTAN) at *t*-th moment. $T_1$ , $T_2$ , $T_3$ and $T_4$ are respectively the thresholds corresponding to different fault severity degrees, where $T_4$ corresponds to the upper limit of monitoring data alarm. $T_1$ , $T_2$ , and $T_3$ are determined by the box plot analysis results and the alarm upper limit, and their calculation formula is as follows.

$$T_1 = Q_{\frac{3}{4}},\ \ T_2 = Q_{\frac{3}{4}} + \frac{1}{3}\nabla,\ T_3 = Q_{\frac{3}{4}} + \frac{2}{3}\nabla,\ \ T_4 = Q_{\frac{3}{4}} + \nabla = T_{\mathrm{alarm}} \tag{14}$$

where $Q_{\frac{3}{4}}$ is the third quartile of monitoring data sequence; $\nabla = T_{\mathrm{alarm}} - Q_{\frac{3}{4}}$, $T_{\mathrm{alarm}}$ is the alarm upper limit.

For the presence of multiple monitoring data sequences, the highest severity degree is taken as the equipment FMEA severity degree, i.e., $D_t = \max(D_{t,i})$.

## 4. Experimental Results and Discussion

This section first introduces the dataset used in the experiment, then describes the experimental methods and hyperparameter settings, and finally discusses and analyzes the experimental results.

### 4.1. DATASET DESCRIPTION

The dataset used in the experiment is from the real operation data of coal-fired power plant unit. Taking the motor fault of induced draft fan as an example, the temperature sensor data of six key measuring points of induced draft fan motor were collected from the coal-fired unit monitoring system for application verification. The six key measuring points are phase A temperature (2 monitoring points), phase B temperature (2 monitoring points), and phase C temperature (2 monitoring points) of induced draft fan motor stator winding. The alarm upper limit at each monitoring point is 90 ℃ . The data covers the continuous operation data of a month in 2025. The sampling interval is 30 seconds, and a total of 89280 time series records are obtained. In this article, the non overlapping sliding window strategy is used to construct the motor health state dataset of induced draft fan, which is used to train and evaluate the predictive ability of the deep model on the change trend of equipment health state. The window length is set to 10 time steps (i.e., covering the 5-minute running period), and a total of 8928 samples are generated. These samples were randomly divided into training set and test set according to the ratio of 4:1. Each sample is a multivariate time sequence segment with a size of $10 \times 6$. For each sample, the first nine time steps are used as the model input (i.e., the model input size is $9 \times 6$), and the temperature value of the tenth time step is used as the target output (the corresponding model output size is $1 \times 6$). This strategy can promote the deep model to learn the temporal evolution rule of equipment degradation, and provide reliable health status representation for subsequent severity assessment.

## 4.2. EXPERIMENTAL METHODS AND HYPERPARAMETER CONFIGURATION

In order to verify the usefulness of the proposed deep model, comparative experiments of four methods were carried out. The four methods are Single-LSTM, Multi-LSTM, Res-LSTM, and DTAN. The difference between DTAN and Res-LSTM is that DTAN adds attention modeling mechanism, while Res-LSTM does not. The difference between Res-LSTM and Multi-LSTM is that Res-LSTM has an identity shortcut, while Multi-LSTM does not. Multi-LSTM is mainly formed by stacking multiple LSTM layers, while Single-LSTM is a single LSTM layer. Table 1 shows the hyperparameter settings related to the DTAN structure. Wherein, TSA represents the time-series attention module. In the table, the hyperparameters of LSTM layer and FC layer in TSA are given in brackets. If there are three parameters in brackets, they are LSTM layer parameters, representing the number of input features, the number of hidden state features and the number of LSTM cells. If there are two parameters in brackets, they respectively represent the number of input and output neurons in the full connection layer. Each TSA has two FC layers with the same number of input and output neurons. The output feature map is expressed in two dimensions (that is, a tensor expressed in the form of "sequence length $\times$ hidden size") or a vector. For each LSTM layer, the hyperparameters of mult-LSTM, Res-LSTM and DTAN are consistent, while the LSTM layer parameter in Single-LSTM is set to [6,16,1].

Table I. Hyperparameters related to the structure of DTAN

| DTAN | LSTM/FC | Size of output feature map |
|---|---|---|
| Input | / | 9×6 |
| TSA_1 | [6,16,2]<br>[9,9] | 9×16 |
| TSA_2 | [16,16,2]<br>[9,9] | 9×16 |
| TSA_3 | [16,16,2]<br>[9,9] | 9×16 |
| TSA_4 | [16,16,2]<br>[9,9] | 9×16 |
| FC | [16,6] | 6 |

The training related hyperparameters are set as follows. The total number of iterations is set to 100. At the beginning of the training, the learning rate is set to 0.1. After each 30 iterations, the learning rate is reduced to one tenth of the original. The batch size is set to 128. All experiments used Adam as the optimizer, and the weight attenuation value was set to 0.0001.

In addition, all deep models use mean square error (MSE) as the loss function in the training stage. In order to eliminate the influence of random factors, all experiments were repeated for 5 times. The hyperparameters of all deep models remained unchanged in the experiment. All deep models are implemented with the PyTorch (version 1.8.0) open source framework. All experiments were conducted on a computer configured with Intel(R) Core(TM) i7-10700 CPU @ 2.90GHz and NVIDIA GeForce GTX 1060Ti GPU.

### 4.3. EXPERIMENTAL AND RESULTS DISCUSSION

#### *4.3.1. Performance comparison on health state prediction*

Table 2 shows the mean square error (MSE) results of the four methods considered, i.e., Single-LSTM, Multi-LSTM, res-LSTM and DTAN (the method proposed in this article), which are repeated five times on the same test set.

Table II. MSE of different methods on test set

| Methods | Test 1 | Test 2 | Test 3 | Test 4 | Test 5 | Mean |
|---|---|---|---|---|---|---|
| Single-LSTM | 0.0399 | 0.0203 | 0.0141 | 0.0108 | 0.0118 | 0.0194±0.0121 |
| Multi-LSTM | 8.7229 | 0.5966 | 0.0271 | 0.0091 | 0.0133 | 1.8738±3.837 |
| Res-LSTM | 0.0072 | 0.0067 | 0.0064 | 0.0051 | 0.0041 | 0.0059±0.0013 |
| DTAN | 0.0070 | 0.0059 | 0.0046 | 0.0044 | 0.0046 | **0.0053±0.0011** |

It can be seen that the proposed DTAN achieves the lowest average MSE of 0.0053, which is significantly better than all the comparison models. The Single-LSTM exhibits high volatility (mean=0.0194 ± 0.0121), indicating its weak generalization ability, which stems from its single-layer structure finding it difficult to effectively model complex temporal dependencies in multivariate sensor signals. Generally, the feature extraction capability of deep networks can be enhanced by stacking multiple LSTM layers. However, the average MSE of Multi-LSTM is significantly lower than that of Single-LSTM, and its training process stability is poor, such as the MSE is as high as 8.7229 in the first test and the average MSE standard deviation is as high as 3.837, which may be due to the gradient disappearance or over fitting caused by the lack of identity shortcut, resulting in serious performance degradation in individual experiments. Instead, Res-LSTM effectively alleviates the difficulty of gradient propagation by introducing identity shortcut, significantly improving the training stability of deep networks, with an average MSE of 0.0059 ± 0.0013, which is significantly better than Single-LSTM and Multi-LSTM, confirming the effectiveness of identity shortcut in deep time series modeling. The proposed DTAN can obtain the best test performance is mainly because DTAN further introduces the time-series attention (TSA) mechanism on the basis of Res-LSTM, which dynamically weights the information of key time steps, so that the deep network focuses more on the pattern highly related to health degradation, so as to obtain better test performance than Res-LSTM. These experimental results show that the proposed DTAN can more accurately capture the degradation features of equipment health status, and realize the accurate characterization of equipment health status.

#### *4.3.2. FMEA severity assessment*

In an induced draft fan motor fault alarm, the corresponding sensor data was retrieved and it was found that the sensor monitoring data exceeded the alarm threshold at the corresponding time of the fault alarm, as shown in Figure 3, which led to the alarm signal.

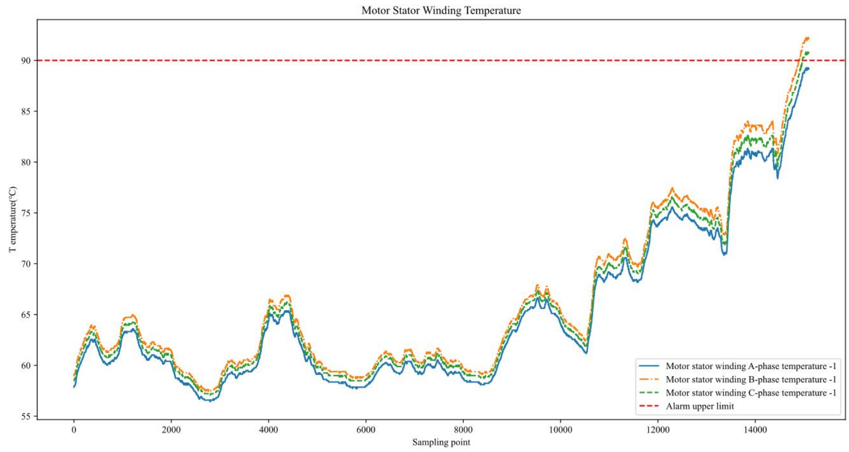


*Figure 3*. Variation curve of sensor monitoring data

The proposed DTAN is used for predictive FMEA severity assessment, and the assessment results are shown in Figure 4. It can be seen that before the first 12200 sampling point, the motor fault severity degree of induced draft fan is 1, indicating that the equipment is in normal operation state. From 12200 sampling point to 14672 sampling point, the motor fault severity degree of induced draft fan has changed from degree 2 to degree 4, showing a process from early fault symptoms to aggravation of fault severity. At 14672 sampling point, the motor fault severity degree of the induced draft fan reached degree 5, which was consistent with the time when the stator winding temperature of the motor exceeded the alarm upper limit, and coincided with the time point of the fault alarm signal. The above results show that the proposed method can accurately quantify the  motor fault severity of induced draft fan, accurately identify the actual time of fault occurrence. More importantly, the proposed method can alert potential risks through the degree 3/4 fault severity warning before the fault occurs, striving for a valuable intervention window for the operation and maintenance personnel.

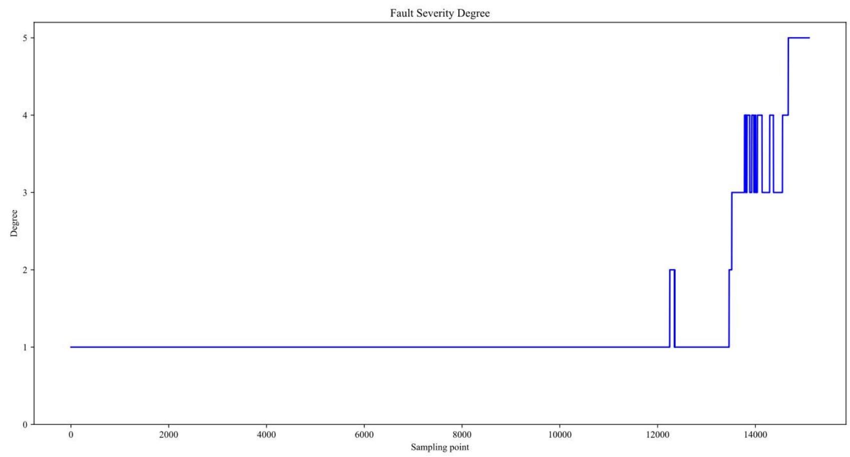


*Figure 4.* Evaluation results of fault severity degree

## 5. Conclusion

This article presents a dynamic quantitative evaluation method of equipment FMEA severity based on deep temporal attention network (DTAN). By stacking multiple time-series attention (TSA) modules, DTAN models the long-term dependencies of multidimensional monitoring signals, and uses the attention mechanism to dynamically identify the key features highly related to health degradation in the time step dimension, so as to achieve accurate characterization of equipment health status. The engineering prior knowledge is creatively integrated, combining statistical analysis feature values with alarm thresholds of operation and maintenance system to classify fault severity degrees, and establishing a mapping mechanism from health status to FMEA severity degrees, ultimately achieving dynamic, objective, and refined quantification of fault severity.The experimental results on the real motor health state dataset of coal-fired power plant unit show that the proposed method not only achieves excellent performance in the accuracy and stability of health state prediction, but also successfully realizes FMEA severity evaluation based on time-series monitoring signals, effectively solving the core problems of traditional FMEA severity scoring subjectivity, static solidification, and difficulty in linking with real-time operating status. The proposed method provides a feasible path for FMEA to shift from static experience driven to dynamic data knowledge dual driven paradigm, which has important engineering application prospects for promoting intelligent operation and maintenance, and improving equipment reliability and safety.

## Acknowledgements

This work was supported by the Special Fund of the 5th Electronics Research Institute of MIIT (Grant No.25Z11, No.24Z03, No.25Z03).

## References

G. Li, X. Wang and C. Zuo. Research on multi-source data combined with fault diagnosis and prediction algorithm of power equipment. 2025 IEEE 7th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Yinchuan, China, 2025, 1031-1034.

H. Han, H. Li, X. Wu, H. Yang and J. Qiao. Generalized Mapping Fault Diagnosis for Industrial Machine Based on Penalty Gating LSTM. IEEE Transactions on Industrial Informatics, 2025, 21(11): 9159-9170.

B. Zhong, M. Zhao, S. Zhong, et al. Deep exponential excitation networks: toward stronger attention mechanism for weak fault diagnosis. Structural Health Monitoring, 2024, 23(6): 3850-3866.

J. Wang, H.Mokhlis, N.Mansor, et al. Smart Fault Detection, Classification, and Localization in Distribution Networks: AI-Driven Approaches and Emerging Technologies. IEEE Access, 2025,. 13: 141664-141693.

A. Kadechkar and H. Grigoryan. FMEA 2.0: Machine Learning Applications in Smart Microgrid Risk Assessment. 2024 12th International Conference on Smart Grid (icSmartGrid), Setubal, Portugal, 2024, 629-635.

J. Shi, Z. Yang, C. Zhaoxsx and H. Zhou. A Multi-Condition Fault Diagnosis Logic Generation Method Based on Extended FMEA. 2025 7th International Conference on System Reliability and Safety Engineering (SRSE), Changchun, China, 2025, 107-111.

B. Zhong, M. Zhao, S. Zhong, et al. Mechanical compound fault diagnosis via suppressing intra-class dispersions: A deep progressive shrinkage perspective. Measurement, 2022, 199.

B. Zhong, M. Zhao, L. Wang, et al. DCSN: Focusing on hard samples mining in small-sample fault diagnosis of marine engine. Measurement, 2024, 235.

S. Fu, L. Lin, Y. Wang, et al. MCA-DTCN: A novel dual-task temporal convolutional network with multi-channel attention for first prediction time detection and remaining useful life prediction. Reliability Engineering & System Safety, 2024, 241.

Y. Tian and X. Zhang. Research on the Application of FMEA in Civil Aircraft Safety Analysis. 2024 Global Reliability and Prognostics and Health Management Conference, Beijing, China, 2024, 1-5.

M. Ahsan, M. Waqar Hassan, J. Rodriguez and M. Abdelrahem. Enhanced Fault Diagnosis in Rotating Machinery Using a Hybrid

CWT-LeNet-5-LSTM Model: Performance Across Various Load Conditions. IEEE Access, 2025, 13: 1026-1045.
Y. Pi, Y. Chen, S. Wang, et al. Development and Engineering Application of Fault Early Warning Model for Large Hydropower Turbine Based on Principal Component Analysis and Long Short-Term Memory Network. IEEE Sensors Journal, 2025, 25(16): 30470-30483.