Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio

doi: 10.3850/978-981-18-8071-1_P691-cd

Knowledge and Data Fusion-driven for Offshore Wind Turbine Gearbox Fault Diagnosis

Qingqing Xu

College of Safety and Ocean Engineering, China University of Petroleum (Beijing), China. E-mail: xuqq@cup.edu.cn

Hao Liu

College of Safety and Ocean Engineering, China University of Petroleum (Beijing), China. E-mail: liuhao@student.cup.edu.cn

Yingchun Ye

College of Safety and Ocean Engineering, China University of Petroleum (Beijing), China. E-mail: yeyingchun@cup.edu.cn

Laibin Zhang

College of Safety and Ocean Engineering, China University of Petroleum (Beijing), China. E-mail: zhanglb@cup.edu.cn

Taotao Zhou

College of Safety and Ocean Engineering, China University of Petroleum (Beijing), China. E-mail: zhoutt@cup.edu.cn

At present, the cumulative installed capacity of offshore wind power in China has reached 27.26 million kilowatts, which promotes the clean and low-carbon transformation of energy and helps China achieve the goal of "carbon peak and carbon neutrality". However, the gearbox of offshore wind turbine is affected by its structure, working condition and environment, which leads to high gearbox failure. Traditional fault diagnosis methods are difficult to meet the requirements of offshore wind turbine equipment diagnosis with variable working conditions and multiple fault types. This paper proposes a knowledge and data fusion-driven fault diagnosis method for offshore wind turbine gearbox. It not only classifies the operating data of offshore wind turbine gearbox through convolutional neural networks (CNN), but also uses knowledge graph to display detailed information of faults and perform intelligent question answering. The innovation of this method is that the fault diagnosis method based on data monitoring and the fault type reasoning using knowledge graph are combined to form a comprehensive fault diagnosis model of offshore wind turbine gearbox. This method is applied to the fault diagnosis of gearboxes in Jiangsu offshore wind farms. For the types of cracks, wear and missing teeth of offshore wind turbine gearbox, the accuracy of fault diagnosis based on convolutional neural network model reaches 95%. At the same time, the visual display and intelligent question answering of offshore wind turbine gearbox faults are realized by using the constructed knowledge graph. The results show that the knowledge and data fusion-driven fault diagnosis method for offshore wind turbine gearbox has a good application effect on the intelligent operation and maintenance of offshore wind turbines.

Keywords: Offshore wind turbine gearbox, knowledge and data fusion-driven, fault diagnosis, convolutional neural network, knowledge graph.

1. Introduction

As a clean and sustainable energy, offshore wind power has great development prospects. In 2022, China's electricity consumption reached 8637.2 billion kilowatt hours, an increase of 3.6% year-on-year. The development of offshore wind power can meet the demand of electricity better, and it is also one of the important measures to achieve the goal of "carbon peak and carbon neutrality" and promote the clean and low-carbon transformation of energy.

As the gearbox of offshore wind turbines is an important component of the transmission mechanism of wind turbines, research on gearbox fault diagnosis is of great significance for ensuring the safe and reliable operation of the entire unit. At present, the fault diagnosis of gearboxes is mainly divided into two categories: data-based and knowledge-based (Long et al., 2017). The data-based method is from the data point of view, in-depth mining of fault information and feature extraction, so as to achieve the purpose of fault diagnosis. Common methods include signal processing and artificial intelligence.

Signal processing methods usually include time domain, frequency domain and time-frequency analysis. Time domain analysis is a method of statistical analysis of data in the time domain. The main indicators include peak, peak-to-peak value, variance, effective value, kurtosis and skewness (He et al., 2010). The time domain indicators are convenient to use and is often used in the early diagnosis of gearbox faults, but it cannot judge the specific type of faults. Frequency domain analysis is a signal analysis method based on Fourier transform. Common spectrograms include spectrum, cepstrum, envelope spectrum and order spectrum. Different faults have different frequency components in the frequency domain. According to the composition and amplitude of these frequency components, gearbox faults can be identified and diagnosed. However, for non-stationary signals, it is necessary to analyze the change of spectrum with time. At this time, frequency domain analysis is no longer applicable, and time-frequency analysis has unique advantages for non-stationary signal analysis. Time-frequency analysis methods mainly include short-time Fourier transform, wavelet transform, empirical mode decomposition and variational mode decomposition. However, this method still has the problem of poor adaptability, and the diagnostic results rely heavily on manual experience for judgment.

As a tool of pattern recognition, artificial intelligence has a wide range of applications in the field of fault diagnosis. It takes the results of data preprocessing as input to match and identify fault modes. Artificial intelligence methods mainly include neural network, support vector machine and fuzzy theory. As a nonlinear machine learning network, neural network can be used as both feature extractor and classifier. Taking convolutional neural network as an example, through the constructed CNN model of gearbox fault diagnosis, the extraction and recognition of gearbox fault features can be realized at the same time, and a good diagnosis effect can be obtained (Jing et al., 2017). However, the usual neural network training requires a lot of data, and there are some deficiencies in the processing of small sample problems. Support vector machine has unique advantages in dealing with small sample and nonlinear problems (Cortes and Vapnik, 1995). The fuzzy theory can judge the state of the equipment through the membership degree and provide help for fault diagnosis (Li et al., 2013).

Knowledge-based methods use existing expert knowledge to analyze and diagnose faults, which mainly includes fault tree and expert system. Fault tree analysis is a deductive analysis method. It starts from the state of the fault and infers and predicts the type and probability of the fault step by step (Lee et al., 1985). Fault diagnosis based on expert system uses the experience and knowledge of related fields to build a knowledge base, and judges the fault through computer simulation of expert reasoning and decision-making process (Guo, 2002).

The above methods have good application effects in the fault diagnosis of wind turbine gearboxes, but the diagnosis methods using data or knowledge alone have the problems in complex working conditions and data association (Liu et al., 2022). Furthermore, the current research mainly focuses on data-based fault diagnosis methods, which cannot effectively use the existing prior knowledge, resulting in waste of unstructured knowledge such as fault diagnosis manuals (Nie et al., 2022).

In order to solve the above problems, this paper proposes a knowledge and data fusion-driven fault diagnosis method, which comprehensively utilizes vibration data and expert knowledge to diagnose and intelligently answer gearbox faults. The main innovations are as follows:

- A comprehensive fault diagnosis model of offshore wind turbine gearbox is formed by fusing the fault diagnosis method based on data monitoring and the fault type question answering using knowledge graph.

- Aiming at the complex operating conditions of variable speed and load changes in the gearbox of offshore wind turbines, as well as multiple fault types such as cracks, wear, and missing teeth, the bispectrum with high frequency domain resolution is selected as the input of the convolutional neural network to achieve fault diagnosis of the gearbox under variable speed operating conditions
- Combining the deep learning algorithm with the construction of knowledge graph, the automatic extraction of entities and relationships is realized, which effectively solves the problems of strong dependence on labor and high cost in the construction of traditional rule-based knowledge base.

The research contents of this paper are as follows: Section 2 introduces the knowledge and data fusion-driven fault diagnosis method. Section 3 presents the convolutional neural network fault diagnosis model. Section 4 recommends the knowledge graph model which fuses expert knowledge. Section 5 shows the application results of the knowledge and data fusion-driven fault diagnosis method for offshore wind turbine gearbox.

2. Knowledge and Data Fusion-driven Fault Diagnosis Method

The framework of the knowledge and data fusion-driven fault diagnosis method is shown in Fig. 1, which is divided into two parts: CNN fault diagnosis model and knowledge graph model.

For the CNN-based fault diagnosis model, this paper summarizes it into the following three steps:

- Data collection. A large number of gearbox vibration data are collected from the field or experimental platform, and the data are preprocessed.
- Feature extraction. The one-dimensional vibration signal is transformed into a time-frequency graph (such as bispectrum) with higher frequency domain resolution, and then the convolutional neural network is used to extract the fault features.
- Fault identification. According to the hierarchical features after convolution training, the state of the gearbox is adaptively described and matched with the fault type.

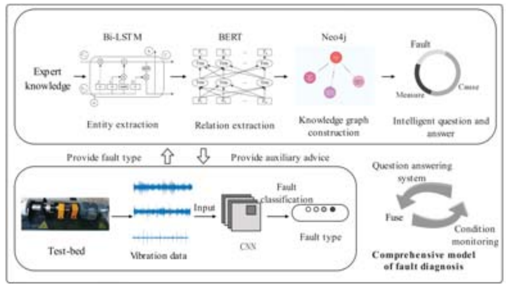


Fig. 1 Framework of the knowledge and data fusion-driven fault diagnosis method

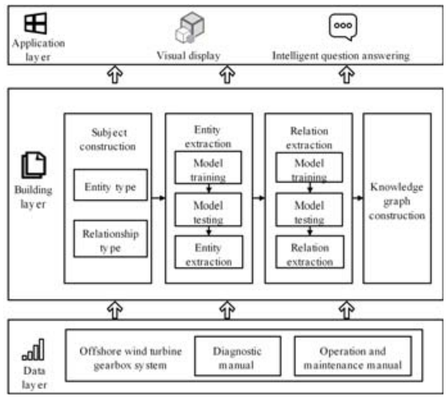


Fig. 2. Construction and application of fault diagnosis knowledge graph for offshore wind turbine gearbox system

The construction and application of knowledge graph for fault diagnosis mainly include: subject construction, entity extraction based on bidirectional long short-term memory (Bi-LSTM), relation extraction based on bidirectional encoder representations for transformers (BERT) pre-training model, visual display and intelligent question answering based on knowledge graph. The knowledge graph model includes three layers: data, building and application. Taking the offshore wind turbine gearbox system as an example, the structure of the fault diagnosis knowledge graph is shown in Fig. 2.

3. CNN Fault Diagnosis Model for Offshore Wind Turbine Gearbox

3.1. *CNN Fault Diagnosis Model*

CNN has the characteristics of hierarchical extraction learning, which can abstract images into specific features, so as to realize image classification. The typical CNN structure is shown in Fig. 3, which mainly includes input layer, convolution layer, pooling layer, fully connected layer and output layer. In the fault diagnosis of offshore wind turbine gearbox, the input layer corresponds to the vibration time-frequency diagram of the planetary gear fault, and the output layer corresponds to the fault type. The fault diagnosis is achieved by the classification of fault time-frequency diagram.

Fig. 3. Convolutional neural network model structure

3.2. *Bispectrum Analysis*

The time-frequency analysis methods of vibration signals mainly include short-time fourier transform time-frequency analysis, wavelet time-frequency analysis and bispectrum analysis. In the comparative analysis of the frequency domain resolution of different faults, the bispectrum has high frequency domain resolution, and there are significant differences between bispectrum of different faults (Xue, 2020). Therefore, in the study of fault diagnosis of offshore wind turbine gearbox, bispectrum can be used as the input image of convolutional neural network.

The specific definition of bispectrum is the two-dimensional Fourier transform of third-order autocorrelation. Bispectrum, as a higher-order spectrum, expresses the correlation between spectral values and two frequency components, which can represent the nonlinearity and non-Gaussianity of the signal and realize the description of the fault characteristics (Yang et al., 2016).

It is assumed that the third-order cumulant is convergent.

$$\sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} |c_{3x}(\tau_1,\tau_2)| < \infty \quad (1)$$

where $x(t)$ is the vibration signal; c_{3x} is the third-order statistics of $x(t)$.

The bispectrum is the discrete Fourier transform of the third-order cumulant, which is expressed as:

$$B_x(\omega_1,\omega_2) = \sum_{\tau_1=-\infty}^{\infty} \sum_{\tau_2=-\infty}^{\infty} c_{3x}(\tau_1,\tau_2) e^{-j(\omega_1\tau_1+\omega_2\tau_2)} \quad (2)$$

For a discrete-time energy-limited random signal $x(t)$, the bispectrum is:

$$B_x(\omega_1,\omega_2) = X(\omega_1)X(\omega_2)X^*(\omega_1,\omega_2) \quad (3)$$

where $X(\omega)$ is the Fourier transform of signal $x(t)$; $X^*(\omega_1)$ is its conjugate complex number (Cheng et al., 2016).

4. Knowledge Graph Model

4.1. *Subject construction*

Knowledge graph is mainly composed of entities and relationships. For fault diagnosis of offshore wind turbine gearboxes, the subject mainly includes: fault type, fault characteristics, fault causes, and fault measures.

4.2. *Entity Extraction Based on Bi-LSTM*

In the knowledge graph, entities correspond to specific nodes. Entity extraction is the first step in the construction of knowledge graph, and the most widely used algorithm is Bi-LSTM (Deng, 2022).

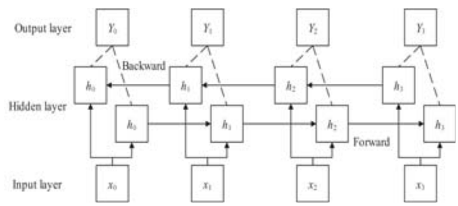


Fig.4. The structure of Bi-LSTM network

Bi-LSTM network connects two LSTM layers with opposite timing directions. In the process of entity extraction, the network will use the information above and below at the same time to achieve accurate entity extraction. The structure of the Bi-LSTM network is shown in Fig. 4.

4.3. *Relation Extraction Based on BERT Model*

In the knowledge graph, the relationship is a directed line segment connecting two entities. Relation extraction of unstructured data is usually based on entity extraction. Relation extraction is the judgment of the relationship between entities.

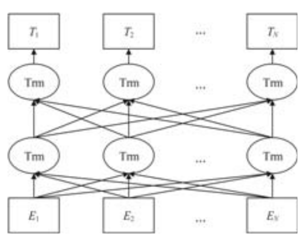


Fig. 5. The structure of BERT model

In the algorithm of relation extraction, the pre-training model based on BERT has a good extraction effect (Zhang et al., 2021). The model structure of BERT is shown in Fig. 5. *E* represents the input word vector, Trm represents the Transformer encoder, and *T* represents the output word vector.

4.4. *Knowledge Graph Construction and Intelligent Question Answering System*

After extracting entities and relationships from expert knowledge, the knowledge graph is constructed. The construction of knowledge graph usually relies on graph database. Among many graph databases, Neo4j is easy to operate and has good performance. Therefore, Neo4j is used to construct the knowledge graph for offshore wind turbine gearbox.

Intelligent question answering system is a specific application of knowledge graph, which can classify the questions raised by users and provide answers. The question classification is realized by the Aho-Corasick algorithm. The Aho-Corasick algorithm can perform multi-pattern matching on the questions. The matching information includes the fault type and some keywords (characteristics, causes and measures). After matching the questions, the constructed knowledge graph is used to retrieve and answer the questions.

5. The Application of the Knowledge and Data Fusion-driven Fault Diagnosis Method for Offshore Wind Turbine Gearbox

5.1. *Gearbox fault Diagnosis Based on CNN*

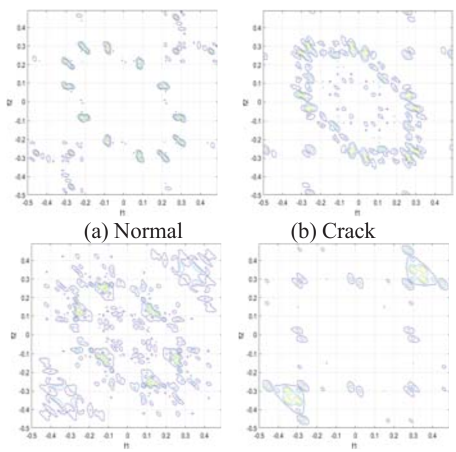
5.1.1. *Data Acquisition of Gearbox*

Due to the complex operating conditions of variable speed and variable pitch of the entire offshore wind turbine, the gearbox system operates under variable speed and variable load conditions. The simulation of the working condition of the gearbox system can be carried out by controlling the speed and loading. Furthermore, the structure of offshore wind turbine gearboxes is mostly planetary gear trains, so this article takes planetary gear trains as the research object. Therefore, an experimental platform for planetary gearbox of offshore wind turbine is built, and the structure is shown in Fig. 6. The experimental system is mainly composed of three-phase variable frequency motor, planetary gearbox, loader and torque speed sensor.

Fig. 6. Planetary gearbox experimental platform for offshore wind turbine

Table 1. Planetary wheel fault data set

Label	Fault type	Working condition (speed-load)	Number	Total
1	Normal	600rpm-5nm	120	240
		1500rpm-5nm	120	
2	Crack	600rpm-5nm	120	240
		1500rpm-5nm	120	
3	Wear	600rpm-5nm	120	240
		1500rpm-5nm	120	
4	Missing teeth	600rpm-5nm	120	240
		1500rpm-5nm	120	



(a) Normal (b) Crack

(c) Wear (d) Missing teeth

Fig. 7. Bispectrum of different faults

In the experimental simulation, the planetary gear is selected as the analysis object, the sampling frequency is 12×1024Hz, and set two kinds of rotational speeds (600rpm and 1500rpm). The vibration acceleration signals of normal, planetary gear crack, planetary gear missing teeth and planetary gear wear were collected respectively. The vibration signal of 20s is collected under each working condition, and it is divided into 120 data samples on average. The single data sample is 2048 points, and there is no overlap between each sample. Therefore, the number of samples in each state of the planetary gear is 240, and the description of the planetary gear fault data set is shown in Table 1. The one-dimensional vibration signal is further converted into bispectrum with higher frequency domain resolution, which is shown in Fig. 7.

5.1.2. *Construction of 2DCNN and Parameter Setting of Gearbox*

According to the needs of gearbox fault diagnosis, this paper designs the structure of two-dimensional convolutional neural network (2DCNN). The 2DCNN consists of three convolutional layers, three maximum pooling layers, one fully connected layer, one softmax layer and one classification output layer. The training parameter setting is shown in Table 2.

Table 2. Parameter settings

Parameter	Value
Batch	60
Learning rate	0.001
Epoch	20
Verification frequency	15

5.1.3 *Convolutional Neural Network Diagnosis Results of Gearbox*

70% of the sample data is selected for training and 30% for testing. The training and verification process of the model is shown in Fig. 8. When the number of iterations is 20, the classification accuracy has reached 90%. When the number of iterations is 30, the classification accuracy has reached 95%. The confusion matrix of the verification data is shown in Fig. 9, in which the horizontal axis represents the predicted fault type of the planetary gear and the vertical axis represents the real fault type of the planetary gear. It can be seen from the diagram that the method proposed in this paper has a good diagnostic rate for planetary gear faults in offshore variable condition simulation. The planetary gear normal and planetary gear cracks are more easily identified.

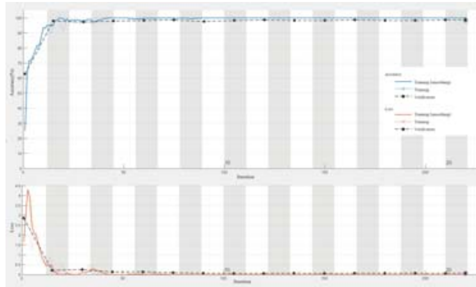


Fig. 8. Training and validation of CNN model

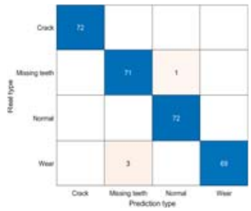


Fig. 9. Confusion matrix

5.2. *Gearbox Fault Intelligent Question Answering System Based on Knowledge Graph*

5.2.1. *Data Information of Gearbox*

The expert knowledge data used in the gearbox fault intelligent question answering system mainly come from the gearbox fault diagnosis manual and the offshore wind turbine operation and maintenance manual. The content mainly includes: fault type, fault characteristics, fault causes and fault measures. The text data is pre-processed to form a corpus of 67 sentences with multiple faults of gearbox.

5.2.2 *Entity Extraction Results*

According to the results of topic construction, the open source data labeling tool Label Studio is used to label the collected corpus according to the BMEO format. The entity labeling results are shown in Table 3, and the visualization is shown in Fig. 10.

Fig. 10. Visualization of entity labeling results (The Chinese sentence in Fig. 10 means the reason for the formation of broken teeth is that the tooth surface hardness is not enough.)

Table 3. Entity labeling results

Chinese character	Label	Chinese character	Label
断	B-Fault	齿	B-Characteristic
齿	E-Fault	面	M-Characteristic
的	O	的	M-Characteristic
形	O	硬	M-Characteristic
成	O	度	M-Characteristic
原	O	不	M-Characteristic
因	O	够	E-Characteristic
是	O	。	O

The labeled training corpus is exported and trained in the Bi-LSTM model. The model training parameters are shown in Table 4. The effect of training is evaluated by F1 score. The F1 score obtained after training is 0.897, indicating that the Bi-LSTM network has a good effect on entity extraction.

Table 4. Model training parameters

Parameter	Value
Embedding layer size	101
Hidden layer size	107
Batch	5
Learning rate	0.001
Epoch	200

5.2.3. *Relation Extraction Results*

The open source data labeling tool Label Studio is used to label the relationship between entities, and the results are shown in Fig. 11. Fault-characteristic, Fault-measure and Fault-reason are different relation labels. In the training process, in order to increase the robustness of the relation extraction system, an "Unknown" relation label is added for relation extraction.

Fig. 11 Visualization of relation labeling results (The Chinese sentence in Fig. 11 means the reason for the formation of broken teeth is that the tooth surface hardness is not enough.)

The labeled relationship data is put into the BERT pre-training model for training. By fine-tuning the training method, the specific parameters of the pre-training are adjusted to obtain a better relationship extraction effect. The training parameters are shown in Table 5. The F1 score after training is 0.884, indicating that the BERT pre-training model has a good effect on entity extraction.

Table 5. Training parameters

Parameter	Value
Batch	8
Learning rate	0.00001
Epoch	5

5.2.4. *Knowledge Graph Construction and Intelligent Question Answering System*

The construction of knowledge graph is mainly based on Neo4j software. The knowledge graph of fault diagnosis for offshore wind turbine gearbox is constructed by the results of labeling and model extraction. The node information of the knowledge graph broken tooth fault is shown in Fig. 12.

Fig. 12 Knowledge graph of broken tooth fault

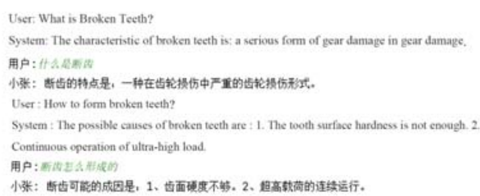


Fig. 13 Intelligent question answering of broken tooth fault

After completing the construction of knowledge graph, the intelligent question answering system for gearbox fault diagnosis is further generated. The Aho-Corasick algorithm is used to perform multi-pattern matching on the question, and the fault types and some keywords (characteristics, causes and measures) contained in the question are extracted. Furthermore, the corresponding answers are given by using the information stored in the knowledge graph. The intelligent question answering results for broken teeth fault are shown in Fig. 13.

6. Conclusion

This paper has proposed a knowledge and data fusion-driven fault diagnosis method for offshore wind turbine gearbox is proposed, which realizes the comprehensive utilization of vibration data and expert knowledge. The constructed CNN diagnostic network can identify the fault type under variable working conditions, and the fault diagnosis accuracy of the model can reach 95%. The knowledge graph constructed can realize the visual display of faults and question answering of fault-related knowledge for the diagnosis results of CNN network. The results show that the knowledge and data fusion-driven fault diagnosis method for offshore wind turbine gearbox has a good application effect on the intelligent operation and maintenance of offshore wind turbines.

Acknowledgement

This paper is supported by CNOOC Energy Development Limited Company-China University of Petroleum (Beijing) joint innovation fund (GD2021ZCAF0021), and China petroleum science and technology innovation fund (No. 2021DQ02-0801).

References

Cheng J., W. Wang, and S. He (2016). Application of the dual spectrum analysis method in fault diagnosis of wind turbine bearing. *Process Automation Instrumentation*, 27-29.

Cortes C., and V. Vapnik (1995). Support-vector networks. *Machine Learning*, 273-297.

Deng C. (2022). Research on entity relation extraction technology based on deep learning. Working Paper, Guangdong University of Technology.

Guo H. (2002). Design of fault diagnosis expert system for wind power generation system. *Energy Conservation*, 36-38.

He Z., J. Chen, T. Wang, and F. Chu (2010). Theories and Applications of Machinery Fault Diagnostics. *Higher Education Press*.

Jing L., M. Zhao, P. Li, and X. Xu (2017). A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement*, 1-10.

Lee W., D. Grosh, F. Tillman, and C. Lie (1985). Fault tree analysis, methods, and applications-a review. *IEEE transactions on reliability*, 194-203.

Li H., Y. Hu, C. Yang, Z. Chen, H. Ji, and B. Zhao (2013). An improved fuzzy synthetic condition assessment of a wind turbine generator system. *International Journal of Electrical Power & Energy Systems,* 468-476.

Liu J., L. Gao, Y. Sun, X. Feng, and H. Ji (2022). Fault diagnosis method for equipment driven by knowledge and data fusion. *Journal of Zhengzhou University (Natural Science Edition)*, 39-46.

Long X., P. Yang, H. Guo, and X. Wu (2017). Review of fault diagnosis methods for large wind turbines. *Power System Technology*, 3480-3491.

Nie T., J. Zeng, Y. Cheng, and L. Ma (2022). Knowledge graph construction technology and its application in aircraft power system fault diagnosis. *Acta Aeronautica et Astronautica Sinica*, 46-62.

Xue X. (2020). Fault diagnosis of planetary gearbox based on deep learning and transfer learning. Working Paper, Taiyuan University of Technology.

Yang S., B. Tan, and R. Zhou (2016). A jamming identification method against radar deception based on bispectrum analysis and fractal dimension. *Journal of Xi'an Jiaotong University*, 128-134.

Zhang C., X. Gu, R. Li, Y. Li, and W. Liu (2021). Construction method for financial personal relationship graphs using BERT. *Journal of Frontiers of Computer Science and Technology*, 137-143.