Artificial Intelligence-Based Technique for Fault Detection and Diagnosis of EV Motors: A Review

Wangjie Lang, *Student Member, IEEE*, Yihua Hu, *Senior Member, IEEE*, Chao Gong, *Student Member, IEEE*, Xiaotian Zhang, *Student Member, IEEE*, Hui Xu, and Jiamei Deng, *Senior Member, IEEE*

***Abstract*— The motor drive system plays a significant role in the safety of electric vehicles as a bridge for power transmission. Meanwhile, to enhance the efficiency and stability of the drive system, more and more studies based on AI technology are devoted to the fault detection and diagnosis (FDD) of the motor drive system. This article reviews the application of AI techniques in motor FDD in recent years. AI-based FDD is divided into two main steps: feature extraction and fault classification. The application of different signal processing methods in feature extraction is discussed. In particular, the application of traditional machine learning and deep learning algorithms for fault classification is presented in detail. In addition, the characteristics of all techniques reviewed are summarized. Finally, the latest developments, research gaps, and future challenges in fault monitoring and diagnosis of motor faults are discussed.**

***Index Terms*— AI-based techniques, deep learning, machine learning (ML), motor fault.**

I. Introduction

FAULT detection and diagnosis (FDD) is a condition monitoring technique that can be used to identify the local or overall operating condition of an electric motor, detecting early faults, and making predictions. The purpose of this technology is to detect faults and distinguish different types of faults to make decisions in advance to prevent the happening of hazards. In addition, modern industrial technology has been developed significantly; electric motors have been widely used in different areas of industrial systems, especially playing a key role in the powertrain of electric vehicles. Fig. 1 illustrates the basic architecture of an electric vehicle drive system. The conventional powertrain of an electric vehicle consists of electrical and mechanical systems. The electrical part includes the electric motor, the battery management unit, the power electronic devices, and the controller module, while the mechanical part includes the gearbox and the wheels.

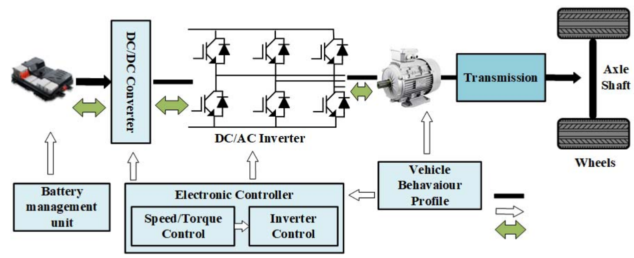


Fig. 1. Basic structure of the motor drive system.

Therefore, the motor is the most important component in the conversion of electrical energy to mechanical energy. The operational status and related parameters of the electric machine can be monitored and acquired in real time through an online system. However, as the operating time increases, the possibility of motor failure under different operating conditions raises, threatening the reliability and safety of electric vehicles.

Electric motors suffer from several types of faults [1]. Generally, the two common forms of faults are the rotor and bearing faults in electric motors. The failure of one component of a motor will lead to a chain reaction, which can fail or even paralysis of the entire equipment system [2]. Therefore, it has become imperative to enhance the safety and reliability of motor drive systems. The effective diagnosis of electrical or mechanical faults in electric motors is necessary [3]–[6]. For noninvasive motor FDD techniques, different signals can be analyzed, including temperature analysis, vibration analysis, infrared analysis, current and voltage analysis, electromagnetic field analysis, ultrasonic analysis, and so on. Vibration analysis [7]–[9] and motor current signature analysis (MCSA) [10], [11] are two of the most popular research in this field. The stator current analysis is known for providing noninvasive condition monitoring for EV motors [12]–[14]. Accurate detection of potential or existing motor faults is an essential measure to maintain safe machine operation. Among the theoretical methods and approaches that have been studied, the technical strategy of motor fault diagnosis in different external environments and operating conditions is a guarantee of improving the reliable operation of the equipment system, which will bring certain limitations [15], [16].

Manuscript received May 22, 2021; revised August 3, 2021; accepted August 26, 2021. Date of publication September 3, 2021; date of current version March 22, 2022. This work was supported by the Newton Advanced Fellowship, U.K., under Grant NAF\R1\191153. *(Corresponding author: Chao Gong.)*

Wangjie Lang, Yihua Hu, Chao Gong, and Xiaotian Zhang are with the Department of Electronic Engineering, University of York, York YO10 5DD, U.K. (e-mail: wl1526@york.ac.uk; yihua.hu@york.ac.uk; 1452101806@qq.com; xz2764@york.ac.uk).

Hui Xu and Jiamei Deng are with the School of Built Environment, Engineering and Computing, Leeds Beckett University, Leeds LS6 3QS, U.K. (e-mail: xuhuicunt2001@gmail.com; j.deng@leedsbeckett.ac.uk).

Digital Object Identifier 10.1109/TTE.2021.3110318

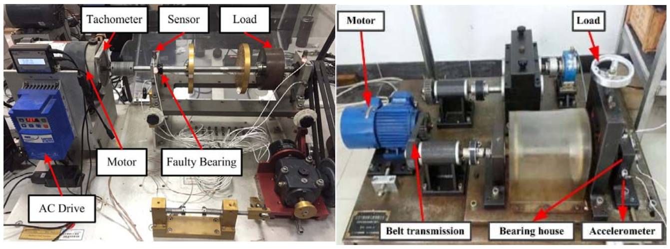


Fig. 2. Experimental rig [42], [106].

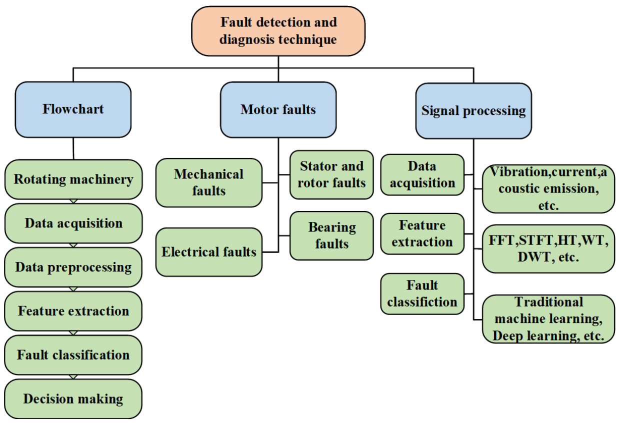


Fig. 3. General flowchart of FDD techniques.

In general, the motor FDD techniques are divided into the model- and data-based methods. Model-based methods for motor fault detection are common in industrial applications. The simplest approach is based on input and output signal processing. If the output of the motor is outside the normal range of variation, the motor is considered faulty or about to fail. Mathematical models are used to describe the output amplitude, phase, frequency, and correlation with the source of the fault, and then, these quantities are processed to determine the location and cause of the fault [17]. Moreover, classical methods based on state estimation or process parameter estimation are also commonly used for motor fault detection [18]. The advantage of these methods is that they can drill down in the basic dynamic characteristics of the motor system for real-time diagnosis, but an accurate mathematical model of the motor is required. Therefore, such methods are difficult to implement when the motor system model is uncertain or nonlinear. Fig. 2 indicates the experimental rig for the application of motor drive system FDD techniques, which includes the motor, controller, load, sensors, and so on.

In recent years, the data-based approach has gained popularity due to its high practicality. It is a suitable method for incorporating artificial intelligence into FDD. Fig. 3 illustrates the general flowchart of FDD techniques. AI-based FDD is divided into two main steps: 1) feature extraction and 2) fault classification. In signal-based feature extraction engineering, the time-domain signal is usually converted to the frequency domain signal by the discrete Fourier transform. Simultaneously, time and frequency signal analysis becomes extremely important when the motor is under dynamic and transient conditions. Short-time Fourier transforms (STFTs), Hilbert transforms (HTs), wavelet transform (WT), and other signal processing techniques have been widely proposed. Furthermore, the time-domain FDD analysis is the best choice when using minimal computational resources to process extremely complex and intermittent signals. In addition, AI-based fault classification methods include machine learning (ML), deep learning, fuzzy neural networks, genetic algorithms, and hybrid algorithms, which can address issues that cannot be solved by traditional fault diagnosis methods [19].

Artificial intelligence is the key point of data-driven technologies. Artificial intelligence-based studies have been extensively carried out in industry applications. Artificial neural networks (ANNs) use activation functions to predict interactions between artificial neurons. The related weights and biases are used to model the biological neural system. Nonlinear characteristics can be extracted and combined using ANNs, which can be employed to perform classification and regression tasks in the motor drive system. In [20], ANNs were applied to track and diagnose external faults in three-phase induction motors (IMs). The data of stator voltage, current, and motor speed were used to train the network. Furthermore, the identification and diagnosis of IM faults implemented in supervised and unsupervised neural networks have been achieved in [21] as well. The application of neural networks in the classification of partial discharges in motor insulation was studied in [22], where the fault of an IM is determined by the imbalance of the inductor current. Aksoy *et al.* [23] investigated the state estimation of IMs based on classical ML algorithm with nonlinear state estimators, which is based on stator current and rotor angular velocity measurements. In addition, DL can model and obtain accurate classification and predictions for complex fault types. Machine fault diagnosis has been effectively implemented using several common deep architectures, including autoencoder (AE), recurrent neural network (RNN), generative adversarial network (GAN), convolutional neural network (CNN), and deep confidence network (DBN). Research on artificial intelligence-based techniques is significant for providing valuable research directions, but there is no comprehensive overview of the application of artificial intelligence-based techniques in motor FDD. It is important to investigate the contributions of different scholars in this field.

This article aims to review the methods of motor FDD based on artificial intelligence techniques. The full article is divided into eight sections. Section II describes the common electric motor and fault types. The different fault monitoring techniques are discussed in Section III. The popular data processing feature extraction methods are presented in Section IV. Section V introduces classical ML and deep learning algorithms applied for fault classification. Section VI presents the other hybrid algorithms with artificial intelligence. In Sections VII and VIII, suggestions for future development and an overall summary are given, respectively.

II. Electric Motor and Fault Types

In the advanced powertrain system, IMs and permanent magnet synchronous motors (PMSMs) are the two major types of electric vehicle motors. IMs are widely used in industry and are characterized by stator and rotor made of laminated silicon

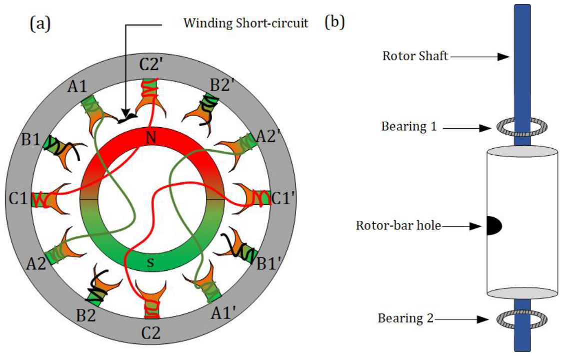


Fig. 4. Typical faults: (a) stator WSC and (b) CRF.

steel sheets and encapsulated with aluminum caps at both ends. Between the stator and the rotor, there are no mechanical elements in connection with each other, and the structure is simple with reliable operation. IMs are more efficient and appropriate for high-speed operation than dc motors of the same power. The principle of operation is to generate an induction current in the rotor through the rotating magnetic field of the stator to produce electromagnetic torque, and no magnetic field is generated in the rotor; therefore, the rotor speed is less than the synchronous speed. Therefore, it is also called the asynchronous motor. Due to the complex mechanical structure of dc motors, which restricts the overall vehicle performance, synchronous motors are gradually becoming popular drive motors. The rotor of PMSM produces a fixed direction magnetic field, and the stator rotating magnetic field drags the rotor magnetic field to rotate, so the rotor speed is equal to the synchronous speed. PMSM has the advantages of high efficiency, high starting torque, and high power factor.

Electric motors and drives are affected by many different types of faults.

1) Stator faults can be classified as stator open phase, and the short circuit leads to stator imbalance or increased resistance connections. Loose electrical parts or poor connections can cause heat and eventually fire.
2) Rotor electrical faults, which include rotor open phase, rotor imbalance due to short circuits or increased resistance connections and broken bars or cracked end rings in IMs, and rotor magnetic failures, such as demagnetization in permanent magnet synchronous machines.
3) Rotor mechanical faults, such as bearings damage, bent shaft, eccentricity, and misalignment.
4) Faults of the power electronic components in the motor drive system.

Fig. 4 illustrates the two typical faults of stator winding short-circuit (WSC) and rotor crack fault (CRF). Listed below are the main categories of faults that can be detected with the MCSA. The frequency caused by each type of fault depends on the specific characteristics of the motor and the operating conditions.

A. Stator Winding Fault

Most winding faults result from the growth of uncorrected turn-to-turn defects. Long-term thermal aging and eventually insulation failure are the major causes of turn-to-turn defects. It may result in the opening, shorting, or grounding of one or more winding circuits, excessive heating, and machine failure.

B. Bearing Faults

Bearing faults can be caused by several factors.

1) Due to the huge output load torque, there is a high vibration in the rotor, which leads to large fatigue stress.
2) Incorrect bearing installation.
3) High currents in the bearing directed by shaft voltage cause lubrication to deteriorate.
4) Heat transferred to the shaft, resulting in friction and pollution.

Bearing faults eccentricate the rotor, resulting in imbalanced magnetic forces and increased bearing stresses. Because the shaft dynamics are impacted by the distorted air gap between the stator and rotor, as well as variations in bearing stiffness, bearing failure is one of the reasons for excessive motor vibration.

Air-gap eccentricity occurs when the air gap distance between the rotor and the stator is not uniform. The two types of abnormal air-gap eccentricity are static and dynamic abnormal air-gap eccentricity. In the case of static eccentricity, the location of the minimal radial air gap is constant, but, in the case of dynamic eccentricity, the position of the minimal air gap follows the rotation of the rotor. When the rotor bars retreat or approach the stator magnetic fields, the current in the stator varies.

C. Broken Rotor Bar Faults

Broken rotor bars are caused by faults in the rotor bars and end rings (BRB). Frequent starting under rated voltage, thermal imbalance, overload at the beginning (thermal stress), and unbalanced magnetic tension are all potential causes of BRB. These faults induce localized heating or arcing in the rotor, as well as vibration from the expansion and bending of the rotor.

III. Fault Monitoring Techniques

MCSA is a condition monitoring approach for diagnosing electric motor issues. The idea was first presented in the early 1970s for use with inaccessible motors in nuclear power plants and motors located in dangerous regions. It has rapidly gained recognition in the industry in recent years. Motors are operated under load and evaluated online without stopping production under regular operating circumstances. MCSA can be used as a preventative maintenance technique to identify typical motor faults before they become serious, which prevents costly catastrophic faults, production interruptions, and extend motor life. MCSA is an electrical characterization (ESA) approach that may be used to assess electric motors, generators, power transformers, and other electrical equipment. The MCSA is used to detect faults by monitoring the stator current of the motor. Only one of the three-phase supply currents of the motor is typically monitored with a single stator current monitoring device. The MCSA employs the motor stator winding as a transducer to acquire signals (induced currents) from the rotor while also revealing information about the condition of the stator. A current sensor (clamp-on probe and current transformer) with a resistive shunt at its output senses the motor current and records it in the time domain. Ideally, the motor current should be a pure sine wave. However, the motor current contains numerous harmonics. The motor current signal is further modulated by the different electrical and mechanical faults present in the motor, resulting in extra sideband harmonics. Faults in motor components cause abnormalities in the magnetic field, affecting the mutual and self-inductance of the motor. In the motor supply current spectrum, they appear as sidebands around the line (power, grid) frequencies. Motor faults can be diagnosed, and their severity can be validated based on fault characteristics. As the frequency range of interest for MCSA is generally 0–5 kHz, Nyquist's theorem states that a sampling rate of at least 10 000 times per second is required.

MCSA is the best choice for motors under steady-state conditions and rated loads since the current changes are influenced by both the fault and the power supply. The impact of eccentricity and the bearing faults can be reduced using a rapid current controlled inverter. Three effects are present in the faults: 1) the pulse repetition frequency, which is determined by the rotation frequency; 2) the vibration caused by the pulse; and 3) a rise in the total noise level. The most accurate method is to use sensor signals to detect the existence of these faults. As the sensor is close to the fault location, the relationship between the fault and the sensed variable becomes clear. However, in addition to MCSA, many different fault diagnosis methods have been proposed.

One of the essential condition monitoring approaches is the oil and lubrication analysis. Online (particle counting, temperature, viscosity monitoring, and so on) and off-line (oil filter analysis for flow and cleanliness characteristics) methods are both utilized to test and analyze lubricant samples. Lubrication analysis aims to retain oil quality and ensure that the components involved operate in the best possible environment. Lubrication analysis is generally done offline, with samples being examined and tested. Furthermore, when oil filters become overly dirty due to component wear, they might be a reliable sign of faults. Acoustic emission analysis is carried out using sensors that capture the sound generated by the machine using a sound level meter. The pressure levels and vibrations are converted into voltage signals using devices with antialiasing and high sampling rates. The types of data collected by acoustic emission analysis and vibration analysis are the same. The acoustic signal obtained is oscillatory. The acoustic signal characteristics vary depending on the faults of the rotating machinery. The noisy background might introduce additional components and impair the accuracy of fault identification of the monitored component, which is the major drawback of the AE condition monitoring method. The most widely used approach for condition monitoring is vibration analysis. Any change or malfunction in any of the mechanical components will cause the vibration profile to alter. Monitoring the vibration frequency can reveal whether a component is faulty or not. The disadvantages of vibration analysis are measurement errors due to improper sensor mounting, crystal overheating, and expensive proximity probes. Thermal field issues in motors are of great concern, and accurate temperature calculations are critical to the design and operation of motors. In addition, accurate thermal modeling of motors is essential for condition monitoring of the motors. Motor losses can result in high temperatures, which can cause severe thermal stress. There is thermal stress in the rotor end ring and bar, which is most likely to cause motor faults. Therefore, analyzing the effect of faults on the performance characteristics of the motor through the temperature rise of the motor can provide guidance to prevent accidents in operation and maintain safety. However, thermal analysis relies heavily on the accuracy of the sensors and measurements. The analysis of air gap or stray flux measurements can directly indicate the asymmetry of the radial or axial flux of the motor generated by fault-induced anomalies. Researchers have developed methods based on the analysis of external magnetic fields. Noninvasive examination and simplicity implementation are two of the main advantages. The drawbacks of these approaches stem from the complexity of simulating the magnetic field, which is strongly dependent on the electromagnetic behavior of the stator yoke and the motor housing, both of which have significant shielding effects. One of the main reasons flux monitoring has not received as much attention as MCSA is because it fails to provide remote monitoring. However, it is a low-cost alternative method that can compensate for the reliability and variety of fault detection in the limitations of electrical, mechanical, and thermal monitoring. Initially, based on the detection of modes associated with the current Park's vector representation, Park's vector approaches have been effectively utilized for condition monitoring of electromechanical systems. The extended Park's vector method relies on spectral analysis of the ac level of the current Park's vector mode, while, by averaging the current Park's vector method, converter power switching faults are detected when the vector mode is not zero. The major risks of converter diagnostics are load dependency and susceptibility to transients. The phase currents are normalized by the modulus of the Parks' vector, and the absolute value of the phase derivative of the absolute Park's vector is utilized as the detection variable in the power converter self-diagnosis based on the Parks' vector technique. Multiple faults can be diagnosed with extra signal processing and variables.

IV. Feature Extraction

The basic task of feature extraction is to find out the most effective features for fault recognition from plenty of features, achieving the compression of the feature space dimension, i.e., to obtain a set of "fewer but more precise" classification performance with a low probability of classification error. Motor current monitoring is an effective fault detecting technique since the stator current waveform data obtained from a defective motor differ from that collected from a standard motor. Most research on this technique uses different algorithms to decompose and interpret stator current waveforms, including Fourier analysis, linear discriminant analysis (LDA), wavelets, neural networks, and other predictive analysis approaches [24]–[27]. Time and frequency information is used in time–frequency domain strategies to capture transient characteristics [28]. Envelope analysis is a common technique for detecting and diagnosing bearing faults, which is traditionally determined based on analytically determined peaks. The FFT method is effectively implemented in the spectral analysis of the envelope signal [29]. However, the Fourier spectrum of vibration signals generally excludes the descriptive time-varying patterns of the acquired signals. Therefore, FFT fails to meet fault diagnosis requirements in real-time applications [30]. In addition, most methods analyze vibration signals in the time or frequency domain. Vibration monitoring is considered a reliable approach to assess the overall health of a rotor system. The frequency-domain analysis is attractive due to the more detailed information provided about the machine status [31]. However, fault detection is performed by comparing the indirect measurements of external forces based on the dynamic behavior of the machine. The difficulty in fault detection lies in classifying many frequency lines present in the vibration spectrum to extract useful information related to the health status of the motor. Several studies have used dynamic signal analyzers to measure the variance of the spectrum over time to solve this issue [32]. Betta *et al.* [33] developed an intelligent FFT analyzer that selects multiple frequencies in the spectrum as features and generates a reference model under healthy conditions, which is then compared to the monitoring characteristics for fault detection.

Furthermore, the HT is presented as a method for extracting and estimating the envelope of the vibration signal to get the local energy at each instantaneous frequency. As a result, this technique may be used to characterize the energy-frequency distribution of the vibration signal. It is beneficial to extract the characteristics of nonlinear signals [34], [35]. Due to the adaptive and unexpected nature of vibration signals, conventional methods must be based on reliable motor models and cannot be used effectively for vibration signal diagnosis.

The STFT is a popular signal feature extraction approach for converting quasi-steady-state vibration data into a continuous spectrum for neural network model training [36]. Faults can be detected based on the change in the expected value of the vibration spectrum modeling error. However, many potentially unstable frequencies must be tracked since there is no precise method exists for predicting which form of failure will occur. Moreover, vibration spectra often contain puzzling combinations of unusual frequencies that provide little descriptive information about the operating conditions of the motor. Therefore, modern procedures for tracking motor conditions are not reliable or efficient enough. The recent performance of neural networks in dynamically modeled complex systems holds the promise of reducing these issues and achieving better fault detection performance [37]. Neural networks can describe any nonlinear model without knowing the exact form, returning fast results during the recall process. Combined with STFT, an analysis method using vibration spectrum neural network modeling was developed in [38] to extract fault spectrum features for detecting machine faults. For nonstationary processing signals, the WT is a useful approach [39], which has excellent time–frequency characteristics in the local area. The internal generation between the studied signal and the intended wavelet basis will provide detailed information in both time and frequency domains. Due to its multiresolution inspection capabilities, this approach has shown great ability in fault diagnosis of mechanical machinery [40].

The discrete WT (DWT) was presented to extract fault features [41]. The advantage was to reduce the computation time. DWT has been widely applied in motor mechanical fault detection [42], [43] since it can only decompose low-frequency subbands. Wavelet packet transform (WPT) was proposed to decompose the high-frequency band and the low-frequency band in parallel to improve its frequency resolution. Although these methods have achieved great performance, they were still limited by the segmentation scheme. WPT cannot split the frequency of the signal, which may severely damage the transient vibration characteristics [44].

To overcome the dichotomous subdivision scheme and achieve adaptive representation, the empirical modal decomposition (EMD) approach was suggested [45]. The multimodulated vibration signal was decomposed into multiple intrinsic mode functions (IMFs) that were regarded as dominant mode components. It can extract both stationary and nonstationary components of a signal with efficiency. Therefore, it has attracted a lot of attention in signal processing and practical industrial applications. However, the lack of a mathematical theory for EMD methods has led researchers to propose improved methods regarding combinations of other methods in recent years [46]. Garcillanosa *et al.* [47] proposed an empirical WT (EWT), which combined the advantages of WT and EMD methods. The EWT method can perform the identification of weak faults and compound faults. Meanwhile, the EWT method can efficiently analyze the signal and extract internal features. In addition, EMD was an adaptive signal processing method that can be perfectly applied to nonlinear and nonsmooth processes. The main drawback was the pattern mixing problem. In addition, a new method called variational mode decomposition (VMD) was proposed in [48], which assumed that each extracted pattern has a finite bandwidth. Compression is performed around the central frequency of the match. However, VMD cannot support the model in practical applications, and its modulation capability depends heavily on the intrinsic parameter settings [49].

Intrinsic time scale decomposition (ITD) is a new adaptive time–frequency analysis approach that allows a nonsmooth signal to be decomposed into the sum of several intrinsic rotational components. Although it has a better performance over EMD, the approach cannot account for the physical meaning of the intrinsic rotational components. Cheng *et al.* [50] developed a local characteristic scale decomposition (LCD) algorithm based on the ITD method for the physical significance and its intrinsic rotational components. It can decompose any complex signal adaptively into instantaneous physical frequencies [51], [52]. Compared with EMD, the LCD method has better time–frequency localization characteristics and can extract the localization information of the original signal more effectively to obtain the fundamental characteristics of the signal. Since raw feature vectors will reduce the efficiency of the classification system, and incorrect generalization will eliminate useful details, linear methods are usually applied to reduce the number of features.

Both principal component analysis (PCA) and LDA can be utilized for 2-D downscaling, achieving clear visualization and detection of different or unknown fault modes [53]. Many researchers have investigated PCA approaches, considering their drawbacks in working with massive datasets when it seeks the global structure of data [54]. The function vector is made up of D computed features and has been expressed in D-dimensional space. Most of the data in the D-space has a nonlinear structure. In recent years, several learning methods [55], [56] have been carried out to retain the information in a lower d-dimensional space, where $D > d$, to solve this problem. Other methods, such as curvilinear component analysis (CCA), "find" the right form of the subtow cable and vice versa automatically. CCA has been applied in the field of fault diagnosis of motor drive systems. A CCA-based neural network FDD was presented in [57], which started with the selection of the most important features from an initial set formed by computing statistical time features from vibration signals.

The novelty of the method lies in its ability to perform complete fault analysis and diagnosis of various faults in bearings, both local and generalized faults. It is applied to the fault classification structure by performing the feature approximation phase of CCA. The classification task can be accomplished by applying a neural network.

Symmetric component (SC) is a promising method in time-domain FDD analysis and has received much attention, which can provide various information about the motor voltage or current balance, magnitude, and sequence. However, the generality of SC is limited to different fault classes, which may require detailed machine models and may be limited by computer resources. The SC method presented in [58] was based on stator currents and accomplishes signal feature extraction in multiple data processing steps. It is a low-cost computational method without complex computer models. Furthermore, particle filtering is another efficient tool for sequential signal processing that uses point masses and related discrete probability masses to estimate the state probability density function. It is based on the principle of sequential value sampling and Bayesian theory. Particle filtering has recently been extended to the prediction of machines [59]. Since mechanical failures are nonlinear dynamic problems, particle filtration is particularly effective in solving these problems. In addition, some mathematical models have been developed to explain the fault propagation process in most applications. However, the derivation of these models is complex and requires expertise in the degradation process. The typical fault propagation model is the first-order hidden Markov model (HMM), which is characterized by the fact that the current state of the system, which is determined by the previous state. Moreover, due to the increase in feature nodes, the trained system may contain some redundant nodes, which will lead to poor accuracy. Singular value decomposition (SVD) has been proposed to simplify the system. In the SVD process, the signal is converted into a matrix where the singular values represent the nature of the faulty signal [60], [61]. In addition, to fuse the information from all the utilized classifiers, Hoang and Kang [62] proposed a fault diagnosis method based on a motor current signal developing deep learning and information fusion (IF). The novel decision-level IF technique applied the raw signals of multiphase motor currents as direct inputs and extracted features from the motor current signals of each phase to improve classification accuracy. Table I summarizes all mentioned signal process methods for feature extraction.

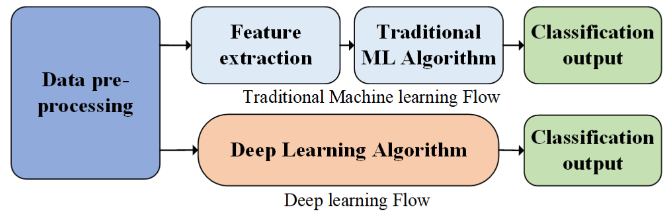


Fig. 5. Flowchart of traditional ML and deep learning.

V. Fault Diagnosis

As the crucial step in FDD, accurate and efficient classification performance is extremely important. As an attractive method, ML includes a variety of model types:

1) supervised learning, which is responsible for analyzing training data and generating inference functions through algorithms;
2) unsupervised learning to draw conclusions from unlabeled data.

The most common unsupervised learning method is cluster analysis, which is mainly used to discover hidden patterns in grouped data. Deep learning is a subdomain of ML, and its related algorithms are inspired by the structure and function of the brain (i.e., ANNs). Deep learning uses a hierarchical framework of nonnearline transformations of input information to construct statistical models and output outcomes. Traditional ML is capable of adapting to a wide range of data quantities, particularly in circumstances with small data volumes. On the other hand, if the amount of data increases rapidly, then the effect of deep learning will be more prominent. Fig. 5 illustrates the flowchart of traditional ML and DL. The application of classic ML and popular deep learning algorithms in motor FDD is introduced in this section.

A. Support Vector Machine (SVM)-Based Approach

An SVM is a binary classification model that attempts to build a line between two classes of points by mapping the feature vector of an instance to certain points in space. SVM is applied in regression analysis for data classification and system parameter estimation.

SVM has various applications, such as handwriting recognition and image recognition. SVM is particularly useful for small samples and database instances. Because of its appealing characteristics and strong analytical results, SVM is becoming increasingly popular in the field of motor drive system

TABLE I
HIGHLIGHT OF FEATURE EXTRACTION METHOD

Reference	Theme	Principle	Highlight
[28] [29] [30] [33]	FFT	• Spectrum analysis of envelope signals • Frequency distribution	• Time-frequency domain conversion • Poor performance for non-stationary signal • No time-frequency analysis
[32] [34] [35]	HHT	• Multi-resolution of frequency scale	• To avoid complex mathematical operations • To analyze signals whose frequency varies with time • High calculation burden
[36] [38]	STFT	• Long and short time window movement	• Time-frequency characteristics are acquired • High-frequency resolution and high time resolution cannot be acquired simultaneously
[39]	WT	• Signal Decomposition and reconstruction	• Time-frequency analysis
[41] [42]	DWT	• Discrete input and output	• To reduce calculation time
[43]	WPT	• Fitting the mutation signal	• Effective splitting of fault-induced resonances and high-frequency band features
[45] [46]	EMD	• Irregular frequency waveforming	• Adaptive data processing • Suitable for non-linear, non-stationary time series
[48] [49]	VMD	• Sub-signal decomposition at different frequencies	• To avoid pattern mixing • Modulation capability dependent on parameter settings
[47]	EWT	• Adaptive wavelet subdivision scheme	• More consistent decomposition
[44] [50] [51]	LCD	• Adaptive decomposition of a complex signal	• Better time-frequency localization characteristics • More efficient extraction of localization information from the original signal • Faster calculation speed
[53] [54]	PCA	• Multidimensional feature mapping	• Signal dimensionality reduction • To reduce model training time • Better accuracy
[57]	CCA	• Information Retention • Self-organizing mapping	• Linear feature extraction • Distance preservation
[58]	SC	• Symmetric component acquisition	• Rapid and accurate computation performance
[60] [61]	SVD	• Matrix decomposition	• To use low-rank approximations to simplify systems

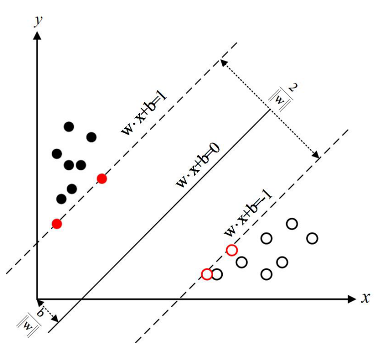


Fig. 6. Optimal hyperplane for binary classification by SVM.

fault diagnosis. Fig. 6 illustrates the optimal hyperplane for binary classification by SVM

$$\min\Phi(w) = \frac{1}{2}\|w\|^2 \tag{1}$$

$$\text{s.t., } y_i(w.x_i + b) \geq 1; \quad i = 1, 2, \ldots, n \tag{2}$$

where w stands for the weight vector, which describes the hyperplane, x_i stands for the input vector, y_i stands for the label associated with y_i, and b stands for a scalar threshold. For the case of nonlinear separation, in high-dimensional spaces, the SVM can construct a hyperplane that allows for linear separation. Only the transformation $\Phi(x)$ from N-dimensional space to Q-dimensional function space of the data is used in SVMs. Since the inner product can be replaced, the mapping translation can be achieved with a kernel operation. By retaining the influence of high-dimensional transformations, it reduces the computational load. This is the kernel function: $k(x, x_i) = \Phi(x).\Phi(x_i)$. Many functions can be used as the kernel function. The Gaussian radial basis function (RBF) is a well-known kernel, and the equation is

$$k(x, x_i) = \exp\left(-\gamma \|x - x_i\|^2\right); \quad \gamma = \frac{1}{2\sigma_b^2} > 0. \tag{3}$$

After optimization, the basic SVM structure can be described as

$$f(x) = \sum_{i=1}^{m} \{y_i \alpha_i k(x, x_i)\} + b. \tag{4}$$

SVM has a nonlinear multilabel classification function. For multiple fault diagnosis, an intelligent fault detection system based on a multivariate integrated incremental SVM (MEISVM) is presented in [63] for multifault diagnosis. This method can detect a variety of faults, including complex compound faults and faults with different severity thresholds. Hu *et al.* [64] proposed a grid search SVM (GSSVM) approach based on redefined dimensionless indicators (RDIs) extracted maximum correlation and minimum redundancy feature selection. Minimum redundancy maximum relevance (mRMR) was used to automatically eliminate redundant and irrelevant features from high-dimensional features to obtain more fault features that indicated the useful information hidden in the vibration signal. The proposed RDI as a new fault feature can effectively solve the shortcomings of the traditional dimensionless index and has a stronger discriminative ability for mechanical faults. Automatically eliminate redundant and irrelevant features in the high-dimensional feature space. The correlation between features and output class labels is maximized, and the redundancy between features is minimized. However, the dimensionless metric is sensitive to faults rather than to operating conditions, whether the classification effect is significant when the operating conditions of the motor change are not given as a validation. A novel intelligent fault detection approach for rolling bearings was proposed based on composite multiscale fuzzy entropy (CMFE) and ensemble SVMs (ESVMs) to extract the nonlinear features that were embedded in the vibration signal [65]. CMFE was utilized to extract hidden nonlinear fault characteristics from rolling bearing vibration signals, and then, ESVM was used to construct a multifault classifier to accomplish an automated intelligent diagnosis of rolling bearings. In [66], several investigations were carried out to record current conditions during various motor power supply activities, such as internal stator WSC failures and supply voltage imbalance at different load speeds. A recursive feature elimination algorithm based on the SVM (SVM-RFE) was applied to select and maximize the number of appropriate features to be used for classification. It was worth noting that two different feature sets were created. One of them contained load-level details, which were intentionally hidden in the other one. The selected features were used to identify various stator winding fault conditions using a support vector regression (SVR)-based classifier. It has been found that the performance of the classifier was better when the load level information was included in the functional layer rather than the load level information was hidden. To make the fault classification algorithm uninterrupted at different load levels, two additional functions were extracted from Park's vector modulus using detrended fluctuation analysis (DFA). Ultimately, the SVR-based classifier was found to be accurate in detecting and classifying internal motor faults. A satisfactory response was shown in distinguishing between supply voltage imbalance and internal stator faults. The full-spectrum cascade analysis of rotating machine vibrations is an effective method. Abrupt changes in acceleration signals of rotor faults can be detected, and the characteristic spectrum of faults is displayed in a full-spectrum cascade diagram. In addition, the fault diagnosis problem becomes more confusing when the current characteristics due to power supply voltage imbalance are significantly similar to those due to internal stator winding faults. It is important to select, rank, and optimize the number of valid features used for classification. RFE uses some mathematical or heuristic rules to train the SVM classifier with features that minimize the margins. This feature elimination process is repeated until some stopping criterion is satisfied. Removing features during the iterative process simplifies the computation, but this may lead to suboptimal solutions. The full spectrum experimental data were used for SVM training, and excellent classification results were obtained [67].

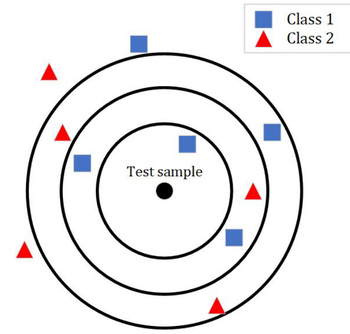


Fig. 7. Structure of KNN.

B. MLP-/KNN-/RF-Based Approach

Multilayer perceptron (MLP) is a simple-structured neural network. The most common MLP includes three layers: the input layer, the hidden layer, and the output layer, with all three layers of the MLP neural network being completely connected, which is widely used in the detection and diagnosis of the motor drive system. Because of its capacity to indirectly detect dynamic nonlinear associations between dependent and independent variables, as well as its ability to detect all possible correlations between predictor variables and multiple training algorithms, the MLP feedforward neural network has been adopted [62]. A three-phase asynchronous motor control method based on RBF-MLP cascaded neural network was proposed in [68]. To obtain rich fault knowledge from stator currents, simple statistical features, such as standard deviation, kurtosis, energy, entropy, and variance, were extracted. PCA was used to pick the most superior functionality to remove obsolete or irrelevant details and reduce the burden of the classification scheme. The classifier is sufficiently robust, i.e., the classification accuracy does not change in the presence of uniform and Gaussian noise in the input and output. The advantage is that good classification performance can be achieved without the requirement for a large amount of accurate measurement data. Furthermore, MLP was utilized to classify interturn short-circuit faults in PMSM stator windings at different speeds, loads, and fault severities. These states led to the complexity of the fault features. Depending on the complexity, the nonlinear relationships between the relevant features are identified to detect different severity levels of interturn short circuits.

Due to its easy implementation and substantial classification efficiency [69], the k-nearest neighbor (KNN) method is a popular classification method in data mining and statistics. The principle structure of KNN is shown in Fig. 7. Recioui *et al.* [70] presented a system for detecting bearing faults in electric motors and monitoring the bearing loss. The approach used spectral kurtosis (SK) and reciprocal association to extract fault features, which were then combined with

TABLE II
HIGHLIGHT OF CLASSICAL ML METHOD

Reference	Theme	Principle	Highlight
[63]	MEISVM	• Sensor information correlation • Pairwise secondary optimization	• Multi-fault diagnosis • Highly rely on input feature data
[64]	mRMR GSSVM	• Maximize feature relevance • Minimize redundancy between features	• Fast calculation and high robustness • Random parameter setting
[65]	CMFE ESVM	• Obtain vector similarity	• Sensitive to nonlinear feature
[67]	FFT SVM	• Structural Risk Minimization	• Acceleration signal analysis application
[68]	RBF MLP	• Competition rules and metric combinations	• Low demand for accurate measurement data • High classification performance
[70]	SK KNN	• Category distance calculation	• To improve speed and space efficiency
[71] [72] [73]	RF	• Random sampling and replacement	• Effective classification ability • Insufficient generalization ability

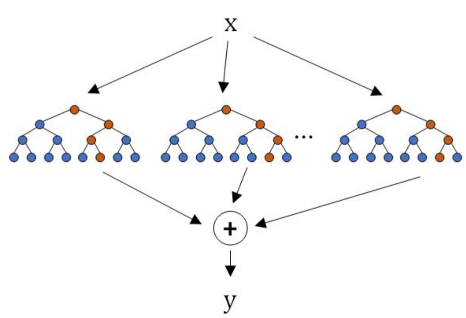


Fig. 8. Basic structure of the RF.

PCA and semisupervised KNN distances to provide health metrics.

Random forest (RF) models are based on the classification and regression tree (CART) groups [71]. RF is a method that uses integrated learning to combine numerous trees, with the fundamental unit becoming a decision tree, each of which is a classifier. The categorization results for an input sample will then be N trees. The RF integrates all classification votes and designates the most voted category as the final output, which is a basic bagging idea. Fig. 8 illustrates the basic structure of the RF. An RF classifier was proposed in [72] for the classification of bearing faults. The statistical features of the bearing vibration signals were computed and fed to the RF and ANN classifiers. Seera *et al.* [73] developed a hybrid model that combines a fuzzy min–max (FMM) neural network with CART for online motion detection and diagnosis tasks. In addition to strong online performance, the FMM-CART produced helpful decision trees to explain the collected functional information.

To sum up, the ML methods mentioned above are all shallow models with simple structures, high computational efficiency, and great classification performance. Table II summarizes the classical ML methods for fault classification.

C. Deep Learning

In contrast to traditional ML methods with manual feature labeling, a large amount of research has focused on how to extract representative features from the original signal. On the other hand, the extracted data may contain redundant or insensitive data. To identify sensitive characteristics, certain dimensionality reduction methods were utilized, which may have an impact on diagnostic findings and computing performance. Most of the studies on intelligent fault diagnosis have produced valuable results, but there are still two obvious drawbacks.

1) Manual extraction of features requires *a priori* knowledge, which requires a lot of practical work to determine and may have greater checking popularity.
2) Traditional ML techniques cannot effectively distinguish complex information in raw data, and the shallow structure of ANNs limits their ability to understand the complex nonlinear relationships hidden in the measurement data.

In terms of a large number of hidden neurons, DNNs can obtain nonlinear representations of data. It has achieved higher performance in the field of motor FDD.

1) Convolutional Neural Network: CNN is a feedforward neural network that mainly stimulates the activity of the visual system of the human brain. Local receptive fields, weight sharing, and spatial domain secondary sampling are the three main architectural principles in the structure of CNN. Therefore, CNN is well suited to processing 2-D data, such as images. Convolutional layers (CLs), pooling layers (PLs), completely connected layers (FL), and SoftMax layers are the four kinds of layers that make up CNN. The following equation can be used to mathematically model the operation of processing input data in the CL:

$$x_j^i = f\left(\sum_{i \in M_j} x_i^{l-1} * k_j^l + b_j^l\right) \tag{5}$$

where x_i^{l-1} indicates the input data of layer l and $*$ denotes the convolution operation. The layer is made up of n kernels, each with its weight matrix and bias vector. Because of its versatility, the rectified linear unit (ReLU) is often used as a CL activation function. The equation for ReLU is

$$f(y) = \max(0, y). \tag{6}$$

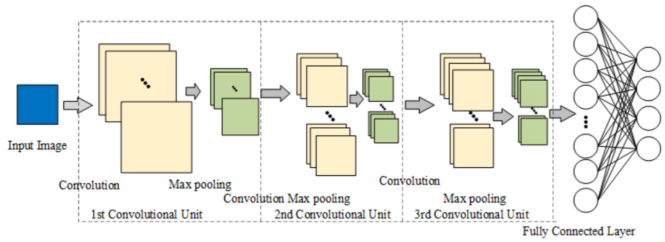


Fig. 9. Basic structure diagram of CNN.

The following equation can be used to express PL mathematical operation:

$$x_j^{l+1} = s\left(x_j^l\right). \tag{7}$$

After the nonlinear combination of features extracted by FL, the SoftMax layer acts as a classifier. The SoftMax function converts an N-dimensional real vector into a set of real vectors (0, 1). The SoftMax function has the following equation:

$$p_i = \frac{\exp(z_i)}{\sum_{j=1}^{N} \exp\left(z_j\right)}. \tag{8}$$

Signal processing techniques with different characteristics are combined with CNN for fault signal diagnosis [74], [75]. Ince *et al.* [74] proposed a novel motion state monitoring framework with an adaptive implementation of 1-D CNN, where the two main modules of traditional fault detection steps included feature extraction and classification were fused into one. The proposed approach extracts the best features through proper training. Therefore, manual parameter adjustment and manual feature extraction are not required, achieving more efficient fault diagnosis capability. Chen *et al.* [76] proposed an integrated strategy based on data-driven and deep learning to deal with initial faults. The average moving technique was introduced into the typical correlation analysis (CCA) framework, making the new residual signal more sensitive to initial faults. Moreover, new test statistics that work closely with Kullback–Leibler divergence (KLD) were proposed from a probabilistic perspective and greatly improved fault detection performance. The fault matrix was defined and used as the input of the CNN, whose feature extraction capability was greatly improved compared with the conventional method, which helped to diagnose the initial faults accurately. The CNN shown in Fig. 9 was used as a construction block multisignal framework. Randomly initializing the designed model, the deep model was trained using the training dataset, minimizing the error between model output predictions and actual labels by iteratively updating the parameters of the DCNN model [77]. The proposed method has one input and output layer, three hidden convolutional units, each of which was followed by a max-pooling operation, and one fully connected hidden layer following the convolutional and max-PLs. ReLu was implemented as an activation function, and CWT was used to convert the time-domain signal into gray-scale images as input. The model was trained and verified with experimental data, including current and vibration signals. The training time was 201 and 156 s, and the accuracy rates of 98.72% and 98.26% were obtained, respectively. In addition, CNN combined with other networks has been proposed. A convolutional RNN (CRNN) was used to diagnose multiple faults of high-speed train (HST) bogies by combining CNN and RNN [78]. The model inherited the functions of both CNN and RNN. A new approach was proposed, which combines CNN and extreme learning machine (ELM). CNN demonstrated strong automated feature extraction capabilities, while ELM was proposed as a quick and efficient classification algorithm [79]. To improve the feature learning capabilities, a CNN with a square pool configuration was built and used as an automated feature extractor in the first level. ELM was further used in the second stage to increase classification accuracy and learning speed. In [80] and [81], the initial vibration signal was input to a deep CNN named deep CNNs with wide first-layer kernels (WDCNN). A large first-layer convolutional kernel and a deep network structure with narrow CLs were the two key features of the WDCNN. The proposed model enhances the accuracy of the current CNN fault diagnosis. Besides, a new model named convolution neural networks with training interference (TICNN) was proposed to solve the fault diagnosis problem in [82]. Without the requirements for a time-consuming manual feature extraction process, TICNN processed the raw vibration signal directly. Without any time-consuming denoising preprocessing, the raw time signal was used as data. Meanwhile, the model was independent of any domain adaptation algorithm or target domain information. An online fault diagnosis scheme was proposed in [83]. Conventional ANNs for fault diagnosis classification tasks were trained offline with historical data, and then, the trained model was used for online fault diagnosis. Due to the limitations imposed by the time consumption associated with the training process of the model, an online fault diagnosis algorithm including two phases was proposed. In the first stage, an SVM was used to separate the healthy data from the faulty data, and in the second stage, a CNN was trained to learn the features used to isolate the faults. Although the proposed method was an end-to-end self-supervised learning model, the performance of the CNN in the second stage was limited by the number of datasets and the training time consumption. The data were used only in the healthy state through the fault diagnosis function in self-supervised learning. The operating condition or operating condition class of the powertrain was defined in the first stage by using a class of SVMs. The generated health classes were used to train a CNN-based classifier. This approach outperformed current algorithms and approaches that use domain feature extraction. A new LeNet-5-based CNN for fault diagnosis was proposed in [84]; three datasets were used to verify CNN models based on LeNet-5 with eight layers, which comprised of three transformed pictures of 64×64, 64×64, and 16×16. On three separate datasets, the presented CNN had prediction accuracies of 99.79%, 99.48%, and 100%. By converting the signal into a 2-D image, the method extracted the features of the converted 2-D image and eliminates the influence of artificial features. The novel data preprocessing approach presented in the literature transformed raw time-domain signal data into 2-D gray-scale pictures without the use of any predetermined parameters, removing as much as possible the knowledge

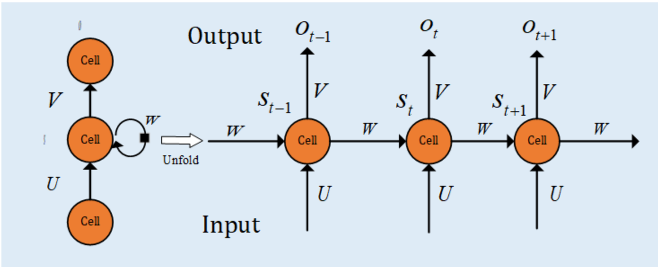


Fig. 10. Basic structure diagram of RNN.

of the expert. However, the limitations of the method in terms of practical applicability included the following aspects. First, the most prevalent fault circumstances must be identified, which may be achieved using a dictionary list type. Otherwise, unknown faults will be categorized incorrectly as known faults. Second, the training procedure took a long time. Transfer learning (TL) can be used to minimize training time in the future based on these constraints.

2) Recurrent Neural Network: RNNs are presented for sequence learning. RNN establishes the connection between units as a directed loop, unlike basic neural networks, where multilayer perceptron can only map from input data to target vectors. Fig. 10 indicates the structure of RNN. Before the time series are input into the RNN, the data are converted into 2-D data according to the size of the time window to maintain their sequential nature. W is a common weight, and W and V are the weight matrices of the input and output layers. In addition, when calculating the state S_t at the current time t, the input value x_i is multiplied by the weight U of the input layer, and then, the result is obtained by multiplying the common weight W, which is summed to be constant. The equation is shown as follows:

$$S_t = f(U * x_t + W * S_{t-1}) \tag{9}$$

$$O_t = f(s_{t-1} * V). \tag{10}$$

Due to the RNN map of the target vector from the entire history of previous inputs, the memory of prior inputs is retained in the internal state of the network. RNN can be trained by backpropagation to accomplish the task of timely supervision. However, the issue of gradient disappearance during the backpropagation of model training hinders the performance of RNN. It means that traditional RNN may not be able to capture long-term dependencies. Therefore, the long short-term memory (LSTM) algorithm is designed to prevent backpropagation errors from disappearing or exploding. To solve the problem of long-term reliance, forgetting gates are incorporated into LSTM. The use of cell state information may be controlled using these applicable forgetting gates. Because it can capture long-term dependencies, LSTM has the advantage over standard RNN for capturing nonlinear dynamics in time series sensory input and learning an efficient representation of machine states. Considering the ability of LSTM to capture remote dependencies and nonlinear dynamics in time series data, LSTM has been effectively applied to fault diagnosis of motor drive systems [80], [85], [86]. Shenfield and Howarth [80] presented the RNN-WDCNN, a new dual-path RNN with a larger initial kernel and deeper CNN routes that can operate on raw temporal information. The WDCNN combined the functions of RNNs with CNNs to capture remote relationships in time series data and remove input signals in high-frequency noise. Besides, Jalayer *et al.* [85] developed a novel approach named convolutional LSTM (CLSTM), which is developed to handle signal-based FDD in rotating machinery. The developed model improved the efficiency of processing multichannel input data and learning its spatiotemporal characteristics. The input channels were subjected to sensitivity analysis, and the results suggest that combining these multidomain characteristics improves the accuracy of the classifier. However, feature engineering on severely unbalanced datasets while finding the optimal hyperparameters to train the classifier faster without compromising its accuracy is a serious topic. The CLSTM architecture was used to process multichannel input data and understand its spatiotemporal characteristics more efficiently. The updated CLSTM supported the FDD approach to properly understand the function of data structures and achieve higher accuracy. Zhao *et al.* [86] proposed a deep neural network architecture for processing raw sensory data, called convolutional bidirectional long-term short-term storage network (CBLSTM). The proposed method consisted of one layer CNN for feature extraction, and the size of the input signal is 100×12. Then, two hidden layer bidirectional LSTMs were implemented to encode the temporal patterns. Two fully connected layers of size 500×600 were used before feeding the representations to the linear regression layer. The activation function was ReLu. One epoch takes 5 s to train, and each sample takes 0.027 s to test. Compared with RNN, deep RNN, LSTM, and so on, the proposed algorithm has a better performance. CBLSTM first used CNN to extract stable and information-rich local features from sequential inputs. Then, the temporal information was encoded using a bidirectional LSTM. The long short-term storage network (LSTM) models linear data and record long-term dependencies, while the bidirectional system captures past and future environments. A stacked fully connected layer and a linear regression layer were constructed on top of the bidirectional LSTM to predict the target values. In addition, two RNN networks were applied as encoder and decoder in place of the general fully connected layer to effectively reduce the dimensionality of the time series data combined with the AE of the RNN [87]. Moreover, Huang *et al.* [53] were used in combination with a variant AE (VAE); variance and noise are added to make the model generation more realistic. The time-domain vibration signals at three different locations were used as inputs, and two RNN networks were developed as encoder and decoder, respectively, instead of the usual fully connected layer. In addition, the variance and noise were added to make the model generation more realistic, in addition to reconstructing the input data. This method reduced the computational cost. However, when the length of the input sequence was too long, it will cause the information to be diluted during the propagation process, thus decreasing the accuracy.

3) Generative Adversarial Network: A GAN is divided into parts: a generator and a discriminator. GAN training aims to

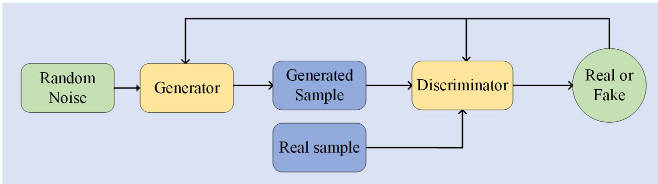


Fig. 11. Basic structure diagram of GAN.

achieve a Nash balance between the generator and discriminator based on the binomial zero-sum game principle. The generator and discriminator in the GAN are differential functions that can be optimized by using any gradient-based approach. The purpose of the generator is to capture the potential distribution hidden in the real sample by using z as input. Therefore, the generator outputs a "fake" sample x_f that is as similar as possible to x_r, thus confusing the discriminator. In addition, the purpose of the discriminator is to distinguish x_f from x_r and identify the actual sample from the generated sample. Due to the conflicting goals of the generator and discriminator, the two parts of the GAN compete and gradually become more powerful during the training process. Once the GAN is trained correctly, the distribution of the generated samples will match the distribution of the actual samples, which makes it difficult for the discriminator to distinguish the difference between x_f and x_r.

Fig. 11 indicates the basic principle of GAN. The nonlinear functions of the discriminator and generator approximation are represented as $G(\cdot)$ and $D(\cdot)$, respectively. The distributions of random noise and actual samples are represented as P_{data} and P_n, respectively. Due to the different training objectives of the discriminator and generator, the objective function is defined, respectively, as

$$\min_G \left\{ L_G(D, G) = E_{Z\sim P_n}\left[\log(1 - D(G(z)))\right]\right\} \quad (11)$$

$$\max_D \left\{ L_D(D, G) = E_{x\sim P_{\text{data}}}\left[\log D(x)\right] + E_{Z\sim P_n}\left[\log(1 - D(G(z)))\right]\right\} \quad (12)$$

where $L_D(D, G)$ and $L_G(D, G)$ represent the objective functions of the discriminator and the generator. Equations (11) and (12) can be combined into a single objective function for the overall training operation of GAN, as follows:

$$\min_G \max_D \left\{ L_D(D, G) = E_{x\sim P_{\text{data}}}\left[\log D(x)\right] + E_{Z\sim P_n}\left[\log(1 - D(G(z)))\right]\right\}. \quad (13)$$

Due to their ability to learn deep representations without deep marking of training results, GANs have received a lot of attention in a wide variety of fields [88], [89]. GAN has also been used in mechanical defect diagnosis in recent years [90], [91]. The data imbalance between different machine health conditions can be addressed [92], [93], and unsupervised fault classification can be achieved by incorporating GAN into certain deep learning algorithms [91]. In general, ML- and DL-based FDDs share a universal drawback, which is the models require a substantial quantity of training data to learn the inherent patterns of normal and faulty data in a customized manner. However, for various reasons, it is hard to acquire actual operational data corresponding to different health conditions: 1) low frequency of failures; 2) inserting faults into rotating equipment is expensive and dangerous; and 3) labeling data is highly time-consuming even if a large amount of data is available. Therefore, GAN is a method that requires only a small amount of sample data to obtain highly accurate training results. GAN can generate sample ML algorithms in an unsupervised manner. Due to the adversarial generator and discriminator, the GAN will learn samples adaptively for training. When both the generator and the discriminator have completed training, the output of the generator will produce samples close to the actual samples even if the input data are random noise. On a variety of image datasets, GAN variants, such as dc-GA and ac-GAN, have demonstrated strong generative efficiency. Lee *et al.* [94] applied the GAN method to the FDD of a motor instead of other traditional resampling methods. To increase the imbalance data of the motor, the GAN-based method generated more realistic samples and improved the accuracy of fault diagnosis compared to use other resampling methods. Due to a complete lack of anomalous samples, single GAN-based fault diagnosis models do not recognize the case of new types of faults. Using stacked noise reduction AEs and GANs, Wang *et al.* [95] proposed a fault diagnosis process (SDAE-GAN) that enabled the discriminator to determine the sample fault type and whether the input data are from a real data distribution. However, learning various data distributions simultaneously is a challenge. Furthermore, the proposed method can eliminate the interaction between the intrinsic distributions of different fault patterns during the training process. Therefore, it has stronger generalization capability compared with a single fault diagnosis method based on GAN. This method is of great significance for solving the problem of difficulty in obtaining fault data in practice. Although FFT requires less expertise than other manual feature extraction methods, it still requires the use of signal processing, and the proposed method cannot be considered as an end-to-end fault diagnosis method. Wen *et al.* [96] studied multiple GANs to understand the data distribution for each health condition and then developed a semisupervised approach to improve the feature extraction capability of each GAN with better generalization capability compared to a single GAN-based approach. The generator and discriminator have a symmetrical structure, which has one input and output layer and two hidden layers. The size of the input layer of the generator is determined by the size of the random noise vector, which is 128×128. ReLu is applied for the activation function of hidden layers, while sigmoid is used in the output layer. Besides, only 200 samples in each dataset are used for training the model. The performance of testing accuracy is greater than 95%, and the standard deviations are below 1.2%.

4) Deep Belief Network (DBN): In contrast to traditional neural network discriminative models, DBNs are probabilistic generative models in which the generative model creates a joint distribution between the observed data and the labels. Layer-by-layer training allows better initial weights to be assigned to the entire network so that the network can be fine-tuned to achieve the best solution. DBNs consist of

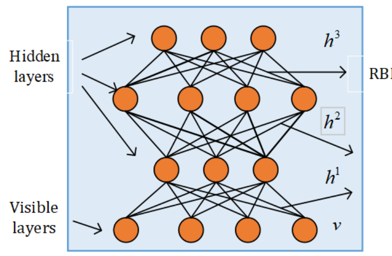


Fig. 12. Basic structure diagram of DBN.

multiple layers of constrained Boltzmann machines (RBMs) and a typical neural network, as shown in Fig. 12. These networks are "constrained" into visible and hidden layers, with connections between the layers, but there are no connections between the internal units. The hidden layer units are learned to recognize the relevance of the higher order data displayed in the visible layer. Through the energy function $E(v, h, \theta)$, the relation weights describe a probability distribution over the joint condition of the visible and hidden cells (v, h)

$$E(v,h,\theta) = -\sum_{j=1}^{m} b_j v_j - \sum_{i=1}^{n} c_i h_i - \sum_{i=1}^{n}\sum_{j=1}^{m} v_j w_{i,j} h_i. \quad (14)$$

The probability distribution of each possible visible and hidden vector pair can be defined by the following energy function similar to that of a general Boltzmann machine

$$p(v,h;\theta) = \frac{1}{Z(\theta)} \exp(-E(v,h;\theta)) \quad (15)$$

where $Z(\theta) = \sum_{v,h} \exp(-E(v,h))$.

The conditional probabilities of hidden and visible cells are given by

$$p(h_i = 1 \,|\, v;\theta) = \frac{1}{\left[1+\exp\left[-c_i - \sum_{j=1}^{m} v_j w_{i,j}\right]\right]} \quad (16)$$

$$p(h_i = 1 \,|\, h;\theta) = \frac{1}{\left[1+\exp\left[-b_i - \sum_{i=1}^{n} v_j w_{i,j}\right]\right]}. \quad (17)$$

The RBM model with binary cells can be learned by a negative log-likelihood gradient as follows:

$$\Delta w_{i,j} = (\Delta w_{i,j} = \eta \langle v_j h_i \rangle_{p(h|v)} - \langle v_j h_i \rangle_{\text{recon}}) \quad (18)$$

$$\Delta b_j = (\langle v_j \rangle_{p(h|v)} \langle v_j \rangle_{\text{recon}}) \quad (19)$$

$$\Delta c_i = (\langle h_i \rangle_{p(h|v)} - \langle h_i \rangle_{\text{recon}}) \quad (20)$$

where η is the learning rate and $\langle h\cdot \rangle_{p(h|v)}$ is the expectation value, which is relative to the conditional distribution $p(h|v)$ reconstruction of the distribution of the model. Superposition in the forward direction RBM learning may be used to learn weights in an unsupervised manner, which is referred to as supervised initialization of the learning parameters. It is the same as having *a priori* knowledge of the supervised learning input data.

Some DBN-based FDD methods are carried out in the motor drive system. A novel hierarchical diagnosis network (HDN) rolling carrying automatic diagnosis method, consisting of two layers of DBN, was proposed in [97]. WPT was used to provide representative features to deal with the nonsmoothness of the fault vibration signal. Determination of fault location and its magnitude level can be achieved by HDN, and weak points were identified in this way to avoid device performance loss and providing information for the reliability of the designed configuration. Shao *et al.* [98] proposed an improved convolutional DBN (CDBN) with compressive sensing (CS). The structure of the proposed algorithm was divided into one gaussian input layer, six convolutions hidden layers, six PLs connected with 12 convolutions hidden layers, and 12 PLs. The output was fed into a softmax classifier for fault classification. The compressed vibration signals were used as input data, and single and compound faults were considered for verification. The performance of testing accuracy was 94.80%, and the standard deviation was 0.53. To improve the analysis efficiency, CS was used to reduce the amount of vibration data. Then, a new CDBN model with Gaussian visible units was developed to enhance the functional learning capability of compressed data. Furthermore, the constructed deep model generalization efficiency was improved using the exponential moving average (EMA) technique. A discriminative DBN (ACO-DDBN) based on ant colony optimization was presented in [99]. The proposed method consists of one input layer, two hidden layers, and one output layer. The input signal is a vibration signal with 1024 data points. It takes 1592.5 s for training the model. After the training process, the training accuracy and the testing accuracy are 94.7% and 91.2%, respectively. The RBM greedy layer-by-layer learning algorithm can effectively pretrain unsupervised models regardless of the amount of available training data. However, the performance of DBNs can be affected by their parameters. Since ants find the best choice of parameters throughout the search process, ACO is suitable for selecting parameters. Therefore, the parameters of the model are obtained by using the ACO method. The composition of the DDBN model can be computed automatically without the optimization process and prior knowledge to improve efficiency.

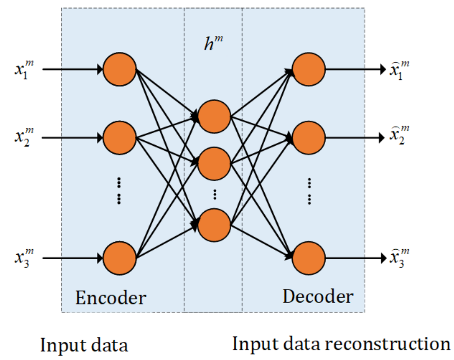


Fig. 13. Basic structure diagram of AE.

5) Autoencoder: The AE neural network is an unsupervised learning algorithm that makes the target value equal to the input value by using a backpropagation algorithm. Fig. 13 illustrates the basic structure diagram of AE. An AE is a neural network that consists of two components: an encoder function and a decoder that generates reconfigurations. Traditionally, AEs have been used for dimensionality reduction or functional learning. An improved prototype structure of the AE has resulted in a DAE that first complete the pretraining of the hidden layer using an unsupervised layer-by-layer greedy training algorithm and then uses BP to tune the entire neural network through systematic parameter optimization. This algorithm greatly reduces the performance metrics of the neural network and effectively improves the undesirable situation where the BP algorithm tends to fall into local minima. In short, compared with the original AE, DAE can increase depth, improve learning ability, and help pretraining. SAE only changes one thing: the input data are fed into the model with noise, and then, the original image without noise can be restored. Through this method, a better robust representation can be extracted.

The dimensionality reduction includes the encoder network. The function reduces the initial high-dimensional data to a low-dimensional representative. The decoder network is the opposite of the encoder network and is part of the restoration process. The function returns the studied low-dimensional features to the high-dimensional data. The equations are shown as follows:

$$h_i = f_{\theta_1}(W_1 x_i + b_1) \tag{21}$$

$$\hat{x}_i = g_{\theta_2}(W_2 h_i + b_2) \tag{22}$$

where W and b are the weights and biases of the network.

The purpose of the AE model is to learn useful hidden representations by minimizing the reconstruction error. Therefore, the parameter sets θ_1 and θ_2 can be optimized by minimizing the reconstruction error L_{rec} while giving n training samples

$$L_{\text{rec}} = \frac{1}{n}\sum_{i=1}^{n}\|x_i - \hat{x}_i\|^2 \tag{23}$$

$$\hat{\theta}_1, \hat{\theta}_2 = \arg\min L_{\text{rec}}(\theta_1, \theta_2) \tag{24}$$

where $\hat{\theta}_1$ and $\hat{\theta}_2$ denote the optimal values of θ_1 and θ_2, respectively. The stochastic gradient descent (SGD) algorithm can be used to solve this optimization problem. The equations are shown as follows:

$$\hat{\theta}_1 \leftarrow \theta_1 - \lambda\frac{\partial L_{\text{rec}}}{\partial \theta_1} \tag{25}$$

$$\hat{\theta}_2 \leftarrow \theta_2 - \lambda\frac{\partial L_{\text{rec}}}{\partial \theta_2} \tag{26}$$

where λ is the learning rate.

A monitoring method based on stacked denoising AEs was proposed in [100]. The stacked denoising AE-based approach can effectively extract robust features from corrupted data and shows well potential in the field of process monitoring. It is suitable for specific health state identification for signals containing environmental noise and fluctuations in operating conditions [101], [102]. A novel method, namely, a local connection network constructed by normalized sparse AE (NSAE-LCN), was proposed for intelligent fault diagnosis [103]. In this approach, the LCN first learned various meaningful features from the vibration signal using NSAE. Then, the LCN generated displacement-invariant features to classify the health condition of the machine based on the learned features. In [104] and [105], a new deep learning algorithm for bearing fault diagnosis was proposed. A combination of discriminative and structural information between different fault conditions in a deep AE model was applied. The method can study the association structure and structural relationship information between multiple fault states, which helped to improve the stability of the deep neural network. To effectively assess the health of a motor, various sensors were installed in different locations to obtain more fault signals. However, due to different sensor arrangements and environmental disturbances, the acquired signals may change, which led to different diagnostic results. To improve the reliability of diagnosis, a signal fusion technique was proposed in [106]. The proposed method combined SAE and DBN. Data features were extracted from sensors and fed into SAE for feature fusion. The SAE has two hidden layers with the same structure and parameters, while DBN has three layers. The training data were vibration signals of under inner race fault and outer race fault with 1260 samples. The classification accuracy was up to 97.82%. For feature fusion, time- and frequency-domain features from various sensor signals are extracted and fed into several two-layer sparse AE (SAE) neural networks. The fused feature vectors are used to machine health indicators. An *et al.* [107] proposed a simple deep learning clustering method, inspired by Hu *et al.* [43], which applied manifold learning to off-the-shelf embeddings to find an alternative model for clustering networks and simply combined the manifold learning method. The proposed model included five layers, and the size of the input layer was equal to the size of input data. The ReLu was applied for the activation function, while the Adam optimizer was implemented. Bearing fault data were used for experimental validation. The dataset contained ten types of bearing health conditions with 10 000 samples. The average classification accuracy was up to 98.99%. Instead of complex clustering networks, the shallow clustering algorithm presents a new deep clustering method for E2LMC, which is based on AE embedding local stream shape learning for unsupervised bearing fault diagnosis.

Table III summarizes the deep learning methods for fault classification. The merits and demerits of deep learning are presented as follows.

Merits:

1) *High Learning Ability:* Deep learning performs well and has a great learning ability.
2) *Strong Adaptability:* Deep learning contains complex neural network layers that can be mapped to any function to solve complicated classification problems.
3) *Well Portability:* Many frameworks can be used.

Demerits:

1) Computationally intensive.
2) Deep learning requires large data and computing power so that the cost is extremely high.
3) High hardware requirements.
4) Complex model design.

VI. Other AI-Based Approaches

In addition to the traditional ML and deep learning introduced above, some other popular artificial intelligence-based

TABLE III
HIGHLIGHT OF DEEP LEARNING

Reference	Theme	Principle	Highlight
[77]	DCNN	• Multiple hidden layers • Learn hierarchical representations	• Deep mining signal features • To require large training dataset
[74]	CNN	• 1-D Convolutional Neural Network • Scalar multiplication and addition • Raw data input	• To decrease computing cost • High efficiency • Insufficient classification accuracy
[75]	Sparse filtering CNN	• Unsupervised two-layer neural network • Softmax regression	• Adaptive learning feature • No manual labeling required
[76]	CCA KLD CNN	• Moving Average Technique • Fault matrix input	• Statistics are more sensitive to initial failures • To improve computational efficiency • No prior knowledge required
[78]	CNN RNN	• Vibration signal input	• Efficient and time-saving
[79]	CNN ELM	• Weighted orthogonality constraint • Randomly generate hidden weights	• To reduce training complexity • To reduce manual intervention • To improve generalization performance
[80]	RNN-WDCNN	• Vibration data input • Domain adaptation and noise suppression	• No feature extraction required • High Robustness
[81]	WDCNN	• Raw vibration signal input • Adaptive batch normalization	• To improve network accuracy • No pre-processing required
[82]	TICNN	• End-to-end learning • 1-D structure input • Direct denoising	• No manual feature extraction process required • High robustness • High accuracy
[83]	SVM-CNN	• Unsupervised fault detection • Online multi-fault classification	• High generalization ability • Robustness
[84] [108]	CNN	• 2-D image conversion • Eliminate the impact of manual features • Easier handling of big data	• Time-consuming training • High hardware conditions • Great training capabilities
[85]	FFT CWT CLSTM	• Multi-channel array input • Input length sensitivity analysis	• Efficient learning of spatio-temporal features • To enhance classification performance
[86] [109]	LSTM CBLSTM	• Original sensing signal composition revealing fault characteristics	• Abstraction and deep learning ability • No expert knowledge required
[87] [53]	AE-RNN	• Time series data dimensionality reduction • Raw vibration signal input	• To reduce dimensionality Effective feature extraction • To improve classification efficiency
[110]	SGU RNN	• Single-gate joint recurrent neural network • Wavelet Packet Decomposition	• To reduce network parameters • To reduce computation time
[94] [96] [111]	GAN	• Unsupervised sample generation • Adaptively learning samples • Generated sample input	• Strong generalization capability • Non-original feature signal input • Not an end-to-end structure
[95]	SDAE-GAN	• Estimating the probability distribution • Extended Diagnostic Sample	• Automatic extraction of effective fault characteristics • Stronger noise immunity • Improved generalization capability
[57]	DAAN	• Feature compression and visualization • Vibration signal input	• Effectively learn domain-invariant features. • Diagnosis under different operating conditions
[98]	CDBN CS	• Reduce vibration data • Layered generation model	• No manual feature extraction required • Great generalizability
[99]	ACO DDBN	• Layer-by-layer unsupervised pre-training • Search for the optimal number of neurons and learning rate of the hidden layer	• Accurate determination of model parameters • To increase the computational burden
[54][106]	SAE DBN	• Feature Fusion • Unsupervised learning weights	• Insensitive to training samples • High classification accuracy
[97]	HBN	• Two-Layer DBN • Vibration signal input	• High reliability • To overcome overlap problem caused by noise
[52] [101]	SDA	• Health status division • Greedy training	• Adaptive mining of salient fault characteristics • High diagnostic accuracy • Strong robustness
[103] [105]	NSAE LCN	• Raw vibration signal input • Obtain translation invariant features	• To generate displacement invariant characteristics • Effectively identify the mechanical health condition
[104]	AE	• Association Matrix	• To dig complex information • High diagnostic accuracy

method. The proposed model included 5 layers and the size of

TABLE IV
HIGHLIGHT OF OTHER AI-BASED METHODS

Reference	Theme	Principle	Highlight
[116]	ANFIS	• If-then rule • Membership functions input	• Nonlinear Approximation Capability • High adaptability
[117]	DT ANFIS	• Rule conversion • Vibration and current signal input	• To reduce data volume • To enhance classification performance • Less accurate classification of the current signal
[118]	ASSBDIM	• Sparse filtering • Singular spectrum analysis • Time series data input	• High accuracy • Low calculation cost • Suitable for online applications
[59]	HMM-ANFIS	• Markov model prediction • Probability density function	• Nonlinear mapping • Real-time state estimation
[119]	Compress data Deep learning	• Nonlinear projection for compression capture • Unsupervised learning	• Do not rely on a priori knowledge • To dig hidden data • Suitable for handling large amounts of data
[134]	Feature Incremental Broad Learning	• Reconfiguration system • Simplify the model	• High diagnostic accuracy • Less training time • The feature extraction method is not optimal
[125]	Infrared Thermography ML	• 2-D thermal image input • Optimal eigenvector matrix transfer	• Non-destructive • Non-contact detection • Performance affected by redundant information noise
[126]	Transfer learning	• Feature transfer • High-dimensional mapping of data	• To reduce feature data requirement • Cannot be applied to multi-target working condition
[127]	TFR	• Smooth blurred plane feature extraction	• To reduce computational burden • To improve feature separability
[120]	GA-PCA LDA ANN	• Estimating high-dimensional hybrid feature sets	• Simultaneous monitoring multiple faults
[28]	GA-ANN	• Weighting and threshold optimization	• Fast structure determination • High classification performance
[131]	AAS	• Unsupervised classification • Data Clustering	• To improve monitoring effectiveness • Removal of redundant or irrelevant information
[128] [132]	GA-AIS	• Spatial transformation	• To reduce computational complexity
[133]	Faster AP-MSDL	• Adaptive parameter estimation • Multi-scale learning	• Short running time • Great fault detection performance

methods are also reviewed in this section, including hybrid algorithms combining fuzzy logic (FL) and ANN methods (ANFIS), TL, compressed sensing (CS), infrared thermography (IRT), temperature estimation strategy, genetic algorithms, and so on. Table IV summarizes the characteristics of each AI-based method.

A. Adaptive Neurofuzzy Inference System (ANFIS)

ANNs have proven to be a reliable technique for diagnosing motion conditions with high learning capabilities. However, ANNs are not interpretable and cannot explain specific decisions to the user in a human-understandable form. Another technique used for FDD is FL. It can imitate human knowledge according to clear and understandable linguistic terms and then convert the linguistic and heuristic terms into complex machine computed values through fuzzy rules and auxiliary functions. The initial parameters and auxiliary functions of if-then rules are usually prepared by experts. Therefore, FL needs to be fine-tuned to obtain an acceptable rule base and optimize the parameters for the available data. The integration of these two methods can solve only one problem from FL or ANN. This method has been applied to motor fault diagnosis [111]. ANFIS [112] is a special neurofuzzy classifier approach that combines the adaptive capabilities of ANNs with an FL qualitative approach. It has been successfully implemented for automatic FDD of IMs [113]. In recent years, ANFIS and its variations with other methods have been widely developed as fault classification techniques. In terms of bearing fault diagnosis, ANFIS with genetic algorithm [114] and ANFIS with WT [115] are two examples of combined algorithms. ANFIS has been used to classify faults in IMs with variable drive speeds [116]. If the measurement data are large and include redundant noise, the accuracy of the output will be greatly reduced when the data are fed directly into the classifier. Feature extraction and selection can minimize the dimensionality of the data by selecting the basic features, which refers to the transition of the current features to the lower dimensional space. However, when feature extraction is completed, each feature set contains several redundant or unnecessary features, as well as significant features in the feature space. Therefore, a feature selection process is required to select the minimum features that define the system state from the entire feature set. As a feature extraction method, decision trees were employed in [117], which was a process for removing redundant features from data to reduce the quantity of data required for efficient learning, classification accuracy, a compact and simply understood knowledge base, and rapid computation. For fault diagnosis of IMs, it integrated CARTs and ANFIS adaptive techniques.

Since insufficient accuracy of the measurement equipment, model errors, or the influence of environmental conditions of the measurement process leads to noise in the signal that usually cannot be avoided, noise filtering mechanisms are important to improve the ability to preprocess the measured data for analysis and extraction of valuable information. Singular spectrum analysis (SSA) and sparse filtering (SSA/SF) are promising tools and the effectiveness of FL depends heavily on the accuracy of the fuzzy set, which involves logical relationships between fuzzy rules and relational functions input and output. Tran *et al.* [118] proposed an ANFIS based on SSA/SF bearing fault detection method, which was combined with an online detection method. Information was extracted from the measurement data with noise by SSA and SF, and a database with tagged data was created. The ANFIS parameters were optimized by the tagged data, and then, faults were detected by the ANFIS comparator. In [59], a higher order particle filter was developed to make predictions based on an m-order HMM that integrates ANFIS and modeling noise. Advanced particle filters were applied to system status modeling. By combining ANFIS and process noise, a high-order HMM was formed that can be used to describe the fault process. Due to the dynamics of the system, an online model adaptive scheme for fault propagation was required. ANFIS was trained with available conditional data to build ANFIS for laminar imaging trend models. A new classification method based on time–frequency representation (TFR) and criterion-based decision-making was proposed in [127]. The fuzzy Doppler delay plane was used as a feature extraction space to reduce the computational effort. This is an error probability model based on a statistical approach for selecting the optimal number of features to be extracted from the streamwise plane and then using the capability of a neural network to learn a nonlinear functional relationship between the input and output. The model is designed to maximize the separability between different classes.

B. Compressed Sensing

The increasing demand for information about complex mechanical systems in recent years has promoted the use of high-dimensional information. To examine health conditions comprehensively, monitoring systems based on traditional sampling theorems are commonly employed. Multiple sensors collect huge amounts of data at high sampling rates over a long operational cycle, placing a strain on both the hardware and the data storage. In the era of big data driving motor health monitoring, how to extract effective information from a large amount of data is an important topic, and CS has attracted widespread attention. This method theoretically obtains all the information contained in the original signal and achieves the reduction of the signal in size. It frees the data from Nyquist theory and enables compressive acquisition by nonlinear projection in the transform space. Thus, the sampled data are greatly reduced and contain all the information. Fault diagnosis is achieved based on the measured data in the compressed domain, which is supported by the depth domain. A new intelligent FDD combining CS and DNN was proposed in [119]. Random projections in the transform domain were used as compressed samples to perform sample compression, and then, a stacked AE network based on SAE was built on mining fault information using a deep network architecture. In addition, the detection of multiple hybrid faults is an essential topic. Aucedo-Dorantes *et al.* [120] presented a diagnostic method for tracking and analyzing the frequency of multiple and combined faults in IM, depending on the measurement and optimization of a high-dimensional hybrid feature set. The proposed method can detect the possibility of multiple and combined faults occurring simultaneously.

C. Infrared Thermography and Temperature Estimation Strategy

When constructing an electric motor or selecting its control approach, the significant thermal stress on possibly defective components of the motor must be taken into consideration. Competitive pressure and high production costs, particularly in the automotive industry, lead engineers to look for new ways to increase the safety of embedded materials. Because overheating can severely damage motors, accurate temperature information must be provided during operation. Among the typically critical components sensitive to overheating, such as stator end windings and bearings, permanent magnets in the rotor are particularly susceptible to damage to the motor. Although sensor-based measurements provide a quick and accurate picture of the machine's thermal state, it is not economically or technically feasible to assess rotor temperature in this way. In particular, direct rotor monitoring techniques, such as IRT [121], [122] or classical thermocouples with slip rings on the shaft [123], have not yet entered industrial mass production. Therefore, based on this model, the study focuses on estimating the rotor and permanent magnet temperatures. In contrast to physically driven estimation methods, ML models will differ from any classical approximation of the underlying thermal theory [124]. Fig. 14 shows the idea of fitting the ML model to the collected test bench data and eventually informing any controller. The more accurate the thermal state information obtained by the control system, the better the ability to monitor critical operations and apply derating power. In addition, suppose rich datasets can be recorded on the test bench or in production in the automotive industry. In that case, engineers can rely on them to model simpler temperature estimators than deep neural networks. The use of domain-specific expertise and instrumentation specifications can be avoided. In addition, certain classical supervised learning algorithms can handle the temperature of essential components in PMSM in real time, as demonstrated by extensive benchmark analysis of many different benchmarks. Alternatively, IRT can

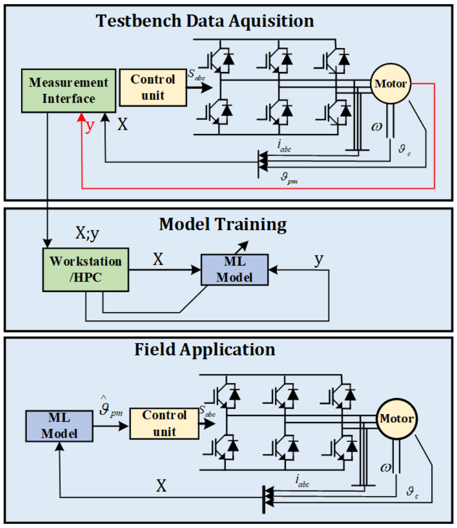


Fig. 14. Entire process test bench.

be used to detect bearing failures in rotating machinery in a nondestructive, noncontact manner. However, performance is limited by the minuscule amount of information and string noise in IR thermal images. To address this problem, Choudhary *et al.* [125] developed a 2-D-DWT-based IRT technique for identifying different bearing faults in IM, such as inner and outer ring cracks and insufficient lubrication, using a 2-D-DWT-based IRT method. PCA was used to decrease the dimensionality of the recovered features to obtain the best feature selection. Then, a mar distance (MD) procedure was used to rank the selected features in order of importance. For fault classification and performance evaluation, these chosen characteristics were input into complex decision trees (CDTs), LDA, and SVMs.

D. Transfer Learning

Fault signal processing includes the separation and extraction of fault features for later use in predictive models. Fault discrimination identifies fault types by testing the fault features obtained through signal processing using a complete training model. Typically, the number of fault samples is usually smaller than the number of regular samples, and the problem of data class imbalance will lead to the overfitting of many data-sensitive classification models. The data imbalance problem can be solved by weighting cost, but, for practical motor fault diagnosis, it is difficult to determine the weighting parameters. Therefore, obtaining as much fault data as possible from data sources is a major issue. Simulation of fault scenarios is the easiest way to obtain fault characterization results. However, it is impractical for mechanical parts of motor equipment to intentionally damage them to achieve fault characteristics. Therefore, it is a challenge to make full use of historical data and ensure the accuracy of the new task model. Xu *et al.* [126] proposed a method to solve the problem of degradation of classification accuracy due to different fault data caused by feature transfer. In feature learning, transfer component analysis (TCA) is a classical approach to solve adaptive problems. Data in both source and target domains are mapped to a high-dimensional replica kernel Hilbert space (RKHS). The data distance between sources is minimized, while interior features are preserved. The TCA data feature extraction method is used to discover standard features in both the source and target domains. Through feature learning, a fault diagnosis model under historical operating conditions can be built to address the problem of insufficient data effectively.

E. GA-Based Approach

Genetic algorithms are search algorithms based on natural selection and genetic principles, characterized by parallel computation and global optimality, as their derivation process is a method of approaching the optimal solution. Moreover, compared with general optimization methods, genetic algorithms require less information to achieve optimal control. Therefore, genetic algorithms are often used to optimize the parameters and structure of neural networks or FL systems. Unal *et al.* [28] presented a diagnostic method based on the envelope, FFT, and backpropagation GA-ANN algorithm for rolling bearing faults to determine the NN structure accurately. The proposed GA-ANN combination strategy can be used to assist experts in designing appropriate, fast, and accurate ANN structures for specific diagnostic problems. Besides, to improve feature separability, several feature selection methods include GA, and particle swarm-based optimization has different drawbacks, which consists of the tendency to fall into local optima and computational inefficiency [129]. The presence of redundant and insignificant features in the high-dimensional feature set can minimize fault identification accuracy. Moreover, unlike traditional classifiers, the artificial immune system (AIS) is a commonly used technique for anomaly detection, independent of the *a priori* knowledge of the model failure modes. For motor bearing faults, traditional fault detection methods are based on solving specific problems, and their design depends on various aspects of the bearing. Abid *et al.* [128] proposed a GA optimization algorithm that combined unique multidomain feature extraction with a supervised artificial intelligence technique (GA-AIS) to train the detector using the minimum amount of information. The proposed method can be used to detect fault types under a wide range of operating conditions. Generally, there is no universal deterministic approach that can be used for all practical fault situations. Negative selection algorithms (NSAs) are used for detectors to distinguish standard data from faulty data. NSA and its variants analyze the vibration and current signals of the motor. Different detector generation strategies form NSA with different variants, including GA optimization, clone selection optimization, and stochastic native [130], [131]. However, the increase in data leads to a decrease in the speed of detector generation and detection.

F. Artificial Ant System (AAS) and Dictionary Learning

A uniform sampling window samples the signal, while the more petite frame width and smaller size are insufficient to accommodate the input signal [132]. Simultaneously, the high-dimensional signal increases the computational burden.

Soualhi *et al.* [8] proposed a new optimization technique based on the AAS. It was successfully validated for motor faults. Besides, dictionary learning is a powerful method for extracting feature conditions. The multiscale dictionary learning method with transformed coefficients performs well in extracting fault signals and requires less learning time compared to ordinary dictionary learning. Zhang *et al.* [133] proposed a faster adaptive parametric multiscale dictionary learning method that adaptively selected the learning scale and simultaneously estimated the sparse coding parameters in dictionary learning. The technique has a relatively short running time and great fault detection performance. Jiang *et al.* [134] developed an effective and responsive motor fault diagnosis method that combined feature extraction, feature incremental generalized learning after collecting the starting sample data, and processing it. These processed data were imported into broad learning to train the network. The network was continuously trained by feature incremental broad learning until the test accuracy was satisfactory.

VII. Discussion and Future Development

AI-based techniques have been proven to be highly effective for FDD in motor drive systems. Deep neural network-based methods have gradually become attractive in industrial applications due to the continuous upgrading of computer hardware. As the layers of neural networks deepen, leading to the disappearance of locally optimal solutions and gradients, pretraining methods alleviate the local optimal solution problem, and deep learning gradually becomes popular, including DBN, CNN, RNN, LSTM, and so on. To overcome gradient disappearance, transfer functions, such as ReLU and max out, replace sigmoid and form the basic form of DNN. DBN simply converts an image matrix into a 1-D vector as input without considering the 2-D structure information of the image. It can represent the distribution of data from a statistical point of view and can reflect the similarity of similar data itself. However, the generative model cannot obtain the optimal classification surface between different categories, resulting in that the classification accuracy may not be as high as the discriminative model. In the area of image recognition, the CNN is a popular study topic. Its weight-sharing network topology resembles biological brain networks, minimizing the complexity of the model and the number of weights. Traditional recognition methods require extensive feature extraction and data reconstruction, so the image may be utilized directly as the input of the model. The signal of each layer of neuron in CNN can only propagate up one layer, and the processing of samples is independent at each moment, called forward neural network. In RNN, the output of a neuron can act directly on itself at the next timestamp. In addition to solving the temporal gradient disappearance, long short-time memory units (LSTMs) are proposed to prevent gradient disappearance by implementing temporal memory function through gate switching. GAN is a generative learning algorithm that can efficiently mine the intrinsic distribution of a dataset using a small number of samples. In addition, the discriminative power of GAN discriminators makes it possible to build integrated fault diagnosis models. Besides, AEs can perform unsupervised learning from data samples and achieve great performance on different datasets without any new feature engineering.

The different signal pattern of the motor state is usually uncertain with changing operating conditions, and how to effectively detect motor state under varying operating conditions is also a major challenge for future research. Traditional static neural networks can be optimized by adjusting the model parameters and structure to obtain higher performance control effects and monitoring performance. In addition, the concept of dynamic neural networks is proposed as a novel research topic in deep learning. Compared to static models with fixed parameters, dynamic neural networks can adapt to different input signals by changing the parameter structure, which offers excellent advantages in accuracy and self-adaptability. Therefore, dynamic neural networks with different state characteristics due to variable load states have great potential in electric vehicle powertrain condition monitoring. Moreover, deep reinforcement learning, a recently popular deep learning algorithm, also offers promising and feasible solutions for control drives of electric motors due to its more efficient learning capability. In conclusion, the adaptive adjustment of neural network parameters and model structure has a wide application space in electric vehicle powertrain systems to obtain reliable and efficient motor drive performance to meet the high requirements of safe driving of electric vehicles. Furthermore, data-driven AI-based algorithms can effectively utilize big data, which are motivated by end-to-end frameworks that provide satisfactory performance with low requirements for domain knowledge. The improvement of robustness and generality of AI-based techniques aims further to enhance the AI-based capability of techniques in FDD applications. They should apply to different tasks, including detection, diagnosis, and prediction in various domains. Therefore, the development of interpretable AI-based technologies has attracted increasing attention in intelligent monitoring. In addition to FDD of motor, they are further used for decision support. Eventually, autonomous condition monitoring and fault warning of the entire electric vehicle powertrain system can be achieved.

VIII. Conclusion

In summary, this article reviews the application of artificial intelligence techniques in motor fault diagnosis. Comprehensive studies of motor faults and their severity are still rare and have so far been limited to diagnosing faults in motors under specific operating conditions. It is difficult to detect faults in motors under light loads. In addition, the accuracy of fault diagnosis may be reduced due to fluctuations in rotor speed during data acquisition at different loads. Therefore, it remains an open challenge to consider the impact of motor operating conditions on AI-based fault diagnosis. Artificial intelligence for pattern recognition or fault diagnosis includes a large number of different types of mathematical tools, i.e., preprocessing, extraction, and selection of appropriate statistical features, and selection of model parameters. It is a challenge to choose which tool is best suited for a particular problem and machine in different situations. Moreover, there is a lot of scope for research in this area since condition monitoring techniques for FDD of rotating machinery have been improved from traditional methods to artificial intelligence methods. Artificial intelligence-based diagnostic systems still have some challenging tasks to accomplish in terms of their efficiency, reliability, computation time, adequate database, and robustness.

Overall, in this article, common electric motor and fault types are presented. The different FDD methods have been also analyzed. The mainstream methods of feature extraction are reviewed, and the application of artificial intelligence techniques in motor fault diagnosis is stated in a large space and reviewed in three parts through theoretical background: traditional ML, deep learning algorithms, and hybrid algorithms. In the deep learning part, almost all algorithms are based on ANNs, developed and changed to achieve different purposes. Finally, the latest developments, research gaps, and future challenges in fault monitoring and diagnosis of motor faults are discussed.

REFERENCES

[1] H. Henao *et al.*, "Trends in fault diagnosis for electrical machines: A review of diagnostic techniques," *IEEE Ind. Electron. Mag.*, vol. 8, no. 2, pp. 31–42, Jun. 2014.

[2] S. S. S. R. S. Duvvuri, "Rotor inter-turn short-circuit fault diagnostics relevant modeling of slip-ring induction motor," in *Proc. IEEE 8th Power India Int. Conf. (PIICON)*, Dec. 2018, pp. 1–6.

[3] R. R. Kumar *et al.*, "Induction machine stator fault tracking using the growing curvilinear component analysis," *IEEE Access*, vol. 9, pp. 2201–2212, 2021.

[4] X. Wang, Z. Wang, Z. Xu, M. Cheng, W. Wang, and Y. Hu, "Comprehensive diagnosis and tolerance strategies for electrical faults and sensor faults in dual three-phase PMSM drives," *IEEE Trans. Power Electron.*, vol. 34, no. 7, pp. 6669–6684, Jul. 2019.

[5] J. Cai and X. Zhao, "Synthetic hybrid-integral-threshold logic-based position fault diagnosis scheme for SRM drives," *IEEE Trans. Instrum. Meas.*, vol. 70, no. 12, pp. 1–8, Oct. 2020.

[6] M. Z. Ali, M. N. S. K. Shabbir, X. Liang, Y. Zhang, and T. Hu, "Machine learning-based fault diagnosis for single- and multi-faults in induction motors using measured stator currents and vibration signals," *IEEE Trans. Ind. Appl.*, vol. 55, no. 3, pp. 2378–2391, May 2019.

[7] X. Wang, S. Lu, and S. Zhang, "Rotating angle estimation for hybrid stepper motors with application to bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 8, pp. 5556–5568, Aug. 2020.

[8] A. Soualhi, G. Clerc, and H. Razik, "Detection and diagnosis of faults in induction motor using an improved artificial ant clustering technique," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 4053–4062, Sep. 2013.

[9] F. Immovilli, C. Bianchini, M. Cocconcelli, A. Bellini, and R. Rubini, "Bearing fault model for induction motor with externally induced vibration," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3408–3418, Aug. 2013.

[10] Z. Wang and C. S. Chang, "Online fault detection of induction motors using frequency domain independent components analysis," in *Proc. IEEE Int. Symp. Ind. Electron.*, Jun. 2011, pp. 2132–2137.

[11] Z. Wang, C. S. Chang, and Y. Zhang, "A feature based frequency domain analysis algorithm for fault detection of induction motors," in *Proc. 6th IEEE Conf. Ind. Electron. Appl.*, Jun. 2011, pp. 27–32.

[12] M. Motahari-Nezhad and S. M. Jafari, "Bearing remaining useful life prediction under starved lubricating condition using time domain acoustic emission signal processing," *Expert Syst. Appl.*, vol. 168, Apr. 2021, Art. no. 114391.

[13] R. C. Ugras, O. K. Alkan, S. Orhan, M. Kutlu, and A. Mugan, "Real time high cycle fatigue estimation algorithm and load history monitoring for vehicles by the use of frequency domain methods," *Mech. Syst. Signal Process.*, vol. 118, pp. 290–304, Mar. 2019.

[14] E. H. El Bouchikhi, V. Choqueuse, and M. Benbouzid, "Induction machine faults detection using stator current parametric spectral estimation," *Mech. Syst. Signal Process.*, vols. 52–53, pp. 447–464, Feb. 2015.

[15] K. A. Loparo, M. L. Adams, W. Lin, M. F. Abdel-Magied, and N. Afshari, "Fault detection and diagnosis of rotating machinery," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 1005–1014, Oct. 2000.

[16] T. W. S. Chow and H.-Z. Tan, "HOS-based nonparametric and parametric methodologies for machine fault detection," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 1051–1059, Oct. 2000.

[17] M. El Hachemi Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 984–993, 2000.

[18] P. S. Priyanga, K. Krithivasan, P. S., and S. V. S. Sriram, "Detection of cyberattacks in industrial control systems using enhanced principal component analysis and hypergraph-based convolution neural network (EPCA-HG-CNN)," *IEEE Trans. Ind. Appl.*, vol. 56, no. 4, pp. 4394–4404, Jul./Aug. 2020.

[19] M. Kaminski, C. T. Kowalski, and T. Orlowska-Kowalska, "Application of radial basis neural networks for the rotor fault detection of the induction motor," in *Proc. IEEE Int. Conf. Comput. Tool (EUROCON)*, Apr. 2011, pp. 1–4.

[20] E. M. T. Eldin, H. R. Emara, E. M. Aboul-Zahab, and S. S. Refaat, "Monitoring and diagnosis of external faults in three phase induction motors using artificial neural network," in *Proc. IEEE Power Eng. Soc. General Meeting*, Jun. 2007, pp. 1–7.

[21] S. Premrudeepreechacharn, T. Utthiyuung, K. Kruepengkul, and P. Puongkaew, "Induction motor fault detection and diagnosis using supervised and unsupervised neural networks," in *Proc. IEEE Int. Conf. Ind. Technol.*, Bangkok, Thailand, Dec. 2002, pp. 93–96.

[22] R. Natarajan, "Failure identification of induction motors by sensing unbalanced stator currents," *IEEE Power Eng. Rev.*, vol. 9, no. 12, pp. 35–36, Dec. 1989.

[23] S. Aksoy and A. Muhurcu, "Recurrent neural network based nonlinear state estimation for induction motors," *J. Int. Rev. Elect. Eng.*, vol. 6, no. 2, pp. 636–645, Apr. 2011.

[24] M. Fernandez-Temprano, P. E. Gardel-Sotomayor, O. Duque-Perez, and D. Morinigo-Sotelo, "Broken bar condition monitoring of an induction motor under different supplies using a linear discriminant analysis," in *Proc. 9th IEEE Int. Symp. Diagnostics Electr. Mach., Power Electron. Drives (SDEMPED)*, Aug. 2013, pp. 162–168.

[25] J. Seshadrinath, B. Singh, and B. K. Panigrahi, "Investigation of vibration signatures for multiple fault diagnosis in variable frequency drives using complex wavelets," *IEEE Trans. Power Electron.*, vol. 29, no. 2, pp. 936–945, Feb. 2014.

[26] W. S. Gongora, H. V. D. Silva, A. Goedtel, W. F. Godoy, and S. A. O. da Silva, "Neural approach for bearing fault detection in three phase induction motors," in *Proc. 9th IEEE Int. Symp. Diag. Electr. Mach., Power Electron. Drives (SDEMPED)*, Aug. 2013, pp. 566–572.

[27] O. Duque-Perez, L.-A. Garcia-Escudero, D. Morinigo-Sotelo, P.-E. Gardel, and M. Perez-Alonso, "Analysis of fault signatures for the diagnosis of induction motors fed by voltage source inverters using ANOVA and additive models," *Electr. Power Syst. Res.*, vol. 121, pp. 1–13, Apr. 2015.

[28] M. Unal, M. Onat, M. Demetgul, and H. Kucuk, "Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network," *Meas., J. Int. Meas. Confederation*, vol. 58, pp. 187–196, Dec. 2014.

[29] D. Yu, J. Cheng, and Y. Yang, "Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings," *Mech. Syst. Signal Process.*, vol. 19, no. 2, pp. 259–270, Mar. 2005.

[30] Z. K. Peng, P. W. Tse, and F. L. Chu, "An improved Hilbert–Huang transform and its application in vibration signal analysis," *J. Sound Vibrat.*, vol. 286, nos. 1–2, pp. 187–205, Aug. 2005.

[31] G. K. Singh and S. A. K. S. Ahmed, "Vibration signal analysis using wavelet transform for isolation and identification of electrical faults in induction machine," *Electr. Power Syst. Res.*, vol. 68, no. 2, pp. 119–136, Feb. 2004.

[32] V. Iglesias, J. Grajal, M. A. Sánchez, and M. López-Vallejo, "Implementation of a real-time spectrum analyzer on FPGA platforms," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 2, pp. 338–355, Feb. 2015.

[33] G. Betta, C. Liguori, A. Paolillo, and A. Pietrosanto, "A DSP-based FFT-analyzer for the fault diagnosis of rotating machine based on vibration analysis," *IEEE Trans. Instrum. Meas.*, vol. 51, no. 6, pp. 1316–1322, Dec. 2002.

[34] V. K. Rai and A. R. Mohanty, "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2607–2615, Aug. 2007.

[35] Y. Qin, S. Qin, and Y. Mao, "Research on iterated Hilbert transform and its application in mechanical fault diagnosis," *Mech. Syst. Signal Process.*, vol. 22, no. 8, pp. 1967–1980, Nov. 2008.

[36] N. Mohanti, *Signal Processing Signal Filtering Detection*. New York, NY, USA: Van Nostrand, 1987.
[37] A. F. Atiya and A. G. Parlos, "New results on recurrent network training: Unifying the algorithms and accelerating convergence," *IEEE Trans. Neural Netw.*, vol. 13, no. 3, pp. 765–786, May 2000.
[38] H. Su and K. T. Chong, "Induction machine condition monitoring using neural network modeling," *IEEE Trans. Ind. Electron.*, vol. 54, no. 1, pp. 241–249, Feb. 2007.
[39] R. Jiang, S. Liu, Y. Tang, and Y. Liu, "A novel method of fault diagnosis for rolling element bearings based on the accumulated envelope spectrum of the wavelet packet," *J. Vibrat. Control*, vol. 21, no. 8, pp. 1580–1593, Jun. 2015.
[40] J. Chen *et al.*, "Wavelet transform based on the inner product in fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process*, vols. 70–71, pp. 1–35, Mar. 2016.
[41] R. G. C. Cunha, E. T. da Silva, and C. M. de Sá Medeiros, "Machine learning and multiresolution decomposition for embedded applications to detect short-circuit in induction motors," *Comput. Ind.*, vol. 129, Aug. 2021, Art. no. 103461.
[42] Y. Hu, X. Tu, F. Li, H. Li, and G. Meng, "An adaptive and tacholess order analysis method based on enhanced empirical wavelet transform for fault detection of bearings with varying speeds," *J. Sound Vibrat.*, vol. 409, pp. 241–255, Nov. 2017.
[43] W. Hu, H. Chang, and X. Gu, "A novel fault diagnosis technique for wind turbine gearbox," *Appl. Soft Comput.*, vol. 82, Sep. 2019, Art. no. 105556.
[44] F. Ding, X. Zhang, W. Wu, and Y. Wang, "A fault feature extraction method of motor bearing using improved LCD," *IEEE Access*, vol. 8, pp. 220973–220979, 2020.
[45] N. Rehman, M. M. Khan, M. I. Sohaib, M. Jehanzaib, S. Ehsan, and K. McDonald-Maier, "Image fusion using multivariate and multidimensional EMD," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 5112–5116.
[46] J.-H. Ahn, D.-H. Kwak, and H.-B. Koh, "Fault detection of a roller-bearing system through the EMD of a wavelet denoised signal," *Sensors*, vol. 14, no. 8, pp. 15022–15038, 2014.
[47] M. M. Garcillanosa, A. M. B. Flores, S. M. M. Jala, and A. J. P. Toleza, "FPGA based powerline and baseline interference removal in electrocardiogram using modified EWT-DWT filtering," in *Proc. 2nd World Symp. Commun. Eng. (WSCE)*, Dec. 2019, pp. 142–146.
[48] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.
[49] X.-B. Wang, Z.-X. Yang, and X.-A. Yan, "Novel particle swarm optimization-based variational mode decomposition method for the fault diagnosis of complex rotating machinery," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 68–79, Feb. 2018.
[50] J. S. Cheng, J. D. Zheng, and Y. Yang, "A nonstationary signal analysis approach—The local characteristic-scale decomposition method," *J. Vibrat. Eng.*, vol. 25, no. 2, pp. 215–220, Feb. 2012.
[51] Z. Wu, J. Cheng, and Y. Yang, "Autonomous dense local feature scale decomposition method and its application," *J. Mech. Eng.*, vol. 6, no. 11, pp. 1464–1470, Dec. 2015.
[52] L. Chenjie, W. Bin, and T. Tao, "Blind estimation algorithm of frequency hopping signal parameters using local feature scale decomposition," *Signal Process.*, vol. 5, no. 3, pp. 308–313, Dec. 2015.
[53] Y. Huang, C.-H. Chen, and C.-J. Huang, "Motor fault detection and feature extraction using RNN-based variational autoencoder," *IEEE Access*, vol. 7, pp. 139086–139096, 2019.
[54] V. Choqueuse, M. E. H. Benbouzid, Y. Amirat, and S. Turri, "Diagnosis of three-phase electrical machines using multidimensional demodulation techniques," *IEEE Trans. Ind. Electron.*, vol. 59, no. 4, pp. 2014–2023, Apr. 2012.
[55] A. O. M. Luiz, L. C. P. Flávio, and E. M. A. Paulo, "Automatic detection of surface defects on rolled steel using computer vision and artificial neural networks," in *Proc. 36th Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Nov. 2010, pp. 1081–1086.
[56] M. I. Chacon-Murguia and S. Gonzalez-Duarte, "An adaptive neural-fuzzy approach for object detection in dynamic backgrounds for surveillance systems," *IEEE Trans. Ind. Electron.*, vol. 59, no. 8, pp. 3286–3298, Aug. 2012.
[57] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega, and H. Henao, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3398–3407, Aug. 2013.
[58] X. F. St-Onge, J. Cameron, S. Saleh, and E. J. Scheme, "A symmetrical component feature extraction method for fault detection in induction machines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7281–7289, Sep. 2019.
[59] C. Chen, B. Zhang, G. Vachtsevanos, and M. Orchard, "Machine condition prediction based on adaptive neuro–fuzzy and high-order particle filtering," *IEEE Trans. Ind. Electron.*, vol. 58, no. 9, pp. 4353–4364, Sep. 2011.
[60] X. An and H. Zeng, "Fault diagnosis method for spherical roller bearing of wind turbine based on variational mode decomposition and singular value decomposition," *J. Vibroeng.*, vol. 18, no. 6, pp. 3548–3556, Sep. 2016.
[61] S.-W. Fei, "Fault diagnosis of bearing based on wavelet packet transform-phase space reconstruction-singular value decomposition and SVM classifier," *Arabian J. Sci. Eng.*, vol. 42, no. 5, pp. 1967–1975, May 2017.
[62] D. T. Hoang and H. J. Kang, "A motor current signal-based bearing fault diagnosis using deep learning and information fusion," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 3325–3333, Jun. 2020.
[63] X. Zhang, B. Wang, and X. Chen, "Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine," *Knowl.-Based Syst.*, vol. 89, pp. 56–85, Nov. 2015.
[64] Q. Hu, X.-S. Si, A.-S. Qin, Y.-R. Lv, and Q.-H. Zhang, "Machinery fault diagnosis scheme using redefined dimensionless indicators and mRMR feature selection," *IEEE Access*, vol. 8, pp. 40313–40326, 2020.
[65] J. Zheng, H. Pan, and J. Cheng, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines," *Mech. Syst. Signal Process.*, vol. 85, pp. 746–759, Feb. 2017.
[66] S. Das, P. Purkait, C. Koley, and S. Chakravorti, "Performance of a load-immune classifier for robust identification of minor faults in induction motor stator winding," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 21, no. 1, pp. 33–44, Feb. 2014.
[67] W. Fengqi and G. Meng, "Compound rub malfunctions feature extraction based on full-spectrum cascade analysis and SVM," *Mech. Syst. Signal Process.*, vol. 20, no. 8, pp. 2007–2021, Nov. 2006.
[68] V. N. Ghate and S. V. Dudul, "Cascade neural-network-based fault classifier for three-phase induction motor," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 1555–1563, May 2011.
[69] S. Zhang, X. Li, M. Zong, X. Zhu, and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1774–1785, May 2018.
[70] A. Recioui, B. Benseghier, and H. Khalfallah, "Power system fault detection, classification and location using the K-nearest neighbors," in *Proc. 4th Int. Conf. Electr. Eng. (ICEE)*, Dec. 2015, pp. 1–6.
[71] Q. Yao, J. Wang, L. Yang, H. Su, and G. Zhang, "A fault diagnosis method of engine rotor based on random forests," in *Proc. IEEE Int. Conf. Prognostics Health Manage. (ICPHM)*, Jun. 2016, p. 3.
[72] R. K. Patel and V. K. Giri, "Feature selection and classification of mechanical fault of an induction motor using random forest classifier," *Perspect. Sci.*, vol. 8, pp. 334–337, Sep. 2016.
[73] M. Seera, C. P. Lim, and C. K. Loo, "Motor fault detection and diagnosis using a hybrid FMM-CART model with online learning," *J. Intell. Manuf.*, vol. 27, no. 6, pp. 1273–1285, Dec. 2016.
[74] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, Nov. 2016.
[75] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.
[76] H. Chen, B. Jiang, T. Zhang, and N. Lu, "Data-driven and deep learning-based detection and diagnosis of incipient faults with application to electrical traction systems," *Neurocomputing*, vol. 396, pp. 429–437, Jul. 2020.
[77] S. Shao, R. Yan, Y. Lu, P. Wang, and R. Gao, "DCNN-based multi-signal induction motor fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 6, pp. 2658–2669, Jun. 2020.
[78] K. Liang, N. Qin, D. Huang, and Y. Fu, "Convolutional recurrent neural network for fault diagnosis of high-speed train bogie," *Complexity*, vol. 2018, pp. 1–13, Oct. 2018.
[79] Z. Chen, K. Gryllias, and W. Li, "Mechanical fault diagnosis using convolutional neural networks and extreme learning machine," *Mech. Syst. Signal Process.*, vol. 133, Nov. 2019, Art. no. 106272.

[80] A. Shenfield and M. Howarth, "A novel deep learning model for the detection and identification of rolling element bearing faults," *Sensors*, vol. 20, no. 18, pp. 1–24, 2020.

[81] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, Feb. 2017.

[82] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018.

[83] J. S. L. Senanayaka, H. Van Khang, and K. G. Robbersmyr, "Toward self-supervised feature learning for online diagnosis of multiple faults in electric powertrains," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 3772–3781, Jun. 2021.

[84] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.

[85] M. Jalayer, C. Orsenigo, and C. Vercellis, "Fault detection and diagnosis for rotating machinery: A model based on convolutional LSTM, fast Fourier and continuous wavelet transforms," *Comput. Ind.*, vol. 125, Feb. 2021, Art. no. 103378.

[86] R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with convolutional bi-directional LSTM networks," *Sensors*, vol. 17, no. 2, p. 273, Jan. 2017.

[87] T. Wong and Z. Luo, "Recurrent auto-encoder model for large-scale industrial sensor signal analysis," in *Proc. Int. Conf. Eng. Appl. Neural Netw.*, Jul. 2018, pp. 203–216.

[88] L. D. Xu and L. Duan, "Big data for cyber physical systems in industry 4.0: A survey," *Enterprise Inf. Syst.*, vol. 13, no. 2, pp. 148–169, Feb. 2019.

[89] Y. Ma, K. Liu, Z. Guan, X. Xu, X. Qian, and H. Bao, "Background augmentation generative adversarial networks (BAGANs): Effective data generation based on GAN-augmented 3D synthesizing," *Symmetry*, vol. 10, no. 12, p. 734, Dec. 2018.

[90] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Trans. Signal Inf. Process.*, vol. 3, no. e2, pp. 1–29, Jan. 2014.

[91] K. Cheng, R. Tahir, L. K. Eric, and M. Li, "An analysis of generative adversarial networks and variants for image synthesis on MNIST dataset," *Multimedia Tools Appl.*, vol. 79, nos. 19–20, pp. 13725–13752, May 2020.

[92] H. Apaydin, H. Feizi, M. T. Sattari, M. S. Colak, S. Shamshirband, and K.-W. Chau, "Comparative analysis of recurrent neural network architectures for reservoir inflow forecasting," *Water*, vol. 12, no. 5, p. 1500, 2020.

[93] C. Zhang, C. Xiong, and L. Wang, "A research on generative adversarial networks applied to next generation," in *Proc. 14th Int. Conf. Comput. Sci. Educ. (ICCSE)*, Aug. 2019, pp. 913–917.

[94] Y. O. Lee, J. Jo, and J. Hwang, "Application of deep neural network and generative adversarial network to industrial maintenance: A case study of induction motor fault detection," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2017, pp. 3248–3253.

[95] Z. Wang, J. Wang, and Y. Wang, "An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition," *Neurocomputing*, vol. 310, pp. 213–222, Oct. 2018.

[96] Y. Ding, L. Ma, J. Ma, C. Wang, and C. Lu, "A generative adversarial network-based intelligent fault diagnosis method for rotating machinery under small sample size conditions," *IEEE Access*, vol. 7, pp. 149736–149749, 2019.

[97] M. Gan and C. Wang, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 72, pp. 92–104, May 2016.

[98] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, and S. Wu, "Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing," *Mech. Syst. Signal Process.*, vol. 100, pp. 743–765, Feb. 2018.

[99] M. Ma, C. Sun, and X. Chen, "Discriminative deep belief networks with ant colony optimization for health status assessment of machine," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 12, pp. 3115–3125, Dec. 2017.

[100] T. Zhang, W. Wang, H. Ye, D. Huang, H. Zhang, and M. Li, "Fault detection for ironmaking process based on stacked denoising autoencoders," in *Proc. Amer. Control Conf. (ACC)*, Jul. 2016, pp. 3261–3267.

[101] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, Jan. 2017.

[102] H. Lee, Y. Kim, and C. O. Kim, "A deep learning model for robust wafer fault monitoring with sensor measurement noise," *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 1, pp. 23–31, Feb. 2017.

[103] F. Jia and Y. Lei, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," *Neurocomputing*, vol. 272, pp. 619–628, Jan. 2018.

[104] W. Mao, W. Feng, Y. Liu, D. Zhang, and X. Liang, "A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis," *Mech. Syst. Signal Process.*, vol. 150, Mar. 2021, Art. no. 107233.

[105] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vol. 72, pp. 303–315, May 2016.

[106] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.

[107] J. An, P. Ai, C. Liu, S. Xu, and D. Liu, "Deep clustering bearing fault diagnosis method based on local manifold learning of an autoencoded embedding," *IEEE Access*, vol. 9, pp. 30154–30168, 2021.

[108] T. Nakazawa and D. V. Kulkarni, "Wafer map defect pattern classification and image retrieval using convolutional neural network," *IEEE Trans. Semicond. Manuf.*, vol. 31, no. 2, pp. 309–314, May 2018.

[109] R. Sabir, D. Rosato, S. Hartmann, and C. Guehmann, "LSTM based bearing fault diagnosis of electrical machines using motor current signal," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2019, pp. 613–618.

[110] W. Tan, Y. Sun, D. Qiu, Y. An, and P. Ren, "Rolling bearing fault diagnosis based on single gated unite recurrent neural networks," *J. Phys., Conf.*, vol. 1601, Jul. 2020, Art. no. 042017.

[111] J. Suwatthikul and S. Sornmuang, "Fault detection and diagnosis of a motor bearing shield," in *Proc. 6th IEEE Int. Conf. Intell. Data Acquisition Adv. Comput. Syst.*, Sep. 2011, pp. 332–335.

[112] H. Abu-Rub, S. M. Ahmed, A. Iqbal, M. Rahimian, and H. A. Toliyat, "Incipient bearing fault diagnostics for inverter fed induction motor drive using ANFIS," in *Proc. 19th Int. Conf. Electr. Mach. (ICEM)*, Sep. 2010, pp. 1–5.

[113] M. Saeidi, J. Zarei, H. Hassani, and A. Zamani, "Bearing fault detection via Park's vector approach based on ANFIS," in *Proc. Int. Conf. Mechatronics Control (ICMC)*, Jul. 2014, pp. 2436–2441.

[114] J. Zarei, H. Hassani, Z. Wei, and H. R. Karimi, "Broken rotor bars detection via Park's vector approach based on ANFIS," in *Proc. IEEE 23rd Int. Symp. Ind. Electron. (ISIE)*, Jun. 2014, pp. 2422–2426.

[115] Y. Lei, Z. He, Y. Zi, and Q. Hu, "Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs," *Mech. Syst. Signal Process.*, vol. 21, no. 5, pp. 2280–2294, 2007.

[116] X. Lou and K. A. Loparo, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," *Mech. Syst. Signal Process.*, vol. 18, no. 5, pp. 1077–1095, 2004.

[117] V. T. Tran, B.-S. Yang, M.-S. Oh, and A. C. C. Tan, "Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 1840–1849, Mar. 2009.

[118] Q. T. Tran, S. D. Nguyen, and T.-I. Seo, "Algorithm for estimating online bearing fault upon the ability to extract meaningful information from big data of intelligent structures," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3804–3813, May 2019.

[119] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 1, pp. 185–195, Jan. 2018.

[120] J. J. Saucedo-Dorantes, A. Y. Jaen-Cuellar, M. Delgado-Prieto, R. D. J. Romero-Troncoso, and R. A. Osornio-Rios, "Condition monitoring strategy based on an optimized selection of high-dimensional set of hybrid features to diagnose and detect multiple and combined faults in an induction motor," *Measurement*, vol. 178, Jun. 2021, Art. no. 109404.

[121] M. Ganchev, B. Kubicek, and H. Kappeler, "Rotor temperature monitoring system," in *Proc. 19th Int. Conf. Electr. Mach. (ICEM)*, Sep. 2010, pp. 1–5.

[122] S. Stipetic, M. Kovacic, Z. Hanic, and M. Vrazic, "Measurement of excitation winding temperature on synchronous generator in rotation using infrared thermography," *IEEE Trans. Ind. Electron.*, vol. 59, no. 5, pp. 2288–2298, May 2012.

[123] C. Mejuto, M. Mueller, M. Shanel, A. Mebarki, M. Reekie, and D. Staton, "Improved synchronous machine thermal modelling," in *Proc. 18th Int. Conf. Electr. Mach.*, Sep. 2008, pp. 1–6.

[124] W. Kirchgässner, O. Wallscheid, and J. Böcker, "Data-driven permanent magnet temperature estimation in synchronous motors with supervised machine learning," 2020, *arXiv:2001.06246*. [Online]. Available: https://arxiv.org/abs/2001.06246

[125] A. Choudhary, D. Goyal, and S. S. Letha, "Infrared thermography-based fault diagnosis of induction motor bearings using machine learning," *IEEE Sensors J.*, vol. 21, no. 2, pp. 1727–1734, Jan. 2021.

[126] W. Xu, Y. Wan, T.-Y. Zuo, and X.-M. Sha, "Transfer learning based data feature transfer for fault diagnosis," *IEEE Access*, vol. 8, pp. 76120–76129, 2020.

[127] T. Boukra, A. Lebaroud, and G. Clerc, "Statistical and neural-network approaches for the classification of induction machine faults using the ambiguity plane representation," *IEEE Trans. Ind. Electron.*, vol. 60, no. 9, pp. 4034–4042, Sep. 2013.

[128] A. Abid, M. T. Khan, and M. S. Khan, "Multidomain features-based GA optimized artificial immune system for bearing fault detection," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 1, pp. 348–359, Jan. 2020.

[129] B. Xue, M. Zhang, and W. N. Browne, "Particle swarm optimization for feature selection in classification: A multi-objective approach," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1656–1671, Dec. 2013.

[130] X. Z. Gao, S. J. Ovaska, and X. Wang, "Genetic algorithm-based detector generation in negative selection algorithm," in *Proc. IEEE Mountain Workshop Adapt. Learn. Syst.*, Logan, UT, USA, Jul. 2006, pp. 133–137.

[131] X. Z. Gao, S. J. Ovaska, X. Wang, and M.-Y. Chow, "Multi-level optimization of negative selection algorithm detectors with application in motor fault detection," *Intell. Autom. Soft Comput.*, vol. 16, no. 3, pp. 353–375, Jan. 2010.

[132] X. Z. Gao, H. Xu, X. Wang, and K. Zenger, "A study of negative selection algorithm-based motor fault detection and diagnosis," *Int. J. Innov. Comput. Inf. Control*, vol. 9, no. 2, pp. 875–901, 2013.

[133] Z. Zhang *et al.*, "Faster multiscale dictionary learning method with adaptive parameter estimation for fault diagnosis of traction motor bearings," *IEEE Trans. Instrum. Meas.*, vol. 70, no. 12, pp. 1–13, Oct. 2020.

[134] S. B. Jiang, P. K. Wong, and Y. C. Liang, "A fault diagnostic method for induction motors based on feature incremental broad learning and singular value decomposition," *IEEE Access*, vol. 7, pp. 157796–157806, 2019.

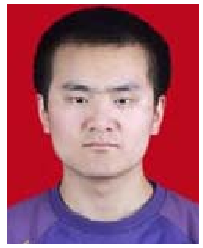
Wangjie Lang (Student Member, IEEE) was born in Zhejiang, China, in December 1998. He received the B.Eng. degree in electrical engineering from the University of Strathclyde, Glasgow, U.K., and Lanzhou University of Technology, Lanzhou, China, in 2020. He is currently pursuing the Ph.D. degree with a major in electrical engineering with the Department of Electronic Engineering, University of York, York, U.K.

His research interests include electrical machine fault detection and diagnosis and EV powertrain engineering-based AI techniques.

Yihua Hu (Senior Member, IEEE) received the B.S. degree in electrical motor drives and the Ph.D. degree in power electronics and drives from China University of Mining and Technology, Xuzhou, China, in 2003 and 2011, respectively.

From 2011 to 2013, he was a Post-Doctoral Fellow with the College of Electrical Engineering, Zhejiang University, Hangzhou, China. From 2013 to 2015, he was a Research Associate with the Power Electronics and Motor Drive Group, University of Strathclyde, Glasgow, U.K. He is currently a Reader with the Department of Electronic Engineering, University of York (UoY), York, U.K. His research interests include renewable generation, power electronics converters & control, electric vehicles, more electric ship/aircraft, smart energy systems, and nondestructive test technology.

Dr. Hu is also an Associate Editor of *IET Renewable Power Generation*, *IET Intelligent Transport Systems*, and *Power Electronics and Drives*.



Chao Gong (Student Member, IEEE) was born in Shandong, China, in February 1991. He received the B.Eng. and M.Eng. degrees in electrical engineering from Northwestern Polytechnical University, Xi'an, China, in 2014 and 2016, respectively. He is currently pursuing the Ph.D. degree with a major in electrical engineering with the Department of Electronic Engineering, University of York, York, U.K.

His research interests include electrical machine design and drives and motion control.

Xiaotian Zhang (Student Member, IEEE) was born in Shandong, China, in August 1996. He received the B.Sc. degree in applied physics from Hohai University, Nanjing, China, in 2018, and the M.Sc. degree in electronic engineering from King's College London, London, U.K., in 2020. He is currently pursuing the Ph.D. degree with a major in electrical engineering with the Department of Electronic Engineering, University of York, York, U.K.

His research interests include advanced AI condition monitoring, fault recognition, and engineering application of AI.

Hui Xu received the B.S., M.Eng., and Ph.D. degrees from the China University of Mining and Technology, Xuzhou, China, in 2005, 2009, and 2012, respectively.

From 2012 to 2013, she was a Post-Doctoral Researcher with Zhejiang University, Hangzhou, China. She is currently a Research Assistant in artificial intelligence and energy with Leeds Beckett University, Leeds, U.K.

Jiamei Deng (Senior Member, IEEE) received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1988, the M.S. degree from Shanghai Institute of Mechanical Engineering, Shanghai, China, in 2001, and the Ph.D. degree from the University of Reading, Reading, U.K., in 2005.

She is currently a Professor of artificial intelligence and energy with Leeds Beckett University, Leeds, U.K., with an emphasis on control systems design, predictive models, data-driven, and power plant safety monitoring.