

REVIEW

Open Access



Signal-Based Intelligent Hydraulic Fault Diagnosis Methods: Review and Prospects

Juying Dai, Jian Tang*, Shuzhan Huang and Yangyang Wang

Abstract

Hydraulic systems have the characteristics of strong fault concealment, powerful nonlinear time-varying signals, and a complex vibration transmission mechanism; hence, diagnosis of these systems is a challenge. To provide accurate diagnosis results automatically, numerous studies have been carried out. Among them, signal-based methods are commonly used, which employ signal processing techniques based on the state signal used for extracting features, and further input the features into the classifier for fault recognition. However, their main deficiencies include the following: (1) The features are manually designed and thus may have a lack of objectivity. (2) For signal processing, feature extraction and pattern recognition are conducted using independent models, which cannot be jointly optimized globally. (3) The machine learning algorithms adopted by these methods have a shallow architecture, which limits their capacity to deeply mine the essential features of a fault. As a breakthrough in artificial intelligence, deep learning holds the potential to overcome such deficiencies. Based on deep learning, deep neural networks (DNNs) can automatically learn the complex nonlinear relations implied in a signal, can be globally optimized, and can obtain the high-level features of multi-dimensional data. In this paper, the main technology used in an intelligent fault diagnosis and the current research status of hydraulic system fault diagnosis are summarized and analyzed. The significant prospect of applying deep learning in the field of intelligent fault diagnosis is presented, and the main ideas, methods, and principles of several typical DNNs are described and summarized. The commonality between a fault diagnosis and other issues regarding typical pattern recognition are analyzed, and research ideas for applying DNNs for hydraulic fault diagnosis are proposed. Meanwhile, the research advantages and development trend of DNNs (both domestically and overseas) as applied to an intelligent fault diagnosis are reviewed. Furthermore, the fault characteristics of a complex hydraulic system are summarized and discussed, and the key problems and possible research ideas of applying DNNs to an intelligent hydraulic fault diagnosis are presented and comprehensively analyzed.

Keywords: Hydraulic system, Intelligent fault diagnosis, Deep neural networks

1 Introduction

Hydraulic systems are widely used in modern machinery owing to a multitude of advantages, such as a fast response, significant load stiffness, large power density, and superior stability. A hydraulic system is often the core component of engineering equipment, such as control and power transmission systems, which are typically operated in the field. A hydraulic system can be damaged by exposure to sunshine, rain forests, and dust particles, among other factors, and by unstable working conditions

such as a high load or severe impact. Therefore, such systems are prone to faults, and if certain initial abnormalities are not located and eliminated in time, they may develop into a functional disability and even lead to a dangerous condition. Therefore, it is extremely important to diagnose and remove such problems in time.

However, a proper hydraulic fault diagnosis remains a challenge. Compared with common mechanical and electrical structures, hydraulic system faults in engineering equipment are more hidden and unclear. It is therefore difficult to obtain fault information, and the mapping relationship between a fault characterization and fault cause is complex. It is thus extremely important

*Correspondence: lgdx_tj@163.com
Field Engineering College, Army Engineering University, Nanjing 210007, China

to research key technologies and methods for achieving hydraulic fault diagnosis.

Currently, fault diagnosis methods can be divided into three categories, mode-, knowledge-, and signal-based methods. Model-based methods must establish the model of the system before a diagnosis; hence, they require a clear understanding of the structure, principle, and mechanism of the diagnosed object [1]. They mainly include a state estimation, parameter estimation, and an equivalent space. However, for a complex hydraulic system, it is difficult to establish a model owing to coupling between the different variables. Knowledge-based methods are based on obtaining a large amount of expert knowledge to simulate an expert reasoning process of a certain model. They are suitable for situations in which the fault reasoning is clear and the decision logic is strong [2]. However, their poor incremental learning ability makes it difficult to add new knowledge to the existing system, and the learning of new samples will be at the cost of giving up original knowledge. Owing to the dynamic complexity of a hydraulic system and the lack of its fault mechanism, prior knowledge is insufficient, thus leading to fewer studies in this area.

Signal-based methods employed to carry out a diagnosis are based on state signals, such as vibrations, sound, temperature, and pressure. If the system has a fault, it should be reflected in the signal. Thus, the essential feature of a fault can be theoretically obtained, provided the features of the signal are appropriately mined and the pattern recognition method is properly carried out. Signal-based methods have been widely studied in intelligent fault diagnosis. First, it benefits from the development of sensor and storage technologies, which allow the monitoring system to collect and store large amounts of offline and online signal samples. The second reason lies in the constant innovation and high-level performance achieved by machine learning algorithms. Machine learning [3] is a type of computer program that can continuously obtain new knowledge and an optimized performance based on incremental learning. When used in a fault diagnosis, the knowledge can be automatically and self-adaptively induced and generalized from the samples obtained. It is therefore an ideal model for feature extraction and pattern recognition.

Currently, a popular research area of machine learning is focused on deep learning, and the most promising approach is the use of deep neural networks (DNNs), which have been widely researched and used in image recognition, speech recognition, and other pattern recognition fields. Their surprising and outstanding performance has aroused the attention of experts in the area of fault diagnosis. Consequently, in this paper, the research status of different technologies and the prospect

of applying signal-based hydraulic intelligent fault diagnosis methods based on a DNN are thoroughly analyzed and discussed. The remaining sections are as follows: In Section 2, the current research status on signal-based intelligent fault diagnosis methods are summarized and analyzed. In Section 3, several typical DNN models that can be potentially used in intelligent diagnosis are listed, and their structure, characteristics, and application are analyzed. In addition, the broad prospect of applying a DNN to a hydraulic fault diagnosis is presented. The research status of typical DNNs used in the field of intelligent fault diagnosis is analyzed in Section 4. In Section 5, the characteristics of complex hydraulic system faults are analyzed, and the key problems and possible research areas when applying DNNs to an intelligent hydraulic fault diagnosis are proposed and thoroughly analyzed. Finally, some concluding remarks are provided in Section 6.

2 Overview of Fault Diagnosis Technology

The fault diagnosis of a hydraulic system is included in a mechanical fault diagnosis. A hydraulic system fault diagnosis needs to combine the structural characteristics of the hydraulic system, the signal manifestation, and other factors. The diagnosis process has not changed. According to the content of the mechanical fault diagnosis, a signal-based intelligent diagnosis is usually a compound method with several processes. Each process and the main methods are shown in Figure 1. As indicated in Figure 1, the common technology during each process of a signal-based intelligent fault diagnosis will be thoroughly discussed.

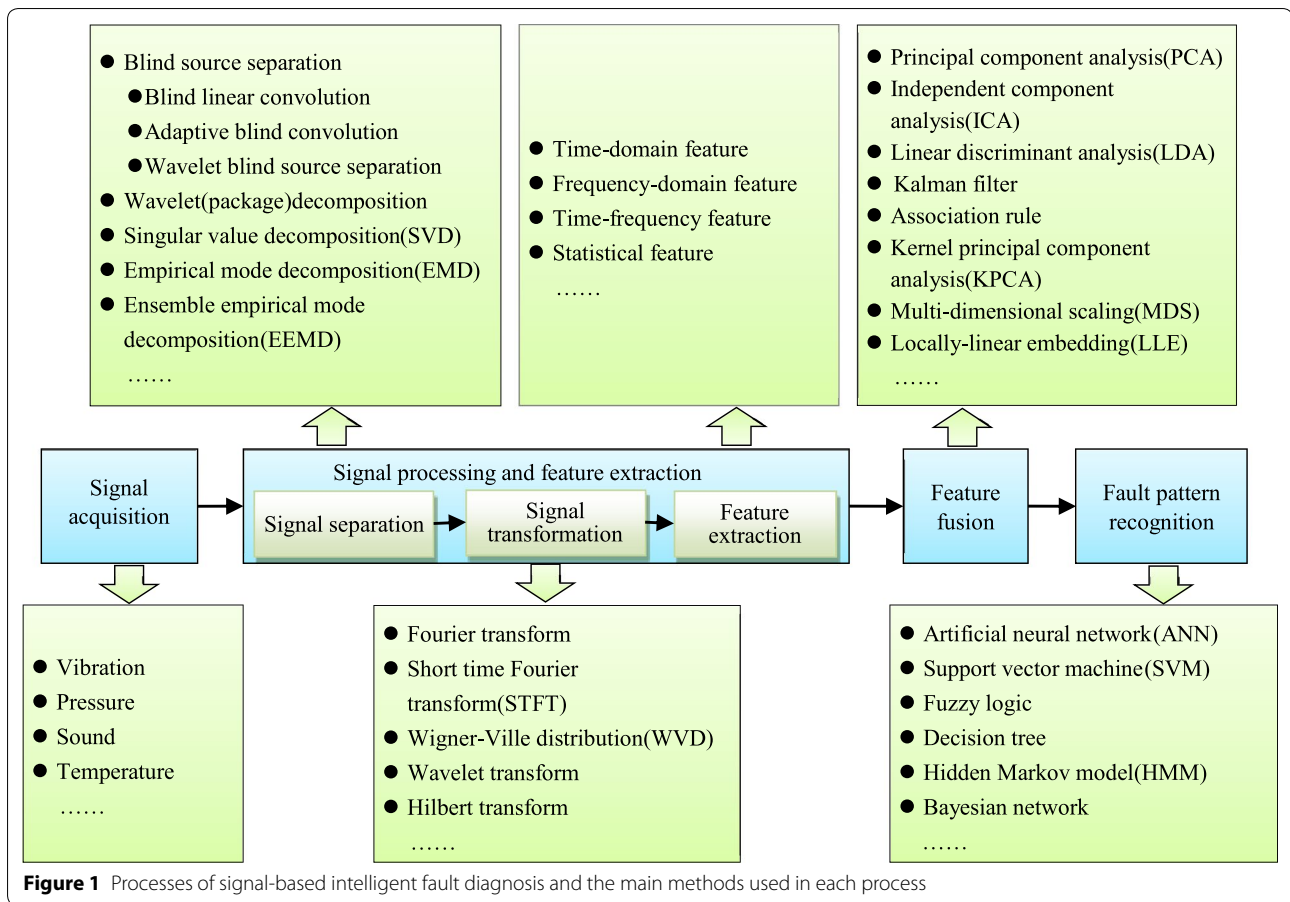
2.1 Main Flow of Fault Diagnosis

2.1.1 Signal Processing and Feature Extraction

Feature extraction from a signal is intensely crucial in a fault diagnosis because it determines the result of the subsequent fault recognition. Different types of signals such as audio, vibration, temperature, and oil debris signals [4] can be used to reflect the state of a machine. Among them, a vibration signal is widely used because it has been deeply researched in theory and is easy to collect. Hence, vibration signals are taken as an example in the following analysis.

(1) Signal separation

The original signal typically needs to be decomposed into several components to de-noise or separate different faults. Typical decomposition methods are shown in Figure 1. These are mainly blind source separation methods, such as wavelet (package) decomposition, singular value decomposition (SVD), empirical mode decomposition (EMD), and ensemble empirical mode decomposition (EEMD).



Zhu et al. [5] proposed a shift-invariant sparse coding method of blind source separation, which was used to extract the pressure pulsation for hydraulic pumps and achieved excellent results. Focusing on the problem of EMD sensitivity to noise, Van et al. [6] put forward a hybrid of a non-local-mean decomposition method. However, the experiment results showed that the above methods are not ideal for a hydraulic fault. The first reason is that the features of a hydraulic composite fault are not necessarily the sum of those of a single fault. The second reason is that the same fault with a different degree of damage does not necessarily show the same frequency components, and occasionally even excites new fault features. Regarding these issues, Figure 2 demonstrates a good case, which indicates the Hilbert marginal spectrum based on the improved EEMD of different fault signals.

(2) Signal processing and feature extraction

The purpose of signal processing is to realize a signal transformation and feature extraction. The extracted features, such as the time-domain, frequency-domain, time-frequency, and statistical features will be used for the fault pattern recognition.

As for a hydraulic vibration signal, owing to its non-stationary and strong time-varying characteristics, a Fourier transform cannot describe how the frequency of the signal changes over time. Goharrizi used a Fourier transform for a leakage detection of a hydraulic cylinder, although the results were unsatisfactory. The author then applied a wavelet [7] and an EMD [8] to decompose the hydraulic vibration signals. In addition, a short-time Fourier transform (STFT) was proposed by Gabor in 1946. Based on the original Fourier transform, a moving window is used during the transformation, which can realize the localization of the local frequency characteristics. Restricted by the Heisenberg uncertainty principle, it cannot achieve a high resolution in the time- and frequency-domains at the same time.

The Wigner-Ville distribution (WVD) [9] has good mathematical operation properties. When used to analyze a non-stationary signal, a good performance is demonstrated most of the time. However, the WVD has an insuperable deflection—cross interference will be generated among the different frequency components. When used to detect some hydraulic faults, such as a cavitation

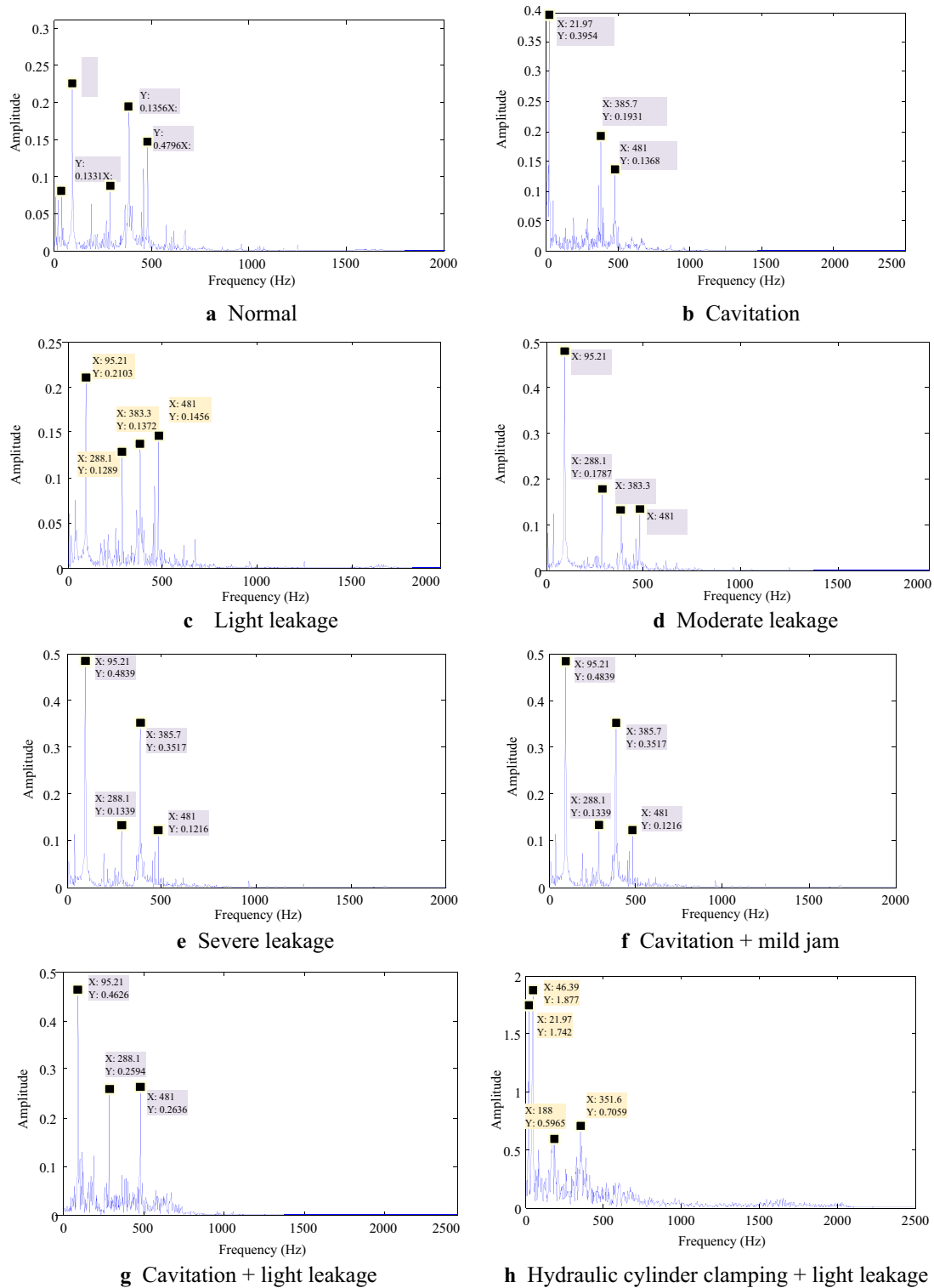


Figure 2 Marginal Hilbert spectrum of various types of hydraulic fault signals

and composite fault, it cannot successfully extract the frequency characteristics.

The wavelet transform inherits the localization idea of STFT. At the same time, it provides a time-frequency window that can automatically and adaptively change with the frequency. Through an expansion and translation operation, WT can realize a high frequency resolution at low frequency and a high time resolution at a high frequency, making it easy to focus on the details of the signal [10]. Thus, it is an ideal tool for a time-frequency analysis of a signal. However, the selection of a wavelet base has significant influence on the results of a time-frequency transformation [11]. Different wavelets are suitable for analyzing different time-frequency characteristics. For inexperienced researchers, how to choose the wavelet base is a challenge.

Consequently, signal processing methods with a multi-scale expression and strong adaptive ability, such as the self-adaptive wavelet, EMD, and EEMD, have become areas of focus. In particular, the latter two are entirely self-adaptive, and thus are more suitable for the non-stability signals of a hydraulic fault. To solve the problems of an end-effect, modal aliasing, and an overshoot-undershoot of the EMD and EEMD, numerous improved algorithms have been suggested. For example, Amirat et al. [12], Zheng et al. [13], and Chai et al. [14] proposed methods based on a signal extension to eliminate the end-effect. In addition, Huang et al. [15] tried to improve the fitting algorithm to reduce the overshoot-undershoot phenomenon of the upper and lower envelopes, and Wang et al. [16] and Chen et al. [17] relieved the modal aliasing by optimizing the decomposing termination conditions. The improved EMD and EEMD can describe the transient characteristics of the signal more clearly, and thus they are more suitable for a hydraulic vibration signal.

(3) Feature fusion

To improve the diagnosis credibility of complex equipment working in an extremely noisy environment, researchers have attempted to combine several feature extraction methods and fuse their extraction results as a basis for diagnosis. For example, Li and Wang [18] put forward an algorithm of feature linear fusion based on affinity propagation (AP) clustering. Its validity was verified on a test platform of an aero-engine rotor. In addition, Li et al. [19] combined an SVD with morphological filtering and applied it to fault feature extraction.

When different features are fused, there will be a high-dimensional vector, thereby causing the consumption of large amounts of computing resources. At the same time, not all features have the same contribution to, and correlation, with a fault diagnosis. Therefore, a dimensionality reduction and sparse representation are required. As

shown in Figure 1, the main methods studied include the principal component analysis (PCA) [20], independent component analysis (ICA), and linear discriminant analysis (LDA) [21]. The above methods are all based on the hypothesis that the samples obey a Gaussian distribution (PCA) or a linear model (ICA or LDA). Nevertheless, the actual vibration signals usually do not satisfy these conditions. Thus, some kernel methods, such as a kernel principal component analysis (KPCA) and kernel linear discriminant analysis (KLDA), have been presented to solve this problem. They use a kernel function to convert data from one space to another to relieve the difficulty of extracting nonlinear features.

In addition to kernel methods, many learning methods have also been widely studied, such as the gray relational analysis (GRA) [22], Kalman filtering [23], and non-negative matrix factor [24]. However, GRA needs to determine the optimal values of the indicators first, and thus is subjective. In addition, the number of calculations of the Kalman filter will increase three times with an increase in the information dimension, and a non-negative matrix decomposition focuses on the extraction of important local features and cannot fully reflect global properties of the signal.

2.1.2 Fault Pattern Recognition

Machine learning is a promising and widely researched classification algorithm owing to its powerful adaptive and incremental learning ability. Among them, an artificial neural network (ANN) and a support vector machine (SVM) are the two most typical classifiers.

In the field of fault diagnosis, the most researched ANNs include a back propagation (BP) network, a radial basis function (RBF) network, an auto-associative neural network (AANN), a Hopfield network, a self-organization mapping (SOM) network, and a Boltzmann machine. To overcome the proneness of following into the local optimum, scholars have put forward many new methods such as an echo state network (ESN) [25], a probabilistic neural network (PNN) [26], and a fuzzy neural network (FNN) [27]. Regarding complex classification problems, there are many bottlenecks that restrict improvement of the ANN performance. Among the main reasons for this, one is an over-fitting owing to a highly strict training goal. If no prior knowledge (experience) is available, it is difficult to determine what training goal is the most suitable. The second reason is that the construction and training of the network are dependent on the experience of the researchers. Thus, prior to 2006, studies related to an ANN remained stagnant, and therefore its performance showed little significant improvement.

An SVM uses an inner product kernel function instead of a nonlinear mapping to a high-dimensional space.

Thus, compared with an ANN, it shows numerous unique advantages in solving nonlinear, high-dimensional pattern recognition problems, particularly when the number of samples is small. Therefore, it has been widely studied in the field of fault diagnosis [28, 29]. Researchers have proposed many improved versions of an SVM for specific diagnostic issues, such as a Gauss kernel SVM [30], particle swarm optimization SVM (PSO-SVM) [31], and EEMD-SVM [32]. However, the application of an SVM to an intelligent fault diagnosis of a complex hydraulic system is limited. One reason for this is that, when faced with large-scale training samples, training of an SVM is difficult to carry out. However, in the real world, an intelligent diagnosis is always based on on-line monitoring, and there are inevitably large numbers of samples applied. As the second reason, an SVM is not suitable for a multi-classification problem. With an increase in system complexity, if also considering the fault superposition and degree of damage, it is inevitable that the number of fault modes will significantly increase.

In addition to an ANN and an SVM, other machine learning algorithms, including logistic regression [33], decision tree [34], Hidden Markov model [35], and Bayesian network [36], have been applied in fault pattern recognition and have obtained some interesting achievements.

2.2 Research Status of Fault Diagnosis for Hydraulic Systems

Fault diagnosis in a hydraulic system is a specific application of the fault diagnosis technology used in a hydraulic system. With the structure of hydraulic equipment becoming more complex, its functionality is also becoming more powerful and its level of automation is increasing, thereby requiring greater reliability of the hydraulic system. The fault diagnosis technology of a hydraulic system is being continuously developed and innovated, and a comprehensive discipline integrating hydraulic control, sensors, decision theory, statistical mathematics, signal processing, artificial intelligence, and pattern recognition has gradually been formed. In retrospect, the following three stages of development have generally occurred.

2.2.1 Subjective Diagnosis

A subjective diagnosis method mainly relies on technical personnel to obtain the condition information of hydraulic equipment through a direct observation or simple diagnostic instrument. Many practical diagnosis methods have been summarized, including sensory diagnosis, instrument detection, parameter measurement, a replacement component method, disassembly component method, logic chain analysis, graph theory, fault tree analysis, and section division method. For example, Lu

[37] elaborated on a method for eliminating on-site faults of a hydraulic system based on a subjective diagnosis. In addition, Ji et al. [38] used graph theory and a block diagram to analyze the fault mechanism and diagnosis method of hydraulic components. Although a subjective diagnosis method achieves simplicity, rapidity, and practicability, it relies too much on the practical experience and personal professional ability of the maintenance personnel. It has a limited diagnostic ability for complex hydraulic systems, and the process of disassembling and assembling the hydraulic equipment damage the lifetime of the equipment.

2.2.2 Fault Diagnosis of Hydraulic System Based on Signal Processing Technology and Mathematical Model

With the development of signal processing technology and faster computer computations, a variety of mature mathematical models and signal processing methods have been introduced into the field of hydraulic diagnosis. A hydraulic fault diagnosis method based on signal processing technology extracts the corresponding features by measuring the hydraulic state parameters, and describes the corresponding relationship between the measured signals and faults through a mathematical model to achieve a diagnosis. This mainly includes a time domain analysis, frequency domain analysis, time-frequency analysis, state estimation, and parameter estimation. For example, Sepehri et al. [39, 40] used an improved Kalman filter to estimate the state of the hydraulic actuator and hydraulic system leakage. In addition, Zhu and Gao [41] discussed the relationship between the change in flow of a hydraulic system and various faults. Du and Zhang [42] and Chen et al. [43] analyzed the vibrations and noises occurring in hydraulic system components within the time or frequency domain, and judged the fault type, degree, and location through a comparison with the time-frequency characteristics of normal signals. Moreover, Jiang et al. [44], Du et al. [45], and Goharrizi and Sepehri [7] used a wavelet transform to analyze the vibration signal of a hydraulic pump, the pressure signal of a hydraulic cylinder, and a hydraulic actuator, respectively.

The type of hydraulic fault diagnosis applied compensates the inefficiency of applying manual data statistics by applying an objective parameter measurement and through the advantage of computer signal processing. With the help of the identification capability of a mathematical model, such a diagnosis has achieved good application results in engineering fields. However, a hydraulic system is a non-linear time-varying system that has certain shortcomings, such as a difficulty in feature extraction, establishing a complex mathematical model, and other factors.

2.2.3 Fault Diagnosis of Hydraulic System Based on Artificial Intelligence

A technology for the diagnosis of a hydraulic system fault, which uses an artificial intelligence (AI) pattern recognition method as its main body and combines various feature extraction methods, has dominated the focus and development trend in this field. Abbott [46] designed a fault diagnosis approach of expert for a hydraulic system used on the NASA space shuttle; Crowther et al. [47] established a neural network identification model for the hydraulic system of a second-order hydraulic actuator. Amin et al. [48] combined multi-feature fusion and fuzzy decision theory to study the on-line health monitoring of hydraulic pumps. Chen et al. [49] acquired the vibration characteristics of hydraulic motors through a second generation wavelet transform and used an SVM to diagnose a fault in a hydraulic system. In addition, Saeed et al. [50] combined a PCA, an ANN, and a multi-adaptive neuro-fuzzy inference system to diagnose hydraulic pipeline faults.

Domestically, Jiang et al. [51] studied a fusion diagnosis strategy for a wavelet analysis, neural network, and fuzzy logic used in a plunger pump. Mou [52] discussed the application of an expert system in the fault diagnosis of a hydraulic system used in coking coal machinery. Lu [53] analyzed and diagnosed the vibration signals of a hydraulic pump by combining an EMD method with a fuzzy C-means clustering analysis. In addition, Tang [54] studied the hydraulic system of a concrete pump truck using an EMD, Hilbert envelope spectrum, and PSO-SVM. Chen et al. [17] studied the application of an EEMD and an SVM in a hydraulic fault diagnosis. Chai et al. [14] used a PCA to deal with a variety of time-frequency characteristics, and then applied a KELM model to diagnose a fault in a hydraulic system.

2.3 Analysis and Comments

At present, with the continuing progress regarding hydraulic control theory, excavator technologies, sensor and test technologies, signal processing methods, artificial intelligence, and pattern recognition methods, fault detection and recognition technologies used in hydraulic systems have made significant progress. Based on a theoretical analysis and a summary of the practical applications in this area, the following issues should not be overlooked.

2.3.1 How to Minimize or Even Avoid the Influence of Human Subjectivity on the Selection of Features

Feature extraction is unavoidably dependent on the experience and knowledge of researchers because signal processing methods and their characteristic parameters are selected artificially, and rely on a large amount of

manpower to extract discriminative features and analyze such features for an accurate fault recognition and classification, which is time-consuming and requires abundant expertise in terms of signal processing and analysis and fault diagnosis. Thus, finding a method that can automatically and adaptively acquire the signal features will clearly be a benefit to fault pattern recognition.

2.3.2 How to Realize Joint Optimization of the Feature Extraction and Pattern Recognition

Feature extraction and pattern recognition are currently conducted separately using independent models, such as a wavelet and decision tree [55], an EMD and ANN [56], an EMD and SVM [57], and an EEMD and a GA-SVM [17]. With these methods, the output of the previous model is the input of the latter. Because each process is independent, a cognitive deviation of different models cannot be globally corrected.

2.3.3 How to Achieve High-Level Transformation and Deep Fusion of State Signal Feature

Although shallow machine learning methods can automatically recognize faults based on the extracted features, which significantly reduce the influence of manual experience and subjectivity in the recognition process and results, the shallow architecture limits the capability of automatically learning high-level features from the input. Thus, traditional machine learning methods depend highly on artificially extracted discriminative features as the input. However, it is difficult to determine the most suitable features to be extracted, and different features may directly lead to different diagnosis results, which is time consuming and unstable.

Because of the structure and closed operation characteristics of the hydraulic system, a fault of a hydraulic system has certain characteristics including a concealment, a large influence of random factors, a complex mapping relationship between the signal characteristics and the system state, and thus it is extremely important to select an appropriate diagnosis method. As the latest achievement and research focus into AI, deep learning has been successfully applied to image recognition, speech recognition, and other complex pattern recognition fields. Owing to its powerful feature extraction, transformation, and fusion and pattern recognition capabilities, deep learning has received increasing attention in the field of fault diagnosis.

3 Deep Neural Network and Its Typical Models

3.1 Deep Neural Network

Hinton first proposed an unsupervised learning algorithm for a deep belief network (DBN) [58], which solved the difficulty of applying a training deep model and

created an upsurge in the amount of research conducted in this field. In 2013, deep learning was considered a top-ten breakthrough by *MIT Technology Review*. At the beginning of 2016, the Baidu deep speech recognition system was also ranked as an annual MIT ten big breakthrough. Each of the above fully reflects the research and application prospects of deep learning.

Deep learning is a type of machine learning [59]. Through multi-layer nonlinear network training, deep learning can realize a complex function approximation, represent a distributed expression of the input, and learn the essential features of the samples. The essence of deep learning is to learn features that are more useful by training a model with many hidden layers (deep structures) and a massive amount of training samples to eventually improve the ability of classification or prognosis.

In general, any mapping structure has the potential to build a deep model, including a local binary pattern (LBP), a PCA, or an LDA. Thus, any nonlinear structure with more than three layers can be considered a deep model. Therefore, the connotation of deep learning is extremely wide, including numerous models and structures. Among all deep models available, the deep neural network (DNN) has been the most rapidly developed and broadly used. The advantages of the DNN are not only reflected in the increase in the number of hidden layers, but also in the optimization of the network structure and an improvement in the training method applied. Although there are many different variants of DNNs, they have essentially evolved from the basic structures of the parent class. Among these, the most researched and widely used include a stacked autoencoder (SAE), a deep belief network (DBN), a convolutional neural network (CNN), and a recurrent neural network (RNN).

3.2 Introduction to Typical DNNs

3.2.1 Stacked Autoencoder

An SAE [60] is used to process high-dimensional complex data, such as a dimensionality reduction or feature learning. The structure of an SAE is stacked using several autoencoders (AEs). An AE is a typical three-layer neural network consisting of an input layer, a hidden layer, and an output layer. As shown in Figure 3, the network structure is divided into coding and decoding processes. Mapping from the input layer to the hidden layer is a coding process, and mapping from the hidden layer to the output layer is a decoding process.

Assume that the input data is X , the reconstructed data is \hat{X} , and the average reconstruction error between X and \hat{X} is as shown in Eq. (1), where n is the number of training samples, \hat{X}^i is a reconstructed sample, X^i is the input, W is the weight, and b is bias. AE extracts and

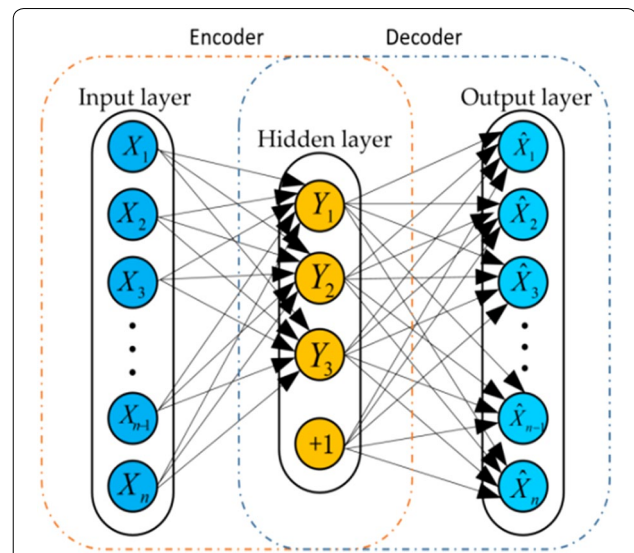


Figure 3 Single autoencoder (AE) architecture

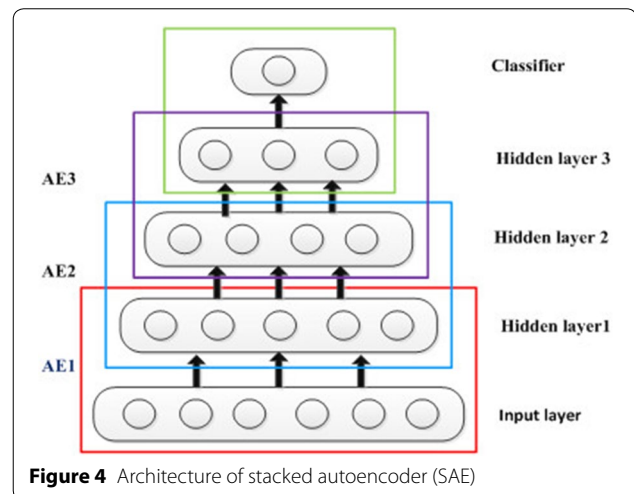


Figure 4 Architecture of stacked autoencoder (SAE)

compresses the input features in the coding part and then restores the compressed features in the decoding part. The aim is to minimize errors between the reconstructed data and the original data to obtain the most essential characteristics of the input. The architecture of an SAE is shown in Figure 4; for the training, an unsupervised self-learning method is applied such that each AE is trained sequentially so that the entire network is trained. The hidden layer parameters of each AE are then stacked to construct the SAE.

$$J(W, b) = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{2} \|\hat{X}^i - X^i\|^2 \right). \tag{1}$$

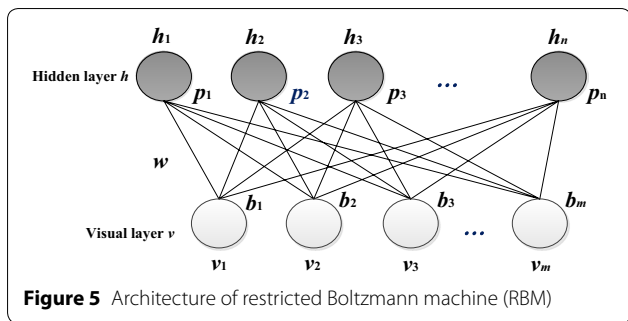


Figure 5 Architecture of restricted Boltzmann machine (RBM)

In addition to the structure of the SAE mentioned above, some deformation models, such as a sparse autoencoder [61] and a denoising autoencoder [62], have been developed. These different structural forms make the extracted features have different sparsity, robustness, and other characteristics. An SAE has been successfully applied to dimension reduction and information retrieval tasks. An SAE has a strong feature expression capability, allowing it to abstract the input in depth layer by layer and thus continuously obtain the essence of the input and a high-level feature expression, often achieving better classification results.

3.2.2 Deep Belief Network

A DBN is a deep neural network with multiple hidden layers, which is stacked using multiple restricted Boltzmann machines (RBMs). An RBM [63] is a typical neural network divided into a visible layer v and a hidden layer h . The nodes of the visible and hidden layers are connected by weight w , and b_m and p_n are the as biases of the corresponding units. The nodes of two layers are fully connected, whereas the nodes of the same layer are not connected. The structure of an RBM is shown in Figure 5. As Figure 6 indicates, a DBN consists of unsupervised RBMs and a supervised softmax classifier. The core idea is to train each RBM through unsupervised learning, and only train a single RBM at a time, using its training result as the input of a later RBM. The unsupervised pre-training method and greedy layer-by-layer learning is helpful in avoiding a problem in which the network falls into the local optimum, and the most essential characteristics of the input information can be obtained.

A DBN has strong autonomous learning and reasoning capabilities. It emphasizes the hidden representation of the learning data and highlights the characteristic expression of the data. The DBN solves the problems inherent to the training of multi-layer neural networks with a traditional BP algorithm as follows: a) a large number of training samples with labels, b) a slow convergence speed, and c) a falling into the local optimum owing to an inappropriate

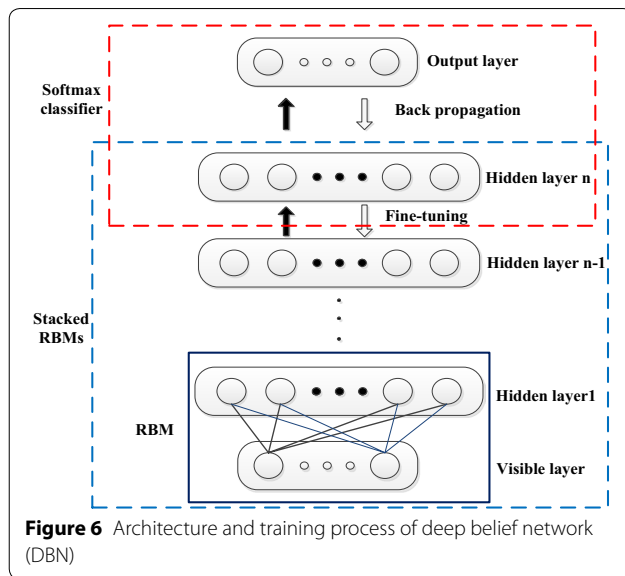


Figure 6 Architecture and training process of deep belief network (DBN)

parameter selection. This is helpful for solving issues that are difficult to deal with through shallow learning algorithms such as high-dimensional, complex, and non-linear expressions of large-capacity data.

3.2.3 Convolutional Neural Network

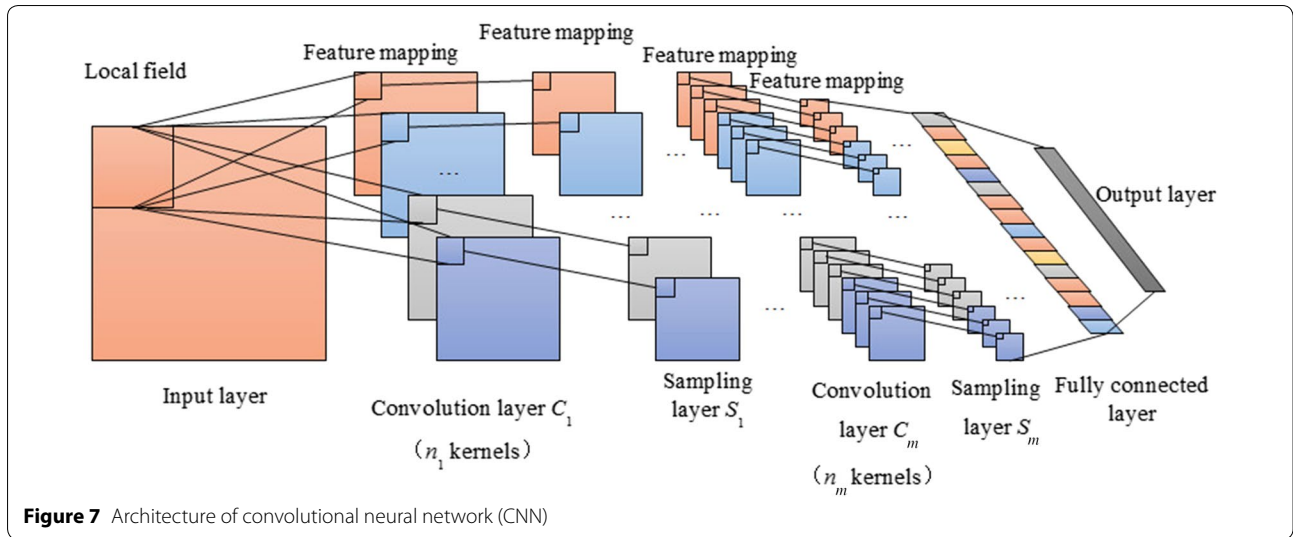
A CNN [64] is a multi-layer supervised learning network. As shown in Figure 7. Its network structure is composed of alternating convolution layers and sampling layers, followed by a full-connection layer and finally a classification layer. Each convolution kernel slides on the data and performs a convolution operation in the local field concurrently, which can explore the characteristics of the original input. In general, the formula for calculating the convolution layer is as follows:

$$x_j^l = f \left(\sum_{i \in M_j} x_i^{l-1} \times k_{ij}^l + b_j^l \right). \tag{2}$$

Among them, l represents the number of layers, k represents the convolution core, M_j represents the input receptive field, and b indicates the bias.

The main function of a pooling operation is to reduce the resolution and dimensions of the feature map, and to some extent increase the robustness of the network to displacement, scaling, and distortion. Pooling can be divided into maximum and average pooling. The form of the lower sampling layer is shown in Eq. (3):

$$x_j^l = f(\beta_j^l \text{down}(x_j^{l-1}) + b_j^l). \tag{3}$$

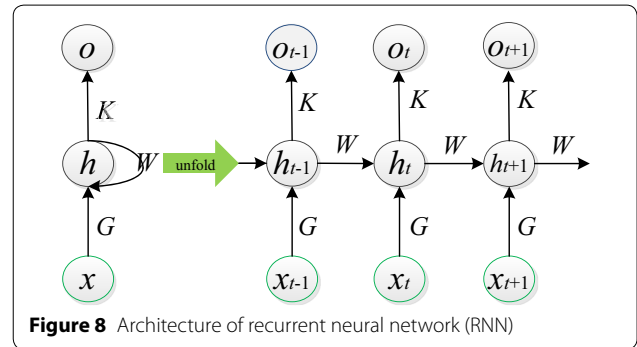


where $down(\cdot)$ is a pooling function and β is a weight coefficient.

As a characteristic of the CNN, the observation features obtained using a local receptive field method are independent of the translation, scaling, and rotation. During the convolution stage, the weight sharing structure reduces the network complexity, which is more obvious when the input features have a high resolution; during the down-sampling stage, the feature is sub-sampled based on the principle of a local correlation, which can effectively reduce the amount of data processing while retaining useful structural information. In particular, the multi-dimensional vector data can be directly input into the network, which avoids the complexity of the data reconstruction during feature extraction.

3.2.4 Recurrent Neural Network

An RNN [65] is one of the most advanced sequential data algorithms available. Because of its internal memory, it is suitable for machine learning issues involving sequential data. According to a certain time series, the structure with rings can be expanded into a sequence network, as shown in Figure 8, where x_t is the input of time t of the network, h_t represents the memory at time t , o_t represents the output of time t , and the direct weight from the input layer to the hidden layer is expressed by G . The original input is abstracted as the input of the hidden layer. In addition, the weight W from one hidden layer to another hidden layer is the memory controller of the network, which is responsible for scheduling the memory, and the weight from the hidden layer to the output layer is K , and the representation learned from the hidden layer will be abstracted through it again as the final output.



The specific calculation process of the forward propagation of an RNN is as follows, where f is a non-linear transformation function, such as \tanh , and g can be a softmax or other function.

$$h_t = f(Gx_t + Wh_{t-1} + b). \tag{4}$$

$$o_t = g(Kh_t + c). \tag{5}$$

The current state h_t of an RNN is determined by the state h_{t-1} of the previous moment and the current x_t . Based on its structural characteristics, it can be concluded that an RNN can best deal with an issue related to the time series. For the sequence data, the data at different times in the sequence can be input into the network in turn, and the output can be the prediction of the next time in the sequence, or the processing result of the information at the current time. RNNs solve the training problem of sequential data beautifully; hence, they have been widely used in natural language processing fields, such as speech and handwriting recognition and machine translation.

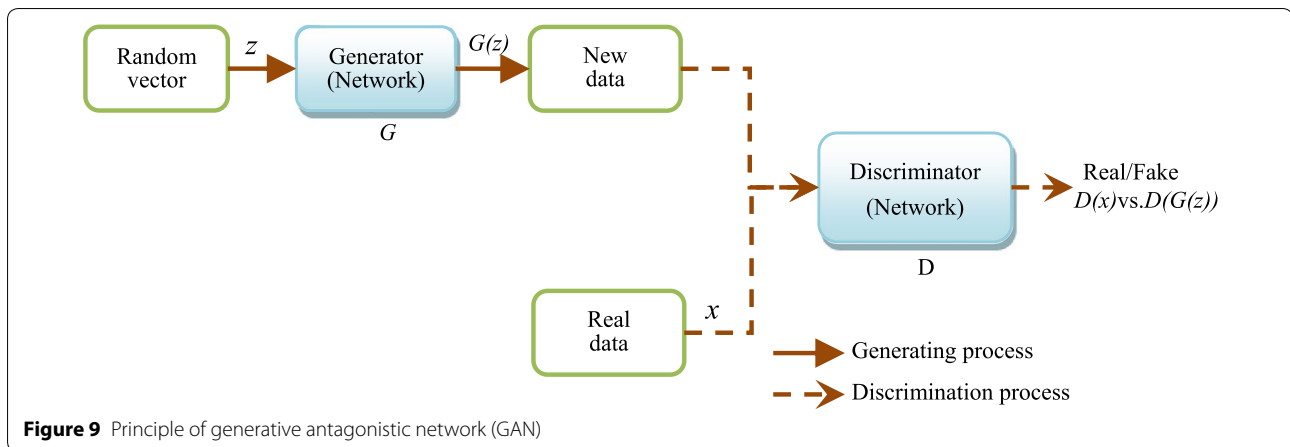


Figure 9 Principle of generative adversarial network (GAN)

3.2.5 Generative Adversarial Network

A generative adversarial network (GAN) [66] is a deep learning model and has become one of the most promising methods for applying unsupervised learning in a complex distribution in recent years. The model produces a good output through the game learning of two modules within the framework: a generative model and a discriminative model. A generative model (G) captures the distribution of the sample data, and a discriminant model (D) is a binary classifier, estimating the probability that a sample comes from the training data rather than the generated data. In addition, G and D are generally non-linear mapping functions, such as a multi-layer perceptron or a CNN. In practice, DNNs are generally used as a G or D. The main function of a GAN is image generation and data enhancement.

Specifically, as shown in Figure 9, a GAN mainly consists of two independent neural networks: a generator and a discriminator. The task of the generator is to sample a vector z from a random uniform distribution and then output the synthetic data $G(z)$; the discriminator will take the real data x and synthetic data $G(z)$ as input and output the probability that the sample is “true.”

The objective function of a GAN is shown in Eq. (6), where $D(x)$ denotes the probability that the discriminator will consider x to be the real sample, whereas $D(G(z))$ denotes the probability that the discriminator will consider the synthetic sample to be false. The form of Eq. (6) can be obtained by adding the logarithms:

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log(D(x))] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))]. \quad (6)$$

A GAN has been one of the few successful technologies in unsupervised machine learning, and this technology is rapidly innovating the ability to conduct generative tasks. GANs have been widely used in industry, from

interactive image editing and three-dimensional shape estimation to robot learning. They have also been applied in language tasks to improve the stability and increase the convenience of the training process.

3.2.6 Summary of Typical DNNs

Table 1 provides a brief summary of the typical DNN models that are commonly used in pattern recognition.

The successful application of a DNN to complex pattern recognition stems mainly from the advantages of its model structure and training method.

- (1) A deep structure with multiple layers is the core aspect of a DNN, which allows it to implement an advanced nonlinear transformation in a layer-by-layer manner. DNNs therefore have an excellent feature extraction capability.
- (2) In a DNN, the feature converter and mode classifier are integrated into a single model. Feature learning is oriented toward pattern recognition, and thus the feature transformation and classification in a DNN are jointly optimized.
- (3) Layer-by-layer greedy unsupervised learning is a self-adaptive learning process that greatly reduces the influence of human subjectivity on the parameter initialization and training result.

3.3 Similarity Analysis of Fault Diagnosis and Other Pattern Recognition Problems

DNNs have been used in 1D (e.g., text, time-series signal), 2D (e.g., images, time-frequency representation), and even 3D (e.g., video, 3D images) pattern recognition. Similarly, as shown in Figure 10, when a DNN is used in hydraulic fault diagnosis, its input may be 1D data, such as a time-domain signal (Figure 10(a)) or frequency spectrum (Figure 10(b)); 2D data, such as a time-frequency

Table 1 Summary and induction of typical DNNs

| Models | Structures | Characters | Applications |
|--------|---|---|--|
| SAE | Stacked AE layers + a classifier | 1) Unsupervised learning 2) Layer-by-layer unsupervised learning, followed by reverse supervised fine-tuning 3) Compression and distributed representation of dataset | 1-dimensional (1D): Data de-noising [67] 2-dimensional (2D): Target recognition [68] 3-dimensional (3D): High spectral image classification [69] |
| DBN | Stacked RBM layers + a classifier | 1) Unsupervised learning 2) Layer-by-layer unsupervised learning, followed by reverse supervised fine-tuning 3) Learning probability distribution on generative structure | 1D: Speech recognition [70] 2D: Radar image recognition [71] 3D: High spectral image classification [72] |
| CNN | Alternately appeared convolution and sampling layers + a classifier | 1) Supervised learning 2) Automatic learning of convolution kernels 3) Obtains the essence of the samples through a local convolution operation | 1D: Speech recognition [73] 2D: Image classification [74] Face recognition [75] 3D: Video classification [76] |
| RNN | Input layer + closed-loop hidden layers + output layer | 1) Unsupervised learning 2) Feedback loop taking the output of the previous moment as input 3) Mainly used for modeling time-series signals | 1D: Text classification [77] Machine translation [78] 2D: Image annotation [79] Emotional test [80] 3D: Video analysis [81] |
| GAN | Generator + discriminator | 1) Semi-supervised learning 2) Learn the probability distribution of the training samples and establish correlation between the input and output | 1D: Image generation [82] 2D: Image retrieval [83] 3D: Video prediction [84] |

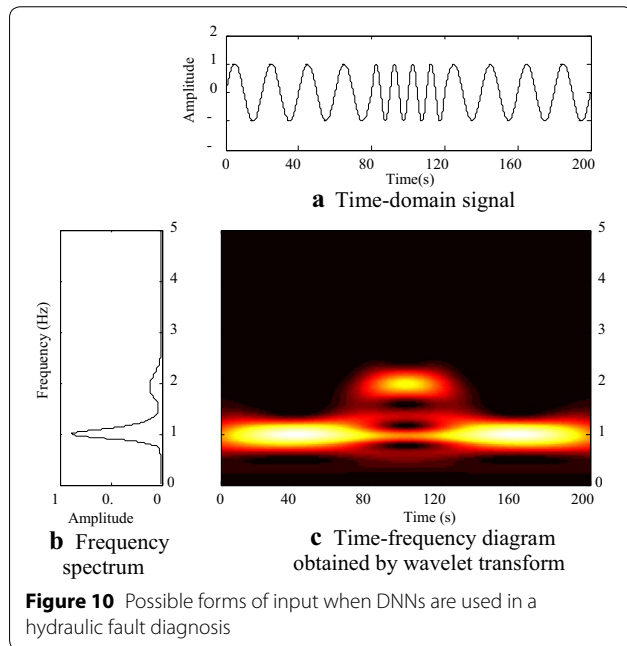


diagram (obtained using a time-frequency joint analysis such as a wavelet transform (Figure 10(c)); and 3D data, including a time-frequency diagram of multiple sensor signals. As shown in Figure 11, the aforementioned signals can be used as the input of 1D DNNs, 2D DNNs, and even 3D DNNs. Figure 12 shows two realization routines (research ideas) based on 1D and 2D DNNs.

4 Research Status of DNNs in Intelligent Fault Diagnosis Field

In 2015, Jia et al. [85] highlighted the advantages and application future of DNNs. Scholars both domestically and abroad have carried out numerous studies on an intelligent fault diagnosis based on a DNN. Owing to the spatial limitations, only selected representative works are summarized herein.

4.1 Research Status of SAE-based Fault Diagnosis

In recent years, an increasing number of scholars have been paying attention to the use of an SAE to realize a fault diagnosis. Shao [86] developed a novel deep autoencoder feature learning method to diagnose a rotating machinery fault. The maximum correntropy is adopted to design a new deep autoencoder loss function for an enhancement of the feature learning from the measured vibration signals. In addition, an artificial fish swarm algorithm is used to optimize the key parameters of the deep autoencoder to adapt to the signal features. The proposed method was applied to the fault diagnosis of a gearbox and electrical locomotive roller bearing and achieved effective and robust diagnosis results. Chen [87] proposed a new multi-sensor data fusion technique that extracts the time- and frequency-domain features from the different sensor signals of rotating machinery, and these features are then input into multiple two-layer SAE neural networks for feature fusion. Finally, the fused

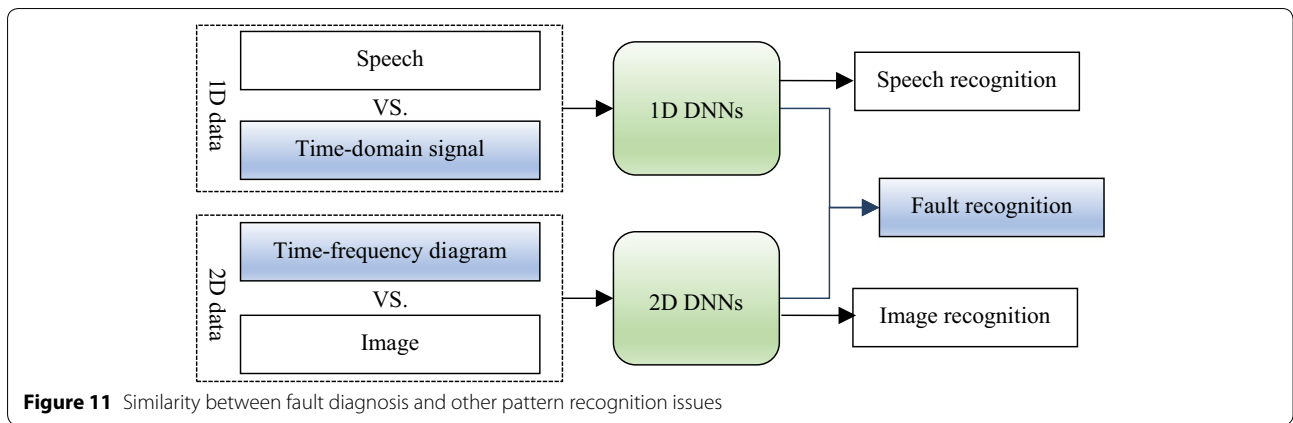


Figure 11 Similarity between fault diagnosis and other pattern recognition issues

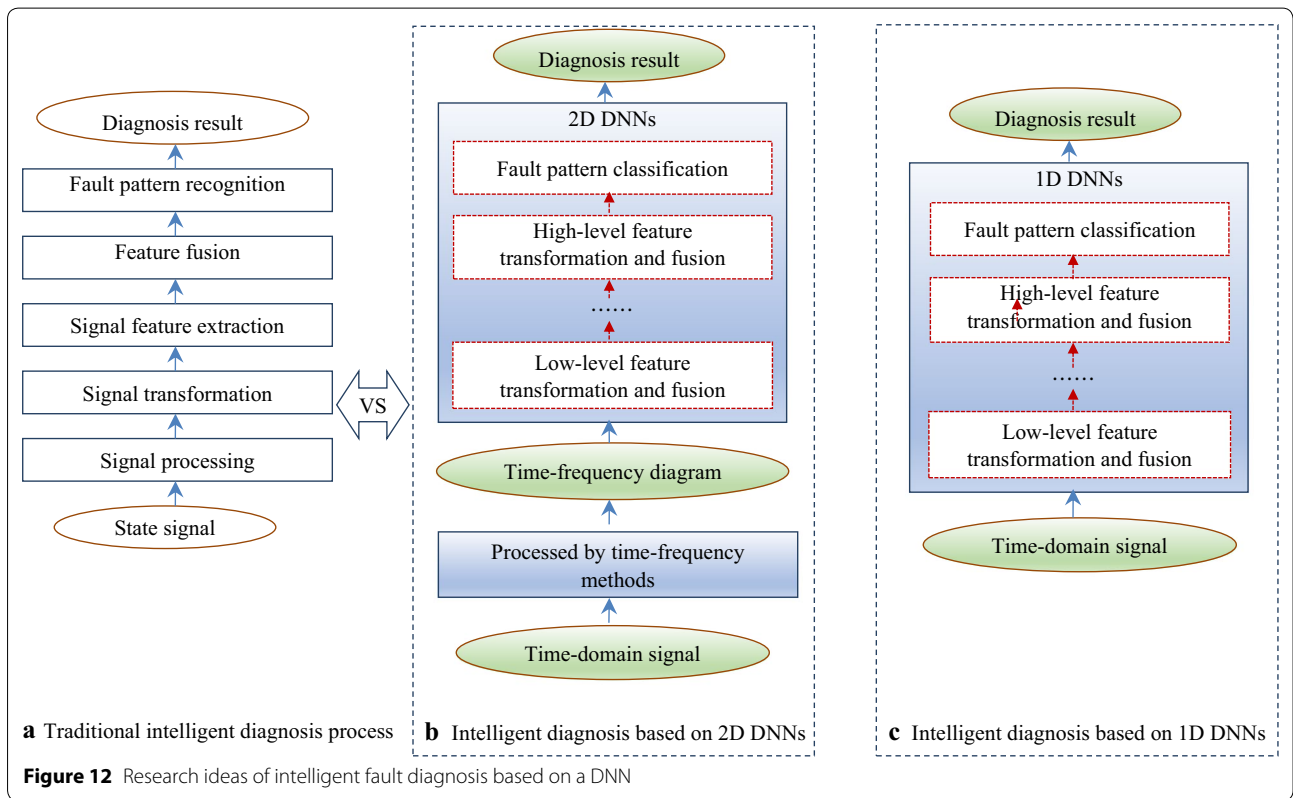


Figure 12 Research ideas of intelligent fault diagnosis based on a DNN

feature vectors are regarded as the machine health indicators, and are used to train a DBN for further classification. Wang [88] proposed a novel continuous sparse autoencoder (CSAE) that adds a Gaussian stochastic unit into the activation function to extract the features of non-linear data. The proposed CSAE is applied to solve the problem of transformer fault recognition and achieve a superior correct differentiation rate. Similar studies have been extended to the fault diagnosis of complex systems, such as a motor [89], centrifugal pump [90], bearing [91,

92], gearbox [93], and rectifier [94], and good results have been achieved.

By using and improving on various SAE deformation algorithms [95–97], scholars have made numerous achievements in terms of fault diagnosis. Since an SAE was first proposed, its main function has been feature-oriented learning and a dimensionality reduction. Thus, when an SAE is used in a fault diagnosis, it can directly learn the sparse and noise reduction of the input signal, extract its robustness characteristics, and then realize a

fault classification. Because of its powerful feature extraction ability, it is suitable for an end-to-end fault diagnosis and can achieve a high performance in the case of a small number of samples.

4.2 Research Status of DBN-based Fault Diagnosis

Since Tamilselvan [98] first applied a DBN to a fault diagnosis of aircraft engine in 2013, an increasing number of scholars both at home and abroad have paid attention to this field and achieved numerous research results. Tran et al. [99] fused a DBN with a TKEO algorithm and applied result to a fault diagnosis of reciprocating compressor valves. They used a GRBM instead of an RBM to overcome the disadvantage in which a traditional RBM can only input binary fault signals; they then combined a DBN with a traditional feature extraction method, achieving a higher level of accuracy. Shao et al. [100] first used a PSO algorithm to optimize the relevant parameters of a DBN, and then applied the optimized DBN algorithm to the fault diagnosis of rotating bearings. In the absence of prior fault information, a better recognition accuracy was achieved, and the difficulty of parameter selection in a DBN model was successfully resolved, providing an extremely good reference.

An increasing number of studies have been conducted on DBN-based fault diagnosis, which has potential advantages when applied to high-speed trains [101], solar cells [102], rolling bearings [103–105], and gearboxes [106]. Moreover, the field of application is widening. Summarizing existing studies, there are two main uses of a DBN for fault detection and recognition: one is to use it as a classifier, and the other is to integrate several steps (feature extraction, feature transformation, information fusion, and pattern recognition) into a single deep structure to realize their joint optimization and complete an intelligent diagnosis. A DBN does not require numerous labeled samples, which not only accelerates the convergence speed of the network, it also achieves excellent diagnostic results, providing technical support for an efficient and in-depth fault diagnosis.

4.3 Research Status of CNN-based Fault Diagnosis

The CNN have achieved significant success in the field of computer vision and image recognition, and has attracted the attention of numerous scholars in the area of intelligent fault diagnosis. Owing to its successful application in the image recognition field, many researchers have treated the fault diagnosis issue as an image recognition problem. Summarizing existing researches, there are mainly two types of studies: The first is converting time-series signals into two-dimensional images as the input of the CNN. Wen et al.

[107] used images of vibration signals as the input of a CNN, and achieved significant improvements in their proposed fault diagnosis method. Hoang et al. [108] adopted a deep CNN structure in the fault diagnosis of rolling bearings and achieved an extremely high accuracy and robustness under noisy environments. The second is using 2D time-frequency images as the input of the CNN for a diagnosis. For example, Guo et al. [109] used the transformation results after a continuous WT as the input of a CNN to diagnose the fault of the rotating machinery. Wang et al. [110] preprocessed an original signal using STFT to obtain a time-frequency diagram, and then applied a CNN to extract the time-frequency features adaptively, to complete the diagnosis. We used a CNN to diagnose a rolling-bearing fault based on a wavelet time-frequency diagram [111, 112].

Fewer studies on a fault diagnosis based on a CNN have been conducted than those based on a DBN or an SAE, the reason for which may be that a CNN is primarily applied to deal with 2D features in its initial application. The real-time state signal is usually a 1D vector. Therefore, some researchers have tried to construct a 1DCNN to process the original signal for a fault diagnosis directly. Turker et al. [113] tested the current of a motor and used a 1DCNN to realize a real-time state monitoring and fault diagnosis. Peng et al. [114] used a 1DCNN to diagnose the faults of HST wheelset bearings with vibration signals and achieved good results. With a 1DCNN, the signal is directly input into the network, and the 1D convolution kernels can be considered different digital filters. The function of a convolution kernel is similar to the sine basis in a Fourier transform or the wavelet basis in a wavelet transform. In addition, the sliding of a kernel is the same as the translation of a wavelet. Their difference lies in the fact that these kernels are acquired automatically and self-adaptively through learning, even unsupervised learning. Thus, a 1DCNN has a better feature extraction capability without a manual intervention.

According to the existing literature, similar to a fault diagnosis method based on a DBN, there are mainly two types of methods: one is to directly use a CNN for feature extraction and fault recognition, and the other is mainly to use a CNN for fault classification. The input of the network can be a time-domain signal, frequency domain feature, time-frequency image, or other extracted feature vector. The unique topological structure of a CNN makes it highly invariant to the translation and scaling of the input sample features. Compared with a DBN and an SAE, a CNN better matches the feature scaling caused by a time shift and change in speed of the mechanical fault signal [115].

4.4 Research Status of RNN-based Fault Diagnosis

The main difference between an RNN and a DBN, an SAE, and a CNN is that an RNN takes full account of the correlation between samples. That is, the pre-and post-samples influence the current sample, and an RNN is suitable for complex equipment or a real-time fault diagnosis. A fault diagnosis based on an RNN was used in the fault diagnosis field earlier than other deep learning methods, and after nearly 10 years of development, its achievements are widespread throughout all relevant fields. Gan [116] explored a model-based recurrent neural network (MBRNN) for use in a fault diagnosis. An MBRNN can use model-based fault detection and isolation (FDI) solutions as a starting point and improve them through training by adapting them to plant nonlinearities. The application of an MBRNN IN in the nonlinear model of an electromechanical governor used in the speed control of large diesel engines indicates that an MBRNN provides better results than “black box” neural networks. Li [117] proposed a fault diagnosis and isolation (FDI) strategy based on a dynamically driven recurrent neural network (DDRNN) architecture for use in situations in which there are thruster/actuator failures in the satellite’s attitude control system. To improve the FDI performance accuracy, the proposed architecture was designed to consist of two DDRNNs. The first determines and diagnoses the presence of a faulty thruster and the second then identifies which thruster is faulty.

In recent years, owing to the powerful feature extraction and pattern recognition capabilities of deep learning, a fault diagnosis algorithm based on an RNN has again attracted extensive attention of scholars. Liu [118] proposed a novel method for a bearing fault diagnosis with an RNN in the form of an autoencoder. With this approach, multiple vibrations of the rolling bearings of the next period are predicted from the previous period by means of a gated recurrent unit (GRU) based denoising autoencoder. Then, for the given input data, the reconstruction errors between the next period data and the output data generated by different GRU-NP-DAEs are used to detect anomalous conditions and classify the fault type. Experiment results indicate that the proposed method achieves a satisfactory performance with strong robustness and high classification accuracy.

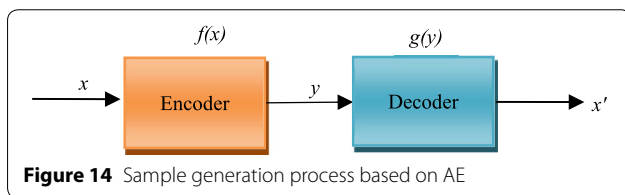
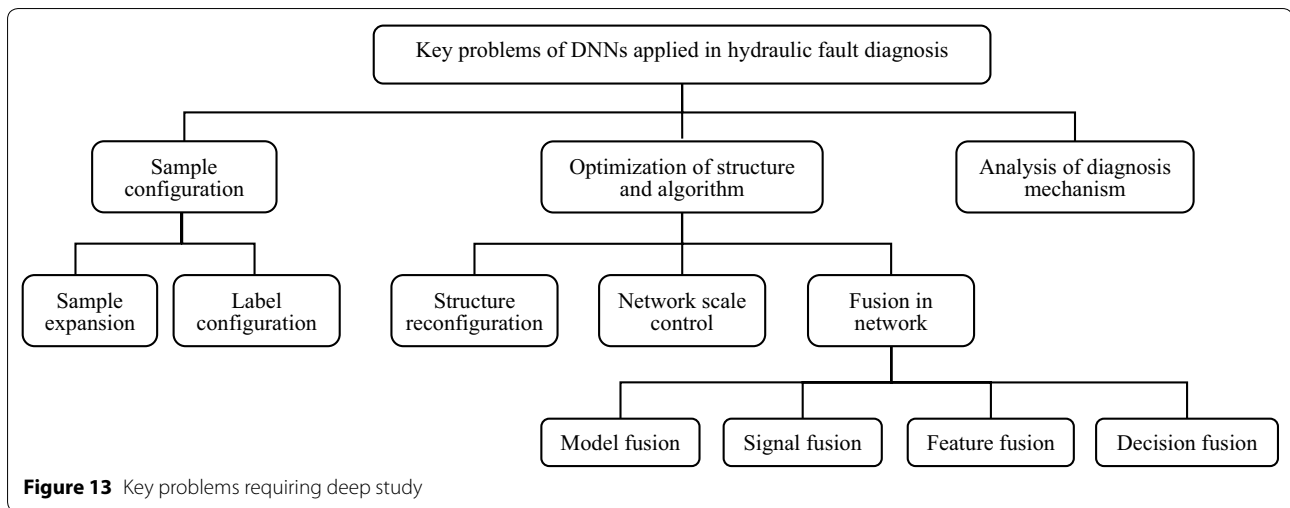
An RNN has incomparable advantages in terms of prediction [119, 120], which has attracted significant attention in the field of deep learning-based fault diagnosis in recent years. Under the background of the large-scale and complex development of a system, an RNN fault diagnosis method will play an increasingly important role.

5 Difficulties and Challenges of DNNs in the Field of Hydraulic Intelligent Fault Diagnosis

5.1 Fault Characteristics of Complex Hydraulic System

Compared with common mechanical and electrical structures, a hydraulic system used in engineering equipment is a highly non-linear system, which is usually complicated in structure, with an electromechanical-hydraulic coupling, and the hydraulic loops intersect each other. The fault mechanisms and forms of the hydraulic components, such as the hydraulic pumps, cylinders, valves, and motors, are complex and diverse. Thus, a hydraulic fault diagnosis remains a challenge. In addition, the hydraulic system used in engineering equipment has the following characteristics:

- (1) The sealing characteristic of the hydraulic system structure results in its faults being concealed, less measurable parameters, and susceptibility to random factors, making it difficult to obtain fault information.
- (2) The motion of the components, such as cylinder reciprocation, and the opening and closing of the control valves, produces a large number of excitation sources, which make the hydraulic state signal demonstrate nonlinear, time-varying, and even impact characteristics [121].
- (3) Owing to the fluid resistance and pressure loss, the vibration transmission mechanism of hydraulic oil is completely different from that used in rigid parts.
- (4) The typical hydraulic faults include stagnation, impact, cavitation, jamming, leakage, and their composite forms. When a composite fault occurs, signals transmitted through complex channels overlap with each other. The mapping relationship between the signal characteristics and the system state is complex. Accordingly, it becomes more difficult to confirm the relationship between the reasons for and the features of a fault [122].
- (5) Complex hydraulic systems consist of many subsystems, and a single component failure may cause a component, subsystem, or system failure in succession. In a short time, multiple different source faults may occur simultaneously, and simple faults occurring at the same time may lead to abnormal functions of multiple subsystems, as well as the concurrency and transformation of multiple types of faults. Therefore, the number of fault modes will rise exponentially with an increase in the system complexity.



5.2 Key Problems of DNNs in Hydraulic Fault Diagnosis

To improve the diagnostic performance of a DNN in a complex hydraulic system further, several key problems need to be deeply studied, as listed in Figure 13.

5.2.1 Expansion of Sample Set

The diagnosis capability of a DNN largely depends on the quality and quantity of the training samples. However, for a hydraulic system, compared with mechanical faults, hydraulic faults are difficult to simulate, and thus there are usually huge amounts of normal signals coexisting with a few fault signals. However, for complex pattern recognition problems such as a hydraulic system fault diagnosis, it is necessary to base a diagnosis on large numbers of samples. Therefore, how to generate more similar samples based on limited fault samples must be considered. Fortunately, in a DNN, prediction and generation models, such as an AE and a GAN, can satisfy this requirement.

As mentioned in Section 3.2.1, an AE consists of two parts including an encoder and a decoder. As shown in Figure 14, an encoder is used to transform and compress the input x into low dimensional codes y . A decoder is applied to recover y for the original input x . Then, the output of decoder, x' , will be compared with x , and the difference between them will be minimized through training. Because a decoder can recover the original data,

it can be used in a fault diagnosis to generate pseudo samples for the training of other deep learning models.

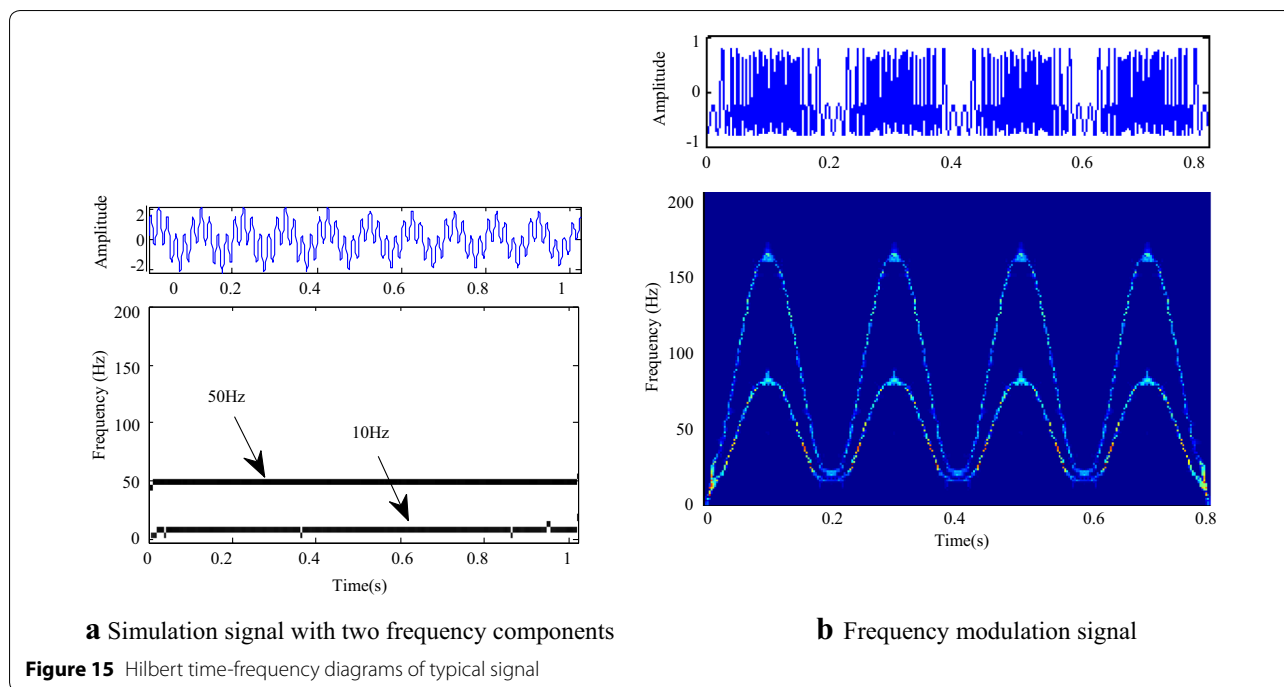
Similarly, as mentioned in Section 3.2.5, a GAN also includes two parts, a generator and a discriminator. The training goal is to make the samples generated by the generator similar to the real samples, such that the discriminator cannot distinguish the real samples from the generated versions. Therefore, the training process is in fact an antagonistic process between the generator and discriminator. It is also a process of identifying the essential features of the samples. Thus, the application of a GAN in a fault diagnosis can also be used to generate pseudo samples to solve the problem of unbalanced samples.

5.2.2 Optimization of Network Structure

(1) Reconfiguration oriented to diagnosis issue

Not all DNNs can be directly used for a fault diagnosis without requiring any changes, and thus some reconfiguration measures should be studied in terms of fault diagnosis.

For example, a shift-invariance is an advantage of a CNN when applied to face recognition. It can eliminate the differences in faces caused by different positions and angles. However, for the time-frequency diagram of the signal, it must be able to locate the frequency precisely while maintaining a shift-invariance on the time axis. Taking the two time-frequency diagrams shown in Figure 15 as examples, in Figure 15(a), there are two frequency components in the simulation signal, which are shown as two straight and parallel lines. It is difficult for a CNN to distinguish them. In Figure 15(b), a frequency modulated signal is shown. The shapes of the two time-frequency lines are similar, and thus a CNN will mistake one for the other and take the former as the stretched



(compressed) result of the latter. Therefore, to solve this type of problem, the structure of the CNN must be reconfigured.

(2) Control of network scale

First, to achieve a better diagnosis result, the scale of the issue should be matched well with that of the network. If the issue is smaller than the network, it may cause an over-fitting. If the issue is larger than the network, it may cause an under-fitting. Thus, the determination of the network scale based on the issue scale is extremely important.

Second, because a hydraulic fault diagnosis is usually a complicated pattern recognition problem, a diagnosis DNN is usually a large-scale network with a large number of neurons, and therefore a large number of parameters. However, for an intelligent built-in diagnostic device, its efficiency will be restricted by the ability and capacity of the hardware. Therefore, it is necessary to obtain a non-redundant and small-scale network structure. Related studies can be carried out from the following aspects:

- 1) The relationship between the width and depth of the network should be studied to find the optimal width-to-depth ratio. Here, the depth is the number of the layers and the width is the scale of a single layer.
- 2) By analyzing the contribution of the neurons during the feature extraction, the less-activated neurons

will be found, and a more simplified structure will be obtained.

(3) Obtain different types of fusion in a network

Four types of fusion in a network can be used, namely, a model fusion, signal fusion, feature fusion, and decision fusion.

1) Model fusion

For a specific problem, to achieve a better recognition and classification, fusing different models in one deep network to benefit from different models comprehensively is an important method, and can be studied from the following aspects.

As shown in Figure 1, in a traditional intelligent diagnosis, a dimensionality reduction and/or sparse processing model is often used in high dimensional data processing and feature selection. Its purpose is to highlight the main characteristics and reduce the difficulty of a calculation as well as the consumption of computing resources. Commonly used models include linear models such as a PCA, an ICA, and an SVD, and non-linear models, such as KPCA, multi-dimensional scaling (MDS), and locally linear embedding (LLE).

In fact, in other pattern recognition areas, there have been some similar studies [123–125]. These studies have shown that a dimensionality reduction achieves an excellent performance in face recognition, and sparse processing is more effective in image processing and audio and video identification. In a previous study,

we tried to integrate a PCA into a CNN and achieved a certain effect. However, considering the nonlinear property of a hydraulic state signal, a nonlinear model may be a better choice.

For an on-line monitoring system, the precise location of the time of the fault's initiation, development, and deterioration is extremely important. In the speech recognition field, the most frequently researched hybrid acoustic models include a DBN-HMM [126] and a CNN-HMM [127]. Such studies have inspired us to combine the time-series models, such as a hidden Markov model (HMM), an RNN, and its variants, namely, long-short term memory (LSTM) and a GRU with a DNN, to describe how the system state changes during an operation. The related issues include the following:

- ① How to select the models based on our problem?
- ② At which layer should the models be applied and how can they be integrated into a DNN?
- ③ How can the effect be evaluated?
- ④ How can an integrated network be further improved?

2) Signal fusion

To provide more evidence for diagnosis, we could apply the following:

- ① Install different sensors at different positions of the hydraulic system to obtain different state signals, including the vibration, pressure, flow, and other factors.
- ② Obtain different features of the same signal by applying different signal transformation methods, such as a Fourier transform, WVD, and wavelet, to obtain the frequency-domain, time-frequency, and statistical characteristics.

Integrating such multi-source or variously formed heterogeneous signals into a DNN to improve the diagnosis reliability is an area worth studying.

3) Feature fusion

Feature fusion means the combinatorial application of features output from different layers.

The feature transformation process of a deep network is similar to that of a wavelet decomposition or an EMD, the decomposition of which is a multi-level separation and a refinement of features. For a diagnosis, all components and features of all levels may be valuable. In the previous study, however, the classification of a deep network, such as a CNN and an SAE, was based only on the output of the last feature transformation layer. The issues worth studying here include the following:

- ① Can the features of different layers reflect the characteristics of a signal from different perspectives?
- ② What characteristics can the output of the different layers represent?
- ③ How can multi-layer features be combinatorially applied to realize a more reliable hydraulic diagnosis?

4) Decision fusion

If used in a diagnosis, the last layer of a DNN should be a classifier, namely, a softmax, an SVM, a boosting, or a K-nearest neighbor (KNN) classifier. The following should be considered:

- ① In view of a specific deep structure, what type of classifier should be selected to match the previous feature extraction layer better, and what type of training algorithm should be designed to realize a joint optimization of the entire network?
- ② To improve the reliability of a diagnosis further, is it possible to configure several different classifiers in the same network to achieve a decision fusion of different classification results? If so, how can the network be trained?

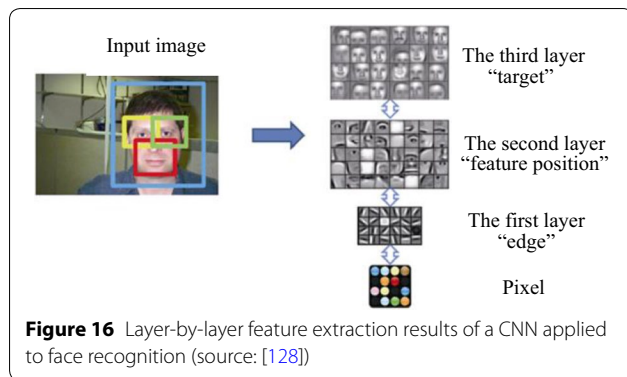
All problems above regarding the optimization of a network structure will correspondingly cause a re-design of the algorithm, including the initialization of the parameters, a pre-processing of the input, a selection, and an optimization of the activation function, among others. Because these problems are extremely specific, and space is limited, the details are not elaborated in this paper.

5.2.3 Analysis of Diagnosis Mechanism

The purpose of an analysis regarding the diagnosis mechanism of a DNN is to answer the following questions:

- ① What is the layer-by-layer transformation result of a DNN?
- ② What is the function of each layer, each neuron, and its parameters?
- ③ Can the transformation result of each layer be effectively interpreted from the perspective of human cognition? Can it in turn inspire humans to have a deeper understanding of hydraulic faults?
- ④ Can we embed human diagnostic knowledge into a network to realize an effective combination of human knowledge and machine learning?

There have been a few similar studies in other pattern recognition fields. For example, as shown in Figure 16, Zeiler et al. [128] analyzed the layer-by-layer transformation results of a CNN used for face recognition,



and found that the layer-by-layer extraction of a CNN includes the contour (outline) information and feature position.

Taking a CNN as an example, if it is used for a fault diagnosis, its input may be a time-domain signal or a 2D time-frequency image. A diagnostic mechanism analysis may be specialized as follows:

- ① What is the state (size and value) of the convolution kernels obtained through training? Can they be considered digital filters as previously supposed? If so, what features does each convolution kernel extract from its input? What are the similarities and differences between a convolution kernel and a traditional digital filter?
- ② What is the relation among the convolution kernels of the same layer? Can these convolution kernels extract all features of the same input? How do they work cooperatively? Is there any redundancy among the convolution results?
- ③ What is the relation among the convolution kernels of two adjacent layers? What features does the latter layer extract from the output of the preceding layer?

6 Conclusions

In this paper, the main technologies used in an intelligent fault diagnosis and the research status of hydraulic system fault diagnosis were analyzed and summarized. Based on a deep analysis of the similarities of a fault diagnosis and other pattern recognition issues, research ideas regarding a hydraulic fault diagnosis based on a DNN were presented. Based on an overview of the research status of fault diagnosis using DNNs in recent years (both domestically and abroad), the characteristics of a complex hydraulic system fault were expounded, and the key problems faced in the application of a DNN to realize a hydraulic fault diagnosis and some possible solutions were presented.

Authors' contributions

JT was in charge of the whole trial and proposed the idea; JD was responsible for the overall work, proposed the idea, performed some experiments and wrote the manuscript; SH and YW performed some of the experiments and contributed to many effective discussions regarding both ideas and paper writing. All authors read and approved the final manuscript.

Authors' information

Juying Dai, born in 1982, is currently a PhD candidate at *Army Engineering University, China*. She received her bachelor degree from *Nanjing University of Aeronautics and Astronautics, China*, in 2007. Her research interests include signal processing and fault diagnosis.

Jian Tang, born in 1977, is an associate professor at *Army Engineering University, China*. She received her PhD degree from *PLA University of Science and Technology, China*, in 2013. Her research interests include signal processing, fault diagnosis and reliability.

Shuzhan Huang, born in 1995, is currently a master candidate at *Army Engineering University, China*.

Yangyang Wang, born in 1995, is currently a master candidate at *Army Engineering University, China*.

Acknowledgements

The authors sincerely thanks to Professor Ting Rui of Army Engineering University for his critical discussion and reading during manuscript preparation.

Competing interests

The authors declare that they have no competing interests.

Funding

Supported by National Natural Science Foundation of China (Grant No. 51705531), and Jiangsu Provincial Natural Science Foundation of China (Grant No. BK20150724).

Received: 14 June 2018 Revised: 22 July 2019 Accepted: 14 August 2019
Published online: 06 September 2019

References

- [1] X Q Li. *Research on key technology of fault prognostic and health management for complex equipment*. Beijing Institute of Technology, 2014. (in Chinese)
- [2] V Venkatasubramanian, R Rengaswamy, S N Kavuri, et al. A review of process fault detection and diagnosis: Part III: Process history based methods. *Computers & Chemical Engineering*, 2003, 27(3): 327–346.
- [3] E Alpaydin. *Introduction to machine learning*. Cambridge MA: The MIT Press, 2004.
- [4] H D M D Azevedo, A M Araújo, N Bouchonneau. A review of wind turbine bearing condition monitoring: state of the art and challenges. *Renewable & Sustainable Energy Reviews*, 2016, 56(4): 368–379.
- [5] H J Zhu, X Q Wang, T Rui, et al. Shift invariant sparse coding for blind source separation of single channel mechanical signal. *Journal of Vibration Engineering*, 2015, 28(4): 625–632. (in Chinese)
- [6] M Van, H J Kang, K S Shin. Rolling element bearing fault diagnosis based on non-local means de-noising and empirical mode decomposition. *Science Measurement & Technology let*, 2014, 8(6): 571–578.
- [7] A Y Goharrizi, N Sepehri. A wavelet-based approach to internal seal damage diagnosis in hydraulic actuators. *IEEE Transactions on Industrial Electronics*, 2010, 57(5): 1755–1763.
- [8] A Y Goharrizi, N Sepehri. Internal leakage detection in hydraulic actuators using empirical mode decomposition and Hilbert spectrum. *IEEE Transactions on Instrumentation & Measurement*, 2012, 61(2): 368–378.
- [9] B Boashash, P Black. An efficient real-time implementation of the Wigner-Ville distribution. *IEEE Trans. on Acoust. Speech Signal Processing*, 1987, 35(11): 1611–1618.
- [10] G S Hu. *Modern signal processing course*. Beijing: Tsinghua University Press, 2004. (in Chinese)

- [11] M J Zhang, J Tang, X H He. *EEMD method and its application in mechanical fault diagnosis*. Beijing: National Defense Industry Press, 2015. (in Chinese)
- [12] Y Amirat, V Choqueuse, M Benbouzid. EEMD-based wind turbine bearing failure detection using the generator stator current homopolar component. *Mechanical Systems & Signal Processing*, 2013, 41(1–2): 667–678.
- [13] J D Zheng, J S Cheng, Y Yang. Improved EEMD algorithm and its application research. *Vibration and Impact*, 2013, 32(21): 21–26. (in Chinese)
- [14] K Chai, M J Zhang, J Huang, et al. Fault diagnosis of hydraulic system based on time-frequency characteristics and PCA-KELM. *Journal of PLA University of Science and Technology*, 2015(4): 394–400. (in Chinese)
- [15] J Huang, J Tang, M J Zhang, et al. An improved EMD based on cubic spline interpolation of extremum centers. *Journal of Vibroengineering*, 2015, 17(5): 2393–2409.
- [16] C Wang, Z L Wang, J Ma, et al. Fault diagnosis for hydraulic pump based on EEMD-KPCA and LVQ. *Vibroengineering Procedia*, 2014, 4(11): 188–193.
- [17] H Chen, M J Zhang, J Huang, et al. Fault diagnosis based on improved EEMD method and GA-SVM for leakage of hydraulic system. *Chinese Hydraulics & Pneumatics*, 2014, (9): 32–38. (in Chinese)
- [18] L M Li, Z S Wang. Feature selection of sudden failure based on affinity propagation clustering. *Advanced Materials Research*, 2012, 586(11): 241–246.
- [19] Z F Li, Y Chai, H F Li. Fault feature extraction method of rolling bearing based on singular value decomposition and morphological filtering. *Application Research of Computers*, 2012, 29(4): 1314–1317. (in Chinese)
- [20] H Yu, F Khan, V Garaniya. A sparse PCA for nonlinear fault diagnosis and robust feature discovery of industrial processes. *Aiche Journal*, 2016, 62(5): 1494–1513.
- [21] D R Huang, C S Chen, G X Sun, et al. Linear discriminant analysis and back propagation neural network cooperative diagnosis method for multiple faults of complex equipment bearings. *Acta Armamentarii*, 2017, 38(8): 1649–1657.
- [22] Q Z Wang, X X Wang. Unified grey relational analysis on transformer DGA fault diagnosis. *Open Mechanical Engineering Journal*, 2014, 8(1): 129–131.
- [23] H M Liu, D W Liu, L Chen, et al. Fault diagnosis of hydraulic servo system using the unscented kalman filter. *Asian Journal of Control*, 2015, 16(6): 1713–1725.
- [24] H Gao, L Liang, X Chen, et al. Feature extraction and recognition for rolling element bearing fault utilizing short-time fourier transform and non-negative matrix factorization. *Chinese Journal of Mechanical Engineering*, 2015, 28(1): 96–105.
- [25] A Czajkowski, K Patan. Robust fault detection by means of echo state neural network. *Advances in Intelligent Systems and Computing*, 2016, 386(8): 341–352.
- [26] H Malik, S Mishra. Application of probabilistic neural network in fault diagnosis of wind turbine using FAST, TurbSim and Simulink. *Procedia Computer Science*, 2015, 58: 186–193.
- [27] Y Chen, Z M Zhen, H H Yu, et al. Application of fault tree analysis and fuzzy neural networks to fault diagnosis in the Internet of Things (IoT) for aquaculture. *Sensors*, 2017, 17(1): 153.
- [28] T Z Wang, J Qi, H Xu, et al. Fault diagnosis method based on FFT-RPCA-SVM for Cascaded-Multilevel Inverter. *ISA Transactions*, 2016, 60(1): 156–163.
- [29] M J Zhang, J Tang, X M Zhang, et al. Intelligent diagnosis of short hydraulic signal based on improved EEMD and SVM with few low-dimensional training samples. *Chinese Journal of Mechanical Engineering*, 2016, 29(2): 396–405.
- [30] Y H Xue, L Zhang, B J Wang, et al. Nonlinear feature selection using Gaussian kernel SVM-RFE for fault diagnosis. *Applied Intelligence*, 2018, 48(10): 3306–3331.
- [31] Y C Xiao, N Kang, Y Hong, et al. Misalignment fault diagnosis of DFWT Based on IEMD energy entropy and PSO-SVM. *Entropy*, 2017, 19(1): <https://doi.org/10.3390/e19010006>.
- [32] X Y Zhang, Y T Liang, J Z Zhou, et al. A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM. *Measurement*, 2015, 69(6): 164–179.
- [33] J Shang, M Y Chen, H Q Ji, et al. Dominant trend based logistic regression for fault diagnosis in nonstationary processes. *Control Engineering Practice*, 2017, 66(9): 156–168. (in Chinese)
- [34] P G Sreenath, G P Kumare, P Sundar, et al. Automobile gearbox fault diagnosis using Naive Bayes and decision tree algorithm. *Applied Mechanics & Materials*, 2015, 813–814: 943–948.
- [35] C Annachhatre, T H Austin, M Stamp. Hidden Markov models for malware classification. *Journal of Computer Virology & Hacking Techniques*, 2015, 11(2): 59–73.
- [36] L Enrique Sucar, C Bielza, E F Morales, et al. Multi-label classification with Bayesian network-based chain classifiers. *Pattern Recognition Letters*, 2014, 41(1): 14–22.
- [37] W L Lu. *Daquan of troubleshooting and repair for practical hydraulic machinery*. Changsha: Hunan Science and Technology Publishing House, 1995.
- [38] G G Ji, N Li, D M Xu. Fault analysis and removal of hydraulic drive system. *Plant Maintenance Engineering*, 1991, 3: 36–38.
- [39] L An, N Sepehri. Hydraulic actuator circuit fault detection using extended Kalman filter. *American Control Conference, IEEE*, 2003, 5(1): 4261–4266.
- [40] L An, N Sepehri. Leakage fault identification in a hydraulic positioning system using extended Kalman filter. *American Control Conference, Proceedings of the IEEE*, 2004, (4): 3088–3093.
- [41] H L Zhu, L Q Gao. Fault diagnosis of hydraulic system based on flow signal. *Journal of Engineering Science*, 2001, 23(1): 66–70.
- [42] Q L Du, K H Zhang. Condition monitoring and fault diagnosis of hydraulic pump based on inherent vibration signals. *Journal of Agricultural Engineering*, 2007, 23(4): 120–123. (in Chinese)
- [43] H X Chen, Patrick S K Chua, G H Lim. Vibration analysis with lifting scheme and generalized cross validation in fault diagnosis of water hydraulic system. *Journal of Sound and Vibration*, 2007, 301: 458–480.
- [44] W L Jiang, S Q Zhang, Y Q Wang. Wavelet transform method for fault diagnosis of hydraulic pump. *Journal of Mechanical Engineering*, 2001, 37(6): 34–37. (in Chinese)
- [45] W Z Du, Z F Zhou, X X Huang. The application of wavelet analysis to hydraulic cylinder leakage detection. *Machine Tools and Hydraulic Pressure*, 2003(6): 318–319. (in Chinese)
- [46] K H Abbott, P C Schutte. Faultfinder: A diagnostic expert system with graceful degradation for onboard aircraft applications. *Mitteilung-Deutsche Forschungs- und Versuchsanstalt fuer Luft- and Raumfahrt*, 1988: 353–370.
- [47] W J Crowther, K A Edge, C R Burrows, et al. Fault diagnosis of a hydraulic actuator circuit using neural networks - An output vector space classification approach. *Proceedings of the Institution of Mechanical Engineers. Part I: Journal of Systems & Control Engineering*, 1997, 212(1): 57–68.
- [48] S Amin, C Byington, M Watson. Fuzzy inference and fusion for health state diagnosis of hydraulic pumps and motors. *Fuzzy Information Processing Society, Nafips Meeting of the North American*. IEEE, 2005.
- [49] H X Chen, Chua P S K, Lim G H. Feature extraction, optimization and classification by second generation wavelet and support vector machine for fault diagnosis of water hydraulic power system. *International Journal of Fluid Power*, 2006, 7(2): 39–52.
- [50] R A Saeed, A N Galybin, V Popov. 3D fluid-structure modelling and vibration analysis for fault diagnosis of Francis turbine using multiple ANN and multiple ANFIS. *Mechanical Systems and Signal Processing*, 2013, 34(1–2): 259–276.
- [51] W L Jiang, Y Q Wang, X D Kong, et al. New progress of fault detection and diagnosis technology for hydraulic system. *China Mechanical Engineering*, 1998, 9(9): 58–60. (in Chinese)
- [52] H W Mou. *Fault diagnosis and expert system research of coking coal machinery hydraulic system*. Hangzhou: Zhejiang University, 2008. (in Chinese)
- [53] C Q Lu. *Fault diagnosis of hydraulic pumps based on HHT and fuzzy C mean clustering*. Qinhuangdao: Yanshan University, 2012. (in Chinese)
- [54] H B Tang. *Research on key technology of fault diagnosis for pumping hydraulic system of concrete pump truck*. Guangzhou: Zhongnan University, 2012. (in Chinese)
- [55] N Saravanan, K I Ramachandran. Fault diagnosis of spur bevel gear box using discrete wavelet features and Decision Tree classification. *Expert Systems with Applications*, 2009, 36(5): 9564–9573.

- [56] J B Ali, N Fnaiech, L Saidi, et al. Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Applied Acoustics*, 2015, 89(3): 16–27.
- [57] D Yao, J W Yang, X Li, et al. A hybrid approach for fault diagnosis of railway rolling bearings using STWD-EMD-GA-LSSVM. *Mathematical Problems in Engineering*, 2016, 2016(10): 1–7.
- [58] G E Hinton, R R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 2006, 313(5786): 504–507.
- [59] Y Lecun, Y Bengio, G Hinton. Deep learning. *Nature*, 2015, 521(7553): 436.
- [60] Y Bengio. Learning deep architectures for AI. *Foundations & Trends in Machine Learning*, 2009, 2: 1–127.
- [61] S Bengio, F Pereira, Y Singer, et al. Group sparse coding. *International Conference on Neural Information Processing Systems*, 2009, 22: 82–89.
- [62] P Vincent, H Larochelle, I Lajoie, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 2010, 11(12): 3371–3408.
- [63] D H Ackley, G E Hinton, T J Sejnowski. A learning algorithm for boltzmann machines. *Cognitive Science*, 1985, 9(1): 147–169.
- [64] J Schmidhuber. Deep learning in neural networks: An overview. *Neural Networks*, 2015, 61: 85–117.
- [65] H R Li, S S Gu. A fast parallel algorithm for a recurrent neural network. *Acta Automatica Sinica*, 2004, 30(4): 516–522.
- [66] C Ledig, L Theis, F Huszar, et al. Photo-realistic single image super-resolution using a generative adversarial Network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: <https://doi.org/10.1109/cvpr.2017.19>.
- [67] X Feng, Y D Zhang, J Glass. Speech feature denoising and dereverberation via deep autoencoders for noisy reverberant speech recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing*, Florence, Italy, May 04–09, 2014: 1759–1763.
- [68] J Maria, J Amaro, G Falcao. Stacked autoencoders using low-power accelerated architectures for object recognition in autonomous systems. *Neural Processing Letters*, 2016, 43(2): 445–458.
- [69] J Zabalza, J C Ren, J B Zheng, et al. Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. *Neurocomputing*, 2016, 214(C): 1062.
- [70] G E Dahl, D Yu, L Deng, et al. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio Speech & Language Processing*, 2012, 20(1): 30–42.
- [71] Z Q Zhao, L C Jiao, J Q Zhao, et al. Discriminant deep belief network for high-resolution SAR image classification. *Pattern Recognition*, 2017, 61(1): 686–701.
- [72] P Zhong, Z Q Gong, S T Li, et al. Learning to diversify deep belief networks for hyperspectral image classification. *IEEE Transactions on Geoscience & Remote Sensing*, 2017, 55(6): 3516–3530.
- [73] T N Sainath, A R Mohamed, B Kingsbury, et al. Deep convolutional neural networks for LVCSR. *IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, May 26–31, 2013: 8614–8618.
- [74] A Krizhevsky, I Sutskever, G E Hinton. ImageNet classification with deep convolutional neural networks. *International Conference on Neural Information Processing Systems*, Lake Tahoe, Nevada, USA, December 03–06, 2012: 1097–1105.
- [75] Y Sun, X G Wang, X O Tang. Deep learning face representation by joint identification-verification. *Advances in Neural Information Processing Systems*, 2014, 27(6): 1988–1996.
- [76] A Karpathy, G Toderici, S Shetty, et al. Large-scale video classification with convolutional neural networks. *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, June 23–28, 2014: 1725–1732.
- [77] G Arevian, C Panchev. Optimising the hystereses of a two context layer RNN for text classification. *International Joint Conference on Neural Networks*, Orlando, Florida, USA, August 12–17, 2007: 2936–2941.
- [78] S Alsenan, M Ykhlef. Statistical machine translation context modelling with recurrent neural network and LDA. *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics*, 2016, 533: 75–84.
- [79] G Lev, G Sadeh, B Klein, et al. RNN fisher vectors for action recognition and image annotation. *Computer Science*, 2015, 9910(9): 833–850.
- [80] T Tsujimoto, Y Takahashi, S Takeuchi, et al. RNN with Russell's circumplex model for emotion estimation and emotional gesture generation. *Evolutionary Computation*, Vancouver, BC, Canada, November 21, 2016: 1427–1431.
- [81] S E Kahou, V Michalski, K Konda, et al. Recurrent neural networks for emotion recognition in video. *International Conference on Multimodal Interaction*, Seattle, Washington, USA, November 09–13, 2015: 467–474.
- [82] K F Wang, Y Lu, Y T Wang, et al. Parallel imaging: a new theoretical framework for image generation. *Pattern Recognition and Artificial Intelligence*, 2017, 30(7): 577–587.
- [83] Y J Liu, C H Dou, Q L Zhao. Hand-drawn image retrieval based on condition generation against network. *Journal of Computer Aided Design and Graphics*, 2017, 29(12): 2336–2342.
- [84] L F Mo, H L Jiang, S P Li. Video prediction based on deep learning: A review. *Journal of intelligent systems*, 2018, (1): 85–96.
- [85] F Jia, Y G Lei, J Lin, et al. Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mechanical Systems & Signal Processing*, 2016, 72(5): 303–315.
- [86] H D Shao, H K Jiang, H W Zhao, et al. A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 2017, 95(10): 187–204.
- [87] Z Y Chen, W H Li. Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. *IEEE Transactions on Instrumentation and Measurement*, 2017, 99(3): 1–10.
- [88] L K Wang, X Y Zhao, J N Pei, et al. Transformer fault diagnosis using continuous sparse autoencoder. *SpringerPlus*, 2016, 5(1): 448.
- [89] L H Wang, Y Y Xie, Y H Zhang, et al. A fault diagnosis method for asynchronous motor using deep learning. *Journal of Xi'an Jiaotong University*, 2017, 51(10): 128–134. (in Chinese)
- [90] G He, Y L Cao, T F Ming, et al. Cavitation state recognition of centrifugal pump based on features of modified octave bands. *Journal of Harbin Engineering University*, 2017, 38(8): 1263–1267, 1302. (in Chinese)
- [91] D T Hoang, H J Kang. A bearing fault diagnosis method based on autoencoder and particle swarm optimization – Support Vector Machine. *Intelligent Computing Theories and Application*. Springer, Cham, 2018.
- [92] S Q Tao, T Zhang, J Yang, et al. Bearing fault diagnosis method based on stacked autoencoder and softmax regression. *34th Chinese Control Conference (CCC)*, IEEE, Hangzhou, China, 2015.
- [93] G F Liu, H Q Bao, B K Han. A stacked autoencoder-based deep neural network for achieving gearbox fault diagnosis. *Mathematical Problems in Engineering*, 2018, 2018: 1–10.
- [94] L Xu, M Y Cao, B Y Song, et al. Open-circuit fault diagnosis of power rectifier using sparse autoencoder based deep neural network. *Neurocomputing*, 2018, 311(10): 1–10.
- [95] X N Zhang, X Zhou, C H Tang. A deep convolutional auto-encoding neural network and its application in bearing fault diagnosis. *Journal of Xi'an Jiaotong University*, 2018. (in Chinese)
- [96] F T Wang, X F Liu, B S Guo, et al. Application of kernel auto-encoder based on firefly optimization in intershaft bearing fault diagnosis. *Journal of Mechanical Engineering*, 2019, 55(7): 58–64. (in Chinese)
- [97] B She, F Q Tian, W G Liang. Fault diagnosis based on a deep convolution variational autoencoder network. *Chinese Journal of Scientific Instrument*, 2018, 39(10): 27–35. (in Chinese)
- [98] P Tamilselvan, P Wang. Failure diagnosis using deep belief learning based health state classification. *Reliability Engineering & System Safety*, 2013, 115(7): 124–135.
- [99] V T Tran, F Althobiani, A Ball. An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks. *Expert Systems with Applications*, 2014, 41(9): 4113–4122.
- [100] H D Shao, H K Jiang, X Zhang, et al. Rolling bearing fault diagnosis using an optimization deep belief network. *Measurement Science and Technology*, 2015, 26(11): 115002.
- [101] J P Xie, Y Yang, T R Li, et al. Learning features from high speed train vibration signals with deep belief networks. *International Joint Conference on Neural Networks*, Beijing, China, July 06–11, 2014: 2205–2210.
- [102] X B Wang, J Li, M H Yao, et al. Solar cells surface defect detection based on deep learning. *Pattern Recognition and Artificial Intelligence*, 2014, 27(6): 517–523. (in Chinese)

- [103] Y F Li, X Q Wang, M J Zhang, et al. An approach to fault diagnosis of rolling bearing using SVD and multiple DBN classifiers. *Journal of Shanghai Jiaotong University*, 2015, 49(5): 681–686, 694. (in Chinese)
- [104] X Q Wang, Y F Li, T Rui, et al. Bearing fault diagnosis method based on Hilbert envelope spectrum and deep belief network. *Journal of Vibroengineering*, 2015, 17(3): 1295–1308.
- [105] M Gan, C Wang, C A Zhu. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mechanical Systems & Signal Processing*, 2016, 72–73(2): 92–104.
- [106] Z Chen, C Li, R V Sánchez. Multi-layer neural network with deep belief network for gearbox fault diagnosis. *Journal of Vibroengineering*, 2015, 17(5): 2379–2392.
- [107] L Wen, X Y Li, L Gao, et al. A new convolutional neural network based data-driven fault diagnosis method. *IEEE Transactions on Industrial Electronics*, 2018, 65(11): 5990–5998.
- [108] D T Hoang, H J Kang. Rolling element bearing fault diagnosis using convolutional neural network and vibration image. *Cognitive Systems Research*, 2019, 53(1): 42–50.
- [109] S Guo, T Yang, G Wei. A novel fault diagnosis method for rotating machinery based on a convolutional neural network. *Sensors*, 2018, 18(5): 1429.
- [110] L H Wang, X P Zhao, J X Wu, et al. Motor fault diagnosis based on short-time Fourier transform and convolutional neural network. *Chinese Journal of Mechanical Engineering*, 2017, 30(6): 1357–1368.
- [111] J H Yuan, T Han, J Tang, et al. Intelligent fault diagnosis method for rolling bearings based on wavelet time-frequency diagram and CNN. *Machine Design and Research*, 2017, (2): 93–97. (in Chinese)
- [112] J H Yuan, J Tang. Intelligent fault diagnosis method for rolling bearings based on MWT and CNN. *Mechanical Transmission*, 2016, (12): 139–143. (in Chinese)
- [113] T Ince, S Kiranyaz, L Eren, et al. Real-time motor fault detection by 1D convolutional neural networks. *IEEE Transactions on Industrial Electronics*, 2016, 63(11): 7067–7075.
- [114] D D Peng, Z L Liu, H Wang, et al. A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains. *IEEE Access*, 2018, 99(12): 10278–10293.
- [115] D Cireşan, U Meier, J Schmidhuber. Multi-column deep neural networks for image classification. *Computer Vision and Pattern Recognition*, IEEE, 2012.
- [116] C Y Gan, K Danaei. Fault diagnosis of the IFAC Benchmark Problem with a model-based recurrent neural network. *IEEE International Conference on Control Applications*, IEEE, 1999.
- [117] L Li, L Y Ma, K Khorasani. A dynamic recurrent neural network fault diagnosis and isolation architecture for satellite's actuator/thruster failures. *International Symposium on Neural Networks*, Springer, Berlin, Heidelberg, 2005.
- [118] H Liu, J Z Zhou, Y Zheng, et al. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Transactions*, 2018, 77(6): 167–178.
- [119] D Željko, M Randić, G Krčelić. A multivariate approach to predicting quantity of failures in broadband networks based on a recurrent neural network. *Journal of Network System and Management*, 2016, 24(1): 189–221.
- [120] C Xu, G Wang, X G Liu, et al. Health status assessment and failure prediction for hard drives with recurrent neural network. *IEEE Trans. on Computers*, 2016, 65(11): 3502–3508.
- [121] Z Zhao, F L Wang, M X Jia, et al. Intermittent chaos and cepstrum analysis based early fault detection on shuttle valve of hydraulic tube tester. *IEEE Transactions on Industrial Electronics*, 2009, 56(7): 2764–2770.
- [122] J Du, S P Wang, H Y Zhang. Layered clustering multi-fault diagnosis for hydraulic piston pump. *Mechanical Systems & Signal Processing*, 2013, 36(2): 487–504.
- [123] S N Borade, R R Deshmukh, S Ramu. Face recognition using fusion of PCA and LDA: borda count approach. *Control and Automation*, Athens, Greece, June 21–24, 2016: 1164–1167.
- [124] Zheng-Ping Hu. Sparse representation algorithm for image recognition based on the combination of structured sparse and atom sparse. *Journal of Signal Processing*, 2013, 29(7): 888–895.
- [125] Z Cui, H Chang, S Shan, et al. Joint sparse representation for video-based face recognition. *Neurocomputing*, 2014, 135(8): 306–312.
- [126] L Badino, C Canevari, L Fadiga, et al. Deep-level acoustic-to-articulatory mapping for DBN-HMM based phone recognition. *Spoken Language Technology Workshop*, Miami, FL, USA, December 02–05, 2013: 370–375.
- [127] O Koller, S Zargaran, H Ney, et al. Deep sign: hybrid CNN-HMM for continuous sign language recognition. *British Conference on Machine Vision*, York, British, September 19–22, 2016.
- [128] M D Zeiler, R Fergus. Visualizing and understanding convolutional networks. *European Conference on Computer Vision*, 2014: 818–833.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
