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, Xiaotian Zhang *Student Member, IEEE*, Hui Xu, 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 of the motor drive system. This paper reviews the application of AI techniques in motor fault detection and diagnosis 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—Motor fault; AI-based techniques; Machine learning; Deep learning.

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.

W. Lang, Y. Hu, C. Gong and X. Zhang are with the Department of Electronic Engineering, University of York, Heslington, York, YO10 5DD, U.K. (E-mail: wl1526@york.ac.uk, yihua.hu@york.ac.uk, 1542101806@qq.com, xz2764@york.ac.uk).
H.Xu and J.Deng are with the School of Built Environment, Engineering and Computing, Leeds Beckett University, Leeds LS6 3QS, U.K. (E-mail: J.Deng@leedsbeckett.ac.uk, Xuhuicunt2001@gmail.com).

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 non-invasive motor fault detection and diagnosis techniques, different signals can be analyzed, including temperature analysis, vibration analysis, infrared analysis, current and voltage analysis, electromagnetic field analysis, ultrasonic analysis, etc. Vibration analysis [7-9] and motor current signature analysis (MCSA) [10] [11] are two of the most popular research in this field. Stator current analysis is known for providing non-invasive 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].
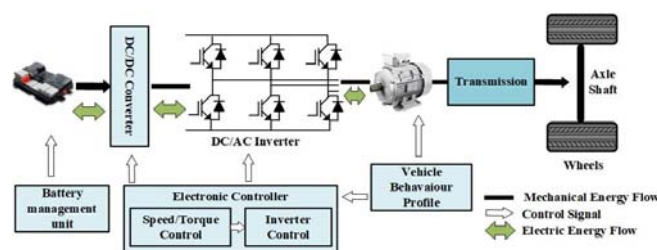

Fig. 1. The basic structure of the motor drive system.

In general, the motor FDD techniques are divided into 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, etc.
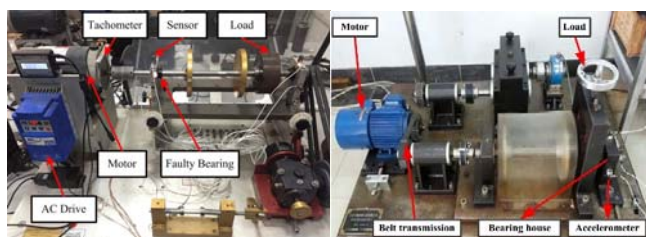

Fig. 2. The experimental rig [42][106].

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. 2. Fault classification. In signal-based feature extraction engineering, the time domain signal is usually converted to the frequency domain signal by 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, Hilbert transforms (HT), wavelet transform (WT) and other signal processing techniques have been widely proposed. Furthermore, 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, 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 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 artificial neural networks, which can be employed to perform

classification and regression tasks in the motor drive system. In [20], Artificial neural networks (ANN) were applied to track and diagnose external faults in three-phase induction motors. The data of stator voltage, current, and motor speed were used to train the network. Furthermore, the identification and diagnosis of induction motor 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 induction motor is determined by the imbalance of the inductor current. [23] investigated the state estimation of induction motors based on classical machine learning 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 fault detection and diagnosis. It is important to investigate the contributions of different scholars in this field.

This paper aims to review the methods of motor fault detection and diagnosis based on artificial intelligence techniques. The full paper 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 machine learning 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.
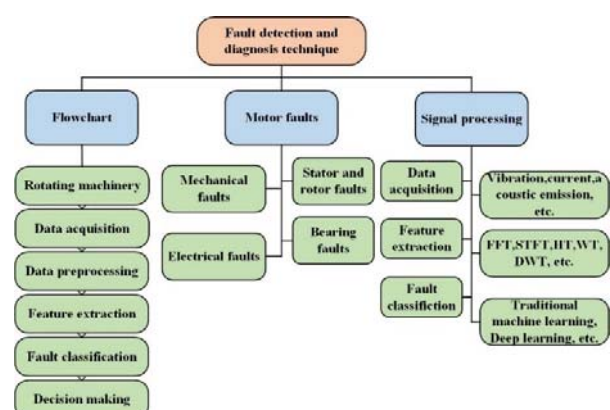

Fig. 3. General flowchart of FDD techniques.

II. Electric Motor and fault types

In the advanced powertrain system, induction motors (IM) and permanent magnet synchronous motors (PMSM) are the two major types of electric vehicle motors. Induction motors are widely used in industry and are characterized by stator and rotor made of laminated silicon 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, the structure is simple with reliable operation. Induction motors (IM) 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, 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 permanent magnet synchronous motor (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.
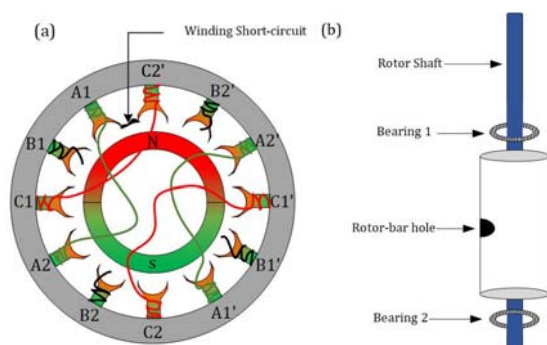

Fig. 4. Typical faults: (a) Stator winding short-circuit (WSC). (b) Rotor crack fault (CRF).

Electric motors and drives are affected by many different types of faults, 1) Stator faults can be classified as stator open phase, the short circuit which 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 induction motors, 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.

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.

Bearing faults: Bearing faults can be caused by several factors. i) Due to the huge output load torque, there is a high vibration in the rotor, which leads to large fatigue stress. ii) Incorrect bearing installation. iii) High currents in the bearing directed by shaft voltage cause lubrication to deteriorate. iv) 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.

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

Motor current signature analysis (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, 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. Due to 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; 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 oil and lubrication analysis. Online (particle counting, temperature and viscosity monitoring, etc.) and offline (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 owing 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. Non-invasive 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 differs 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,

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 application [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]. [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 Hilbert transform (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 short-time Fourier transform (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 Short Time Fourier Transform (STFT), an analysis method using vibration spectrum neural network modeling was developed in [38] to extract fault spectrum features for detecting machine faults. For non-stationary processing signals, the wavelet transform (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 multi-resolution inspection capabilities, this approach has shown great ability in fault diagnosis of mechanical machinery [40].

The Discrete wavelet transform (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 sub-bands. Wavelet packet transform (WPT) was proposed to decompose the high-frequency band and 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 multi-modulated vibration signal was decomposed into multiple intrinsic mode functions (IMFs) that were regarded as dominant mode components. It can extract both stationary and non-stationary 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]. [47] proposed an empirical wavelet transform (EWT), which combined the advantages of wavelet transform 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 non-smooth 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 non-smooth signal to be decomposed into the sum of several intrinsic rotational components. Although it has better performance over EMD, the approach cannot account for the physical meaning of the intrinsic rotational components. [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].

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

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 linear discriminant analysis (LDA) can be utilized for 2D 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 data sets 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 sub-tow 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 of 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 singular value decomposition 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, [62] proposed a fault diagnosis method based on a motor current signal developing deep learning and information fusion (IF). The novel decision-level information fusion (IF) technique applied the raw signals of multi-phase motor currents as direct inputs and extracted features from the motor current signals of each phase to improve classification accuracy. Table I summarize all mentioned signal process methods for feature extraction.

V. FAULT DIAGNOSIS

As the crucial step in fault detection and diagnosis, accurate and efficient classification performance is extremely important. As an attractive method, machine learning 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 sub-domain of machine learning, and its related algorithms are inspired by the structure and function of the brain (ie, artificial neural networks). Deep learning uses a hierarchical framework of non-nearline transformations of input information to construct statistical models and output outcomes. Traditional machine learning 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 machine learning and popular deep learning algorithms in motor FDD are introduced in this section.
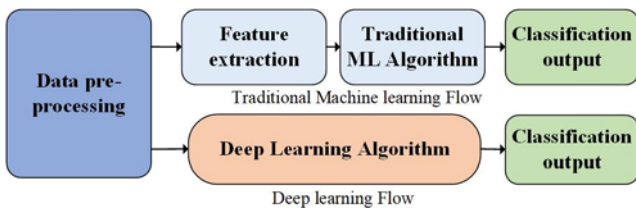

Fig. 5. Flowchart of traditional ML and Deep learning.

1) *Support vector machine (SVM) based approach*

A support vector machine (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, image recognition, etc. 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 fault diagnosis. Fig. 6 illustrates the optimal hyperplane for binary classification by SVM.
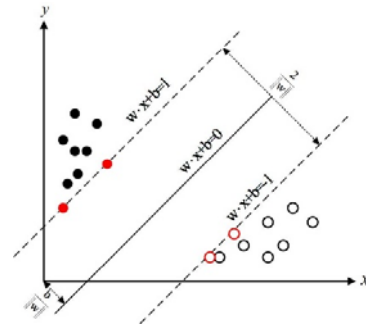

Fig. 6. The optimal hyperplane for binary classification by SVM.

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

$$subject\ to, y_i(w.x_i + b) \geq 1; i = 1,2,\dots,n \qquad (2)$$

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, the equation is:

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

After optimization, the basic SVM structure can be described as:

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

SVM has a nonlinear multi-label classification function. For multiple fault diagnosis, an intelligent fault detection system based on a multivariate integrated incremental support vector machine (MEISVM) is presented in [63] for multi-fault

diagnosis. This method can detect a variety of faults, including complex compound faults and faults with different severity thresholds. [64] proposed a grid search support vector machine (GSSVM) approach based on redefined dimensionless indicators (RDI) 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 support vector machines (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 multi-fault 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 winding short-circuit failures and supply voltage imbalance at different load speeds. A recursive feature elimination algorithm based on the support vector machine (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 the 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. 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 sub-optimal solutions. The full spectrum experimental data were used for support vector machine (SVM) training and excellent classification results were obtained [67].

2) *MLP/KNN/RF-based approach*

Multi-layer 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. Since 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 inter-turn short-circuit faults in PMSM stator windings at different speeds, loads, and fault severity. 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 inter-turn short circuits.
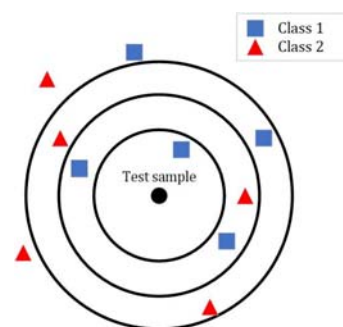

Fig. 7. The structure of KNN.

Due to its easy implementation and substantial classification efficiency [69], the k-nearest neighbors (KNN) method is a

TABLE II
HIGHLIGHT OF CLASSICAL MACHINE LEARNING 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

popular classification method in data mining and statistics. The principle structure of KNN is as shown in Fig. 7. [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 principal component analysis (PCA) and semi-supervised k-nearest neighbor (KNN) distances to provide health metrics.
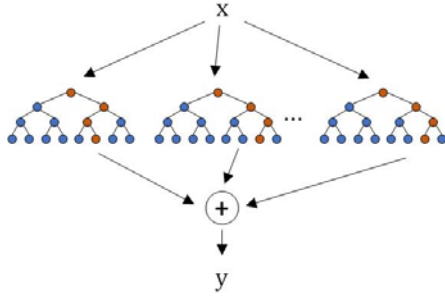

Fig. 8. The basic structure of the random forest.

Random forest (RF) models are based on the classification and regression of tree groups (CART) [71]. Random forest 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 random forest 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 random forest. A random forest 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. [73] developed a hybrid model that combines a fuzzy min-max (FMM) neural network with classification and regression tree (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 machine learning methods mentioned above are all shallow models with simple structures, high computational efficiency, and great classification performance. Table II summarizes the classical machine learning methods for fault classification.

3) *Deep learning*

In contrast to traditional machine learning 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 as well as 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 machine learning techniques cannot effectively distinguish complex information in raw data, and the shallow structure of artificial neural networks 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 fault detection and diagnosis.

A. *Convolutional Neural Network (CNN)*

Convolutional Neural Network (CNN) is a feed-forward 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 two-dimensional data like images. Convolutional layers (CL), pooling layers (PL), 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:

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

where x_i^{l-1} indicates the input data of layer l, $*$ 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:

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

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

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

After the non-linear 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 equation:

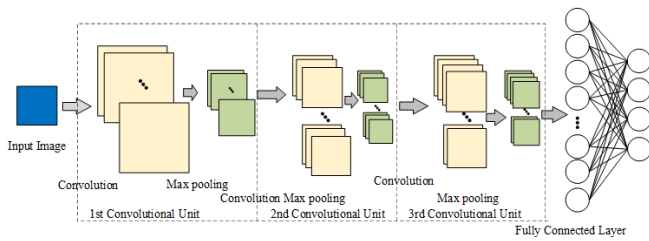$$p_i = \frac{\exp(z_i)}{\sum_{j=1}^{N} \exp(z_j)} \tag{8}$$

Fig. 9. The basic structure diagram of CNN.

Signal processing techniques with different characteristics are combined with CNN for fault signal diagnosis [74] [75]. [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. [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 convolutional neural network (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 multi-signal 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-pooling layers. ReLu was implemented as an activation function, 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 201s and 156s, 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 recurrent neural network (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 that 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] [81], the initial vibration signal was input to a deep convolutional neural network named deep convolutional neural networks with wide first-layer kernels (WDCNN). The. A large first-layer convolutional kernel and a deep network structure with narrow convolutional layers 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 pre-processing, 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 convolutional neural network (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 data sets and the training time consumption. The data was 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 two-dimensional (2-D) image, the method extracted the features of the converted 2-D image and eliminates the influence of artificial features. The novel data pre-processing approach presented in the literature transformed raw time-domain signal data into two-dimensional grayscale pictures without the use of any predetermined parameters, removing as much as possible the knowledge 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 can be used to minimize training time in the future based on these constraints.

B. Recurrent neural network (RNN)

Recurrent neural networks (RNN) 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 two-dimensional 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 below:

$$S_t = f(U * x_t + W * S_{t-1}) \tag{9}$$
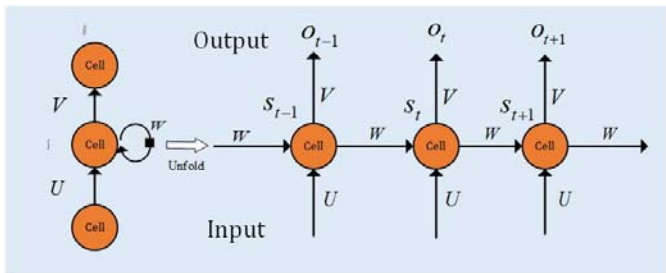
$$O_t = f(s_{t-1} * V) \tag{10}$$

Fig. 10. The basic structure diagram of RNN.

Due to the RNN map the target vector from the entire history of previous inputs, the memory of prior inputs are 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]. [80] presented the RNN-WDCNN, a new dual-path recurrent neural network with a larger initial kernel and deeper convolutional neural network routes that can operate on raw temporal information. The WDCNN combined the functions of recurrent neural networks (RNN) with convolutional neural networks to capture remote relationships in time-series data and remove input signals in high-frequency noise. Besides, [85] developed a novel approach named convolutional long short-term memory (CLSTM) is developed to handle signal-based FDD in rotating machinery. The developed model improved the efficiency of processing multi-channel input data and learning its Spatio-temporal characteristics. The input channels were subjected to sensitivity analysis, and the results suggest that combining these multi-domain characteristics improves the accuracy of the classifier. However, feature engineering on severely unbalanced data sets 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 multi-channel input data and understand its Spatio-temporal characteristics more efficiently. The updated CLSTM supported the FDD approach to properly understand the function of data structures and achieve higher accuracy. [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 bi-directional 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 seconds to train, and each sample takes 0.027 seconds to test. Compared with RNN, deep RNN, LSTM, etc., the proposed algorithm has 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 and 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 autoencoder of the RNN [87]. Moreover, [53] was used in combination with a variant autoencoder (VAE), variance and noise are added to make the model generation more realistic. The time-domain vibration signals at 3 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.

C. Generative Adversarial Network (GAN)

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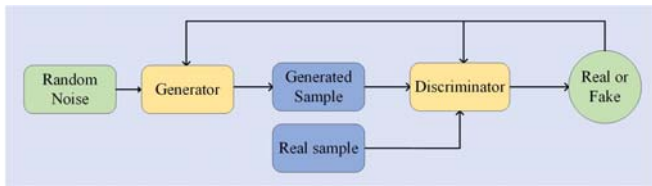

Fig. 11. The basic structure diagram of GAN.

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:

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

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

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 seen in equation (13).

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

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. 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 discriminator have completed training, the output of the generator will produce samples close to the actual samples even if the input data is random noise. On a variety of image datasets, GAN variants such as DC-GA and AC-GAN have demonstrated strong generative efficiency. [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. Owing 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 autoencoders and generative adversarial networks, [95] proposed a fault diagnosis process (SDAE-GAN) that enabled the discriminator to determine the sample fault type and whether the input data is 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. [96] studied multiple GANs to understand the data distribution for each health condition and then developed a semi-supervised 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 sigmod 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%.

D. Deep Belief Network (DBN)

In contrast to traditional neural network discriminative models, deep belief networks (DBN) 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 multiple layers of constrained Boltzmann machines (RBMs), 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.

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

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:

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

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

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

where η is the learning rate and $\langle \cdot \rangle_{p(h|v)}$ 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's the same as having a priori knowledge of the supervised learning input data.
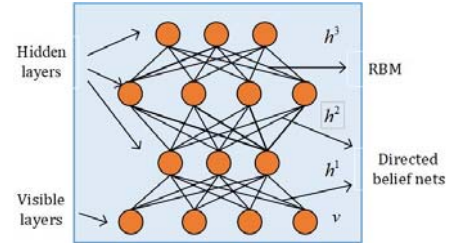

Fig. 12. The basic structure diagram of DBN.

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 non-smoothness 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. [98] proposed an improved convolutional deep confidence network (CDBN) with compressive sensing (CS). The structure of the proposed algorithm was divided into one gaussian input layer, 6 convolutions hidden layers,6 pooling layers connected with 12 convolutions hidden layers and 12 pooling layers. The output was fed into a softmax classifier for fault classification. The compressed vibration signals as input data and there two cases were considered for verification include single fault and compound faults. 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 deep belief network (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 1592.5s 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 pre-train unsupervised models regardless of the amount of available training data. However, the performance of deep belief networks 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.

E. Autoencoder

The autoencoder 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 autoencoder is a neural network that consists of two components: an encoder function and a decoder that generates reconfigurations. Traditionally, autoencoders have been used for dimensionality reduction or functional learning. An improved prototype structure of the autoencoder has resulted in a DAE that first complete the pre-training 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 pre-training. SAE only changes one thing: the input data is 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.
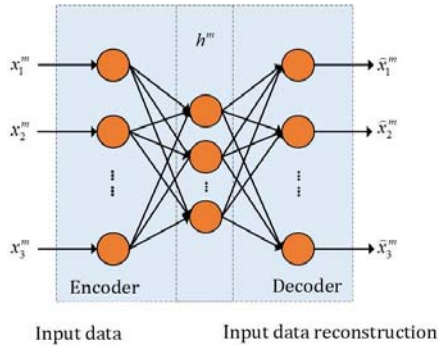

Fig. 13. The basic structure diagram of AE.

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 below:

$$h_i = f_{\theta_1}(W_1 x_i + b_1) \qquad (21)$$
$$\hat{x}_i = g_{\theta_2}(W_2 h_i + b_2) \qquad (22)$$

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

The purpose of the autoencoder 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.

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

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

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 follow:

$$\hat{\theta}_1 \leftarrow \theta_1 - \lambda \frac{\partial L_{rec}}{\partial \theta_1} \qquad (25)$$

$$\hat{\theta}_2 \leftarrow \theta_2 - \lambda \frac{\partial L_{rec}}{\partial \theta_2} \qquad (26)$$

where λ is the learning rate.

A monitoring method based on stacked denoising autoencoders was proposed in [100]. The stacked denoising autoencoder-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 local connection network constructed by normalized sparse autoencoder (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] [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 autoencoder 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 was 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 autoencoder (SAE) neural networks. The fused feature vectors are used to machine health indicators. [107] proposed a simple deep learning clustering method, inspired by [43], which applied manifold learning to off-the-shelf embeddings to find an alternative model for clustering networks and simply combined the manifold learning

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

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 was used for experimental validation. The dataset contained 10 types of bearing health conditions with 10000 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 autoencoder 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 machine learning and deep learning introduced above, some other popular artificial intelligence-based methods are also reviewed in this section including hybrid algorithms combining fuzzy logic and artificial neural network methods (ANFIS), transfer learning (TL), compressed sensing (CS), Infrared thermography and temperature estimation strategy, and genetic algorithms, etc. Table IV summarizes the characteristics of each AI-based method.

A. Adaptive neuro-fuzzy system (ANFIS)

Artificial neural networks (ANN) have proven to be a reliable technique for diagnosing motion conditions with high learning capabilities. However, artificial neural networks are not interpretable and cannot explain specific decisions to the user in a human-understandable form. Another technique used for fault detection and diagnosis is fuzzy logic (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, fuzzy logic needs to be fine-tuned to obtain an acceptable rule base and to optimize the parameters for the available data. The integration of these two methods can solve only one problem from fuzzy logic or ANN. This method has been applied to motor fault diagnosis [112]. Adaptive Neuro-Fuzzy Inference System (ANFIS) [113] is a special neuro-fuzzy

classifier approach that combines the adaptive capabilities of artificial neural networks with a fuzzy logic qualitative approach. It has been successfully implemented for automatic fault detection and diagnosis of induction motors [114]. 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 [115] and ANFIS with wavelet transform [116] are two examples of combined algorithms. ANFIS has been used to classify faults in induction motors with variable drive speeds [117]. If the measurement data is large and includes redundant noise, the accuracy of the output will be greatly reduced when the data is fed directly into the classifier. Feature extraction and selection can minimize the dimensionality of the data by selecting the basic features, which refer 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 [118], 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 induction motors, it integrated Classification and Regression Trees (CART) 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 pre-process 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. [119] 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 criteria-based decision-making was proposed in [128]. 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 acquire a large amount of while data at high sampling rates over a long operating cycle puts a burden not only on the hardware but also on 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 compressed sensing (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 is greatly reduced and contains 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 [120]. Random projections in the transform domain were used as compressed samples to perform sample compression, and then a stacked autoencoder network based on sparse autoencoder (SAE) was built on mining fault information using a deep network architecture. In addition, the detection of multiple hybrid faults is an essential topic. [121] 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 infrared thermography [122] [123] or classical thermocouples with slip rings on the shaft [124], 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, machine learning (ML) models will differ from any classical approximation of the underlying thermal theory [125]. 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 data sets 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, infrared thermography (IRT) can be used to detect bearing failures in rotating machinery in a non-destructive, non-contact manner. However, performance is limited by the minuscule amount of information and string noise in IR thermal images. To address this problem, [126] developed a two-dimensional discrete wavelet transform (2D-DWT) based IRT technique for identifying different bearing faults in IM, such as inner and outer ring cracks and insufficient lubrication, using a two-dimensional discrete wavelet transform (2D-DWT) based IRT method. Principal component analysis (PCA) was used to decrease the dimensionality of the recovered features to obtain the best feature selection. and 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 (CDT), linear discriminant analysis (LDA), and support vector machines (SVM).
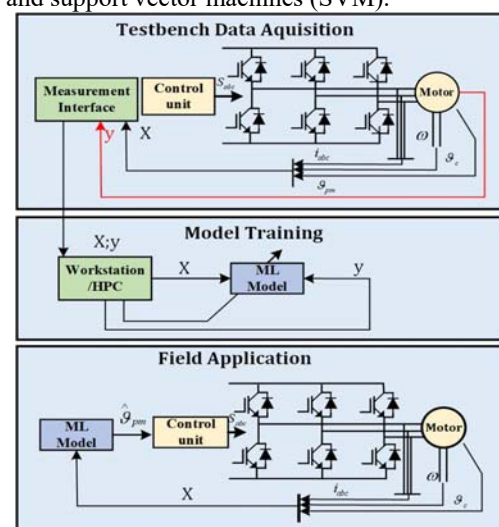

Fig. 14. The entire process test bench.

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. [127] 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 fuzzy logic systems. [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 have different drawbacks consist of the tendency to fall into local optima and computational inefficiency [130]. 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. [129] proposed a GA optimization algorithm that combined unique multi-domain 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 (NSA) 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 [131] [132]. However, the increase of 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 [133]. Simultaneously, the high-dimensional signal increases the computational burden. [134] proposed a new optimization technique based on the Artificial Ant System (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. [135] 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. [136] 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 fault detection and diagnosis 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. Pre-training methods alleviate the local optimal solution problem, and deep learning gradually becomes popular including DBN, CNN, RNN, LSTM, etc. 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 one-dimensional vector as input without considering the two-

TABLE IV
HIGHLIGHT OF OTHER AI-BASED METHOD

Reference	Theme	Principle	Highlight
[117]	ANFIS	• If-then rule • Membership functions input	• Nonlinear Approximation Capability • High adaptability
[118]	DT ANFIS	• Rule conversion • Vibration and current signal input	• To reduce data volume • To enhance classification performance • Less accurate classification of the current signal
[119]	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
[120]	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
[136]	Feature Incremental Broad Learning	• Reconfiguration system • Simplify the model	• High diagnostic accuracy • Less training time • The feature extraction method is not optimal
[126]	Infrared Thermography ML	• 2-D thermal image input • Optimal eigenvector matrix transfer	• Non-destructive • Non-contact detection • Performance affected by redundant information noise
[127]	Transfer learning	• Feature transfer • High-dimensional mapping of data	• To reduce feature data requirement • Cannot be applied to multi-target working condition
[128]	TFR	• Smooth blurred plane feature extraction	• To reduce computational burden • To improve feature separability
[121]	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
[134]	AAS	• Unsupervised classification • Data Clustering	• To improve monitoring effectiveness • Removal of redundant or irrelevant information
[129] [133]	GA-AIS	• Spatial transformation	• To reduce computational complexity
[135]	Faster AP-MSDL	• Adaptive parameter estimation • Multi-scale learning	• Short running time • Great fault detection performance

dimensional 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 the classification accuracy may not be as high as the discriminative model. In the area of image recognition, the convolutional neural network 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 and short time memory units (LSTM) 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, autoencoders can perform unsupervised learning from data samples, achieve great performance on different data sets 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 system 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 fault detection and diagnosis 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 fault detection and diagnosis 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 paper 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. pre-processing, 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 fault detection and diagnosis 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 paper, common electric motor and fault types are presented. The different fault detection and diagnosis methods have been also analyzed. The mainstream methods of feature extraction are reviewed, 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 machine learning, 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," in IEEE Industrial Electronics Magazine, vol. 8, no. 2, pp. 31-42, June 2014.

[2] S. S. S. R. Sarathbabu Duvvuri, "Rotor Inter-turn Short-circuit Fault Diagnostics Relevant Modeling of Slip-Ring Induction Motor," 2018 IEEE 8th Power India International Conference (PIICON), 2018, pp. 1-6,

[3] R. R. Kumar et al., "Induction Machine Stator Fault Tracking Using the Growing Curvilinear Component Analysis," in 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," in IEEE Transactions on Power Electronics, vol. 34, no. 7, pp. 6669-6684, July 2019.

[5] J. Cai and X. Zhao, "Synthetic Hybrid-Integral-Threshold Logic-Based Position Fault Diagnosis Scheme for SRM Drives," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-8, 2021, Art no. 1500608.

[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," in IEEE Transactions on Industry Applications, vol. 55, no. 3, pp. 2378-2391, May-June 2019.

[7] X. Wang, S. Lu and S. Zhang, "Rotating Angle Estimation for Hybrid Stepper Motors With Application to Bearing Fault Diagnosis," in IEEE Transactions on Instrumentation and Measurement, 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] Chang, Z. Wang and C. S, "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] Motahari-Nezhad, M., & Jafari, S. M. (2021). "Bearing remaining useful life prediction under starved lubricating conditions using time-domain acoustic emission signal processing," Expert Systems with Applications, 168(December 2020), 114391.

[13] Ugras, R. C., Alkan, O. K., Orhan, S., Kutlu, M., & Mugan, A. (2019). "Real-time high cycle fatigue estimation algorithm and load history monitoring for vehicles by the use of frequency-domain methods," Mechanical Systems and Signal Processing, 118, 290–304.

[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] Kenn eth A.Loparo, M. L. Adams and M. Farouk Abdel- Magied, et al, "Fault detection and diagnosis of rotating machinery," IEEE Trans. on Industrial Electronics, pp. 2000, 47(5): 1005- 1014.

[16] Tan, Tommy W. S. Chow and H. Zhou, "HOS - based non-parametric and parametric methodologies for machine fault detection," IEEE Trans. on Industrial Electronics, 2000, 47(5): 1051- 1059.

[17] Benbouzid, Mohamed El Hachemi, "A review of induction motors signature analysis as a medium for faults detection," IEEE Trans. on Industrial Electronics, 2000, 47(5): 984- 993.

[18] P. Priyanga S, K. Krithivasan, P. S and S. Sriram V S, "Detection of Cyberattacks in Industrial Control Systems Using Enhanced Principal Component Analysis and Hypergraph-Based Convolution Neural Network (EPCA-HG-CNN)," IEEE Transactions on Industry Applications, vol. 56, no. 4, pp. 4394-4404, July-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," 2011 IEEE EUROCON - International Conference on Computer as a Tool, 2011, pp. 1-4.

[20] E.M. Tag Eldin, H. R. Emara, E.M. Aboul-Zahab, Sh. S. Refaat, "Monitoring and Diagnosis of External Faults in Three-Phase Induction Motors Using Artificial Neural Network," IEEE.trans, 1- 4244-1298-2007.

[21] S. Premrudee preecacharn, T. Utthiyuung, K. Kruepengkul, P. Puongkaew, "Induction Motor Fault Detection and Diagnosis Using Supervised and Unsupervised Neural Networks," IEEE ICIT'OZ, Bangkok Bangkok, THAILAND, 0-7803-7657.91021 - 2002.

[22] R. Natarajan, "Failure Identification of Induction motors by sensing unbalanced stator currents," IEEE Transactions on Energy Conversion, Vol. 4, No. 4, December 1989.

[23] S. Aksoy, A. Mühürcü, "Recurrent Neural Network Based Nonlinear State Estimation for Induction Motors," Journals of International Review of Electrical Engineering, pp. April 2011 (Vol. 6 N. 2) - Papers Part A.

[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, 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. Q. da Silva, "Neural approach for bearing fault detection in three-phase induction motors," in Proc. 9th IEEE Int. Symp. Diagnostics Electr. Mach., Power Electron. Drives, 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] Unal, M., Onat, M., Demetgul, M., & Kucuk, H, "Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network.," Measurement: Journal of the International Measurement Confederation, pp. 58(January 2021), 187–196., 2014.

[29] D. Yu, J. Cheng, Y. Yang, "Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings," Mech. Syst. Signal Process, 19 (3) (2005) 259–270.

[30] Z.K. Peng, P.W. Tse, F.L. Chu, "An improved Hilbert-Huang transform and its application in vibration signal analysis," J. Sound Vib, 286 (2005) 187–205.

[31] Ahmed, G. K. Singh and S. A. K. S, "Vibration signal analysis using wavelet transform for isolation and identification of electrical faults in induction machine," Electr. Power Syst. Res, vol. 68, pp. 119–136, 2004.

[32] V. Iglesias, J. Grajal, M. A. Sánchez and M. López-Vallejo, "Implementation of a Real-Time Spectrum Analyzer on FPGA Platforms," IEEE Transactions on Instrumentation and Measurement, 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, Dec. 2002.

[34] V.K. Rai, A.R. Mohanty, "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert-Huang transform," Mech. Syst. Signal Process, 21 (8) (2007) 2607–2615.

[35] Y. Qin, S. Qin, Y. Mao, " Research on iterated Hilbert transform and its application in mechanical fault diagnosis," Mech. Syst. Signal Process, 22 (Nov 2008) 1967–1980.

[36] N. Mohanti, "Signal Processing Signal Filtering Detection".New York: Van Nostrand, 1987.

[37] Palos, A. F. Atiya and A. G, "New results on recurrent network training: Unifying the algorithms and accelerating convergence," IEEE Trans. Neural Netw, vol. 13, pp. 765–786, May 2000.

[38] Su, H., & Chong, K. T, "Induction machine condition monitoring using neural network modeling," IEEE Transactions on Industrial Electronics, 54(1), 241–249, 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, 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] Guerreiro, R., Cunha, C., & Teodoro, E, "Machine learning and multiresolution decomposition for embedded applications to detect short-circuit in induction motors," Computers in Industry, 129, 2021.

[42] Hu, Y., Tu, X., Li, F., Li, H., & Meng, G. (2017). "An adaptive and touchless order analysis method based on enhanced empirical wavelet transform for fault detection of bearings with varying speeds," Journal of Sound and Vibration, 409(November 2017), 241–255.

[43] Hu, W., Chang, H., & Gu, X. (2019). "A novel fault diagnosis technique for wind turbine gearbox," Applied Soft Computing Journal, 82, 105556.

[44] Ding, F., Zhang, X., Wu, W., & Wang, Y, "A Fault Feature Extraction Method of Motor Bearing Using Improved LCD," IEEE Access, 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," 2014 IEEE International Conference on Image Processing (ICIP), 2014, pp. 5112-5116.

[46] J.H. Ahn, D.-H. Kwak, and B.-H. 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," 2019 2nd World Symposium on Communication Engineering (WSCE), 2019, pp. 142-146.

[48] Zosso, K. Dragomiretskiy and D, "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/ASMETrans. 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] Huang, Y., Chen, C. H., & Huang, C. J, "Motor fault detection and feature extraction using an rnn-based variational autoencoder," IEEE Access, 7, 139086–139096, 2019.

[54] V. Choqueuse, M. E. H. Benbouzid, Y. Amirat, and S, "Diagnosis of three-phase electrical machines using multidimensional demodulation techniques," IEEE Trans. Ind. Electron, vol. 59, no. 4, pp. 2014–2023, Apr. 2012.

[55] L. A. O Martins, F. L. C. Pádua, and P. E. M. Almeida, "Automatic detection of surface defects on rolled steel using computer vision and artificial neural networks," Proc. IEEE IECON, Nov. 2010, pp. 1081–1086.

[56] G.Duarte, M. I. C.Murguia and S, "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] Prieto, M. D., Cirrincione, G., Espinosa, A. G., Ortega, J. A., & Henao, H, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks," IEEE Transactions on Industrial Electronics, 60(8), 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," in IEEE Transactions on Industrial Electronics, vol. 66, no. 9, pp. 7281-7289, Sept. 2019.

[59] C. Chen, B. Zhang, G. Vachtsevanos and M. Orchard, "Machine Condition Prediction Based on Adaptive Neuro-Fuzzy and High-Order Particle Filtering," in IEEE Transactions on Industrial Electronics, vol. 58, no. 9, pp. 4353-4364, Sept. 2011.

[60] Zeng, X. An and H, "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 Transactions on Instrumentation and Measurement, pp. 69(6), 3325–3333, 2020.

[63] Zhang, X., Wang, B., & Chen, X, "Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine," Knowledge-Based Systems, 89, 56–85., 2015.

[64] Q. Hu, X. Si, A. Qin, Y. Lv and Q. Zhang, "Machinery Fault Diagnosis Scheme Using Redefined Dimensionless Indicators and mRMR Feature Selection," IEEE Access, vol. 8, pp. 40313–40326., 2020.

[65] Zheng, J., Pan, H., & Cheng, J, "Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines," Mechanical Systems and Signal Processing, 85(December 2015), 746–759, 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 Transactions on Dielectrics and Electrical Insulation, pp. 21(1), 33–44, 2014.

[67] Fengqi, W., & Meng, G, "Compound rub malfunctions feature extraction based on full-spectrum cascade analysis and SVM," Mechanical Systems and Signal Processing, 20(8), 2007–2021, 2006.

[68] V. N. Ghate and S. V. Dudul, "Cascade neural-network-based fault classifier for three-phase induction motor," IEEE Transactions on Industrial Electronics, 58(5), 1555–1563, 2011.

[69] S. Zhang, X. Li, M. Zong, X. Zhu and R. Wang, "Efficient kNN classification with different numbers of nearest neighbors," IEEE Transactions on Neural Networks and Learning Systems, 29(5), 1774–1785, 2018.

[70] Recioui, A., Benseghier, B., & Khalfallah, H, "Power system fault detection, classification and location using the K-Nearest Neighbors," 2015 4th International Conference on Electrical Engineering, ICEE 2015., 2016.

[71] Yao, Q., Wang, J., Yang, L., Su, H., & Zhang, G, "A fault diagnosis method of engine rotor based on Random Forests," 2016 IEEE International Conference on Prognostics and Health Management, ICPHM 2016, 0–3. , 2016.

[72] Patel, R. K., & Giri, V. K, "Feature selection and classification of the mechanical fault of an induction motor using random forest classifier," Perspectives in Science, 8, 334–337., 2016.

[73] Seera, M., Lim, C. P., & Loo, C. K, "Motor fault detection and diagnosis using a hybrid FMM-CART model with online learning," Journal of Intelligent Manufacturing, 27(6), 1273–1285, 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 Transactions on Industrial Electronics, 63(11), 7067–7075, 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 Transactions on Industrial Electronics, 63(5), 3137–3147, 2016.

[76] Chen, H., Jiang, B., Zhang, T., & Lu, N, "Data-driven and deep learning-based detection and diagnosis of incipient faults with application to electrical traction systems," Neurocomputing, 396, 429–437, 2020.

[77] S. Shao, R. Yan, Y. Lu, P. Wang and R. X. Gao, "DCNN-Based multi-signal induction motor fault diagnosis," IEEE Transactions on Instrumentation and Measurement, 69(6), 2658–2669, 2020.

[78] Liang, K., Qin, N., Huang, D., Fu, Y., & Hu, L, "Convolutional Recurrent Neural Network for Fault Diagnosis of High-Speed Train Bogie," Complexity, 2018.

[79] Chen, Z., Gryllias, K., & Li, W, "Mechanical fault diagnosis using Convolutional Neural Networks and Extreme Learning Machine," Mechanical Systems and Signal Processing, 133, 106272, 2019.

[80] Shenfield, A., & Howarth, M, "A novel deep learning model for the detection and identification of rolling element bearing faults," Sensors (Switzerland), 20(18), 1–24, 2020.

[81] Zhang, W., Peng, G., Li, C., Chen, Y., & Zhang, Z, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," Sensors (Switzerland), 17(2), 2017.

[82] Zhang, W., Li, C., Peng, G., Chen, Y., & Zhang, Z, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," Mechanical Systems and Signal Processing, 100, 439–453. , 2018.

[83] J. S. L. Senanayaka, H. Van Khang and K. G. Robbersmyr, "Towards Self-Supervised Feature Learning for Online Diagnosis of Multiple Faults in Electric Powertrains," IEEE Transactions on Industrial Informatics, 3203(c), 1–1., 2020.

[84] L. Wen, X. Li, L. Gao and Y. Zhang, "A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method," IEEE Transactions on Industrial Electronics, 65(7), 5990–5998, 2018.

[85] Jalayer, M., Orsenigo, C., & Vercellis, "Fault detection and diagnosis for rotating machinery: A model based on convolutional LSTM, Fast Fourier and continuous wavelet transforms," Computers in Industry, 125, 103378. , 2021.

[86] Zhao, R., Yan, R., Wang, J., & Mao, K, " Learning to monitor machine health with convolutional Bi-directional LSTM networks," Sensors (Switzerland), 17(2), 2017.

[87] Luo, T. Wong and Z, "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] Duan, L. D. Xu and L, "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, 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," IEEE Int. Conf, 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] L. Wen, X. Li, L. Gao and Y. Zhang, "A generative adversarial network-based intelligent fault diagnosis method for rotating machinery under small sample size conditions," IEEE Access, pp. 7, 149736–149749, 2019.

[97] Gan, M., Wang, C., & Zhu, C, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," Mechanical Systems and Signal Processing, pp. 72–73, 92–104, 2016.

[98] Shao, H., Jiang, H., Zhang, H., Duan, W., Liang, T., & Wu, S, "Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing," Mechanical Systems and Signal Processing, 100, 743–765, 2018.

[99] M. Ma, C. Sun and X. Chen, "Discriminative Deep Belief Networks with Ant Colony Optimization for Health Status Assessment of Machine," IEEE Transactions on Instrumentation and Measurement, 66(12), 3115–3125, 2017.

[100] Zhang, T., Wang, W., Ye, H., Huang, D., Zhang, H., & Li, M, "Fault detection for ironmaking process based on stacked denoising autoencoders," Proceedings of the American Control Conference, 2016-July, 3261–3267, 2016.

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

[102] Lee, H., Kim, Y., Kim, C.O, "A deep learning model for robust wafer fault monitoring with sensor measurement noise," IEEE Trans. Semicond. Manuf, 30, 23–31, 2017a.

[103] Jia, F., Lei, Y., Guo, L., Lin, J., & Xing, S, "A neural network constructed by deep learning technique and its application to intelligent fault diagnosis of machines," Neurocomputing, 272, 619–628, 2018.

[104] Mao, W., Feng, W., Liu, Y., Zhang, D., & Liang, X, "A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis," Mechanical Systems and Signal Processing, 150, 107233, 2021.

[105] Jia, F., Lei, Y., Lin, J., Zhou, X., & Lu, N, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," Mechanical Systems and Signal Processing, pp. 72–73, 303–315, 2016.

[106] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," IEEE Transactions on Instrumentation and Measurement, 66(7), 1693–1702., 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, pp. 9, 30154–30168., 2021.

[108] T. Nakazawa and D. V. Kulkarni, "Wafer map defect pattern classification and image retrieval using convolutional neural network," IEEE Transactions on Semiconductor Manufacturing, 31(2), 309–314, 2018.

[109] Sabir, R., Rosato, D., Hartmann, S., & Guehmann, C, "LSTM based bearing fault diagnosis of electrical machines using motor current signal," Proceedings - 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019, 613–618, 2019.

[110] Tan, W., Sun, Y., Qiu, D., An, Y., & Ren, P, "Rolling bearing fault diagnosis based on single gated unite recurrent neural networks," Journal of Physics: Conference Series, 1601(4), 2020.

[111] 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, pp. 7, 149736–149749, 2019.

[112] J. Suwatthikul and S. Sornmuang, "Fault detection and diagnosis of a motor bearing shield," Proceedings of the 6th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems, 2011, pp. 332-335.

[113] H. Abu-Rub, S. Moin Ahmed, A. Iqbal, M. Rahimian and H. A. Toliyat, "Incipient bearing fault diagnostics for inverter fed induction motor drive using ANFIS," The XIX International Conference on Electrical Machines - ICEM 2010, 2010, pp. 1-5.

[114] M. Saeidi, J. Zarei, H. Hassani and A. Zamani, "Bearing fault detection via Park's vector approach based on ANFIS," 2014 International Conference on Mechatronics and Control (ICMC), 2014, pp. 2436-2441

[115] J. Zarei, H. Hassani, Z. Wei and H. R. Karimi, "Broken rotor bars detection via Park's vector approach based on ANFIS," 2014 IEEE 23rd International Symposium on Industrial Electronics (ISIE), 2014, pp. 2422-2426,

[116] Lei, Y., He, Z., Zi, Y., & Hu, Q, "Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs," Mechanical Systems and Signal Processing, pp. 21, 2280–2294, 2007.

[117] Lou, X., & Loparo, K. A, "Bearing fault diagnosis based on wavelet transform and fuzzy inference," Mechanical Systems and Signal Processing, pp. 18, 1077–1095, 2004.

[118] Tran, V. T., Yang, B. S., Oh, M. S., & Tan, A. C. C, "Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference," Expert Systems with Applications, 36(2 PART 1), 1840–1849, 2009.

[119] Q. T. Tran, S. D. Nguyen and T. Seo, "Algorithm for estimating online bearing fault upon the ability to extract meaningful information from big data of intelligent structures," IEEE Transactions on Industrial Electronics, 66(5), 3804–3813, 2019.

[120] J. Sun, C. Yan and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," IEEE Transactions on Instrumentation and Measurement, 67(1), 185–195, 2018.

[121] Aucedo-Dorantes, J. J., Jaen-Cuellar, A. Y., Delgado-Prieto, M., Romero-Troncoso, R. de J., & Osornio-Rios, R. A, "Condition monitoring strategy based on an optimized selection of the high-dimensional set of hybrid features to diagnose and detect multiple and combined faults in an induction motor," Measurement, 178(April), 109404., 2021.

[122] M. Ganchev, B. Kubicek, and H. Kappeler, "Rotor Temperature Monitoring System," The XIX International Conference on Electrical Machines, pp. 1–5, 2010.

[123] S. Stipetic, M. Kovacic, Z. Hanic and M. Vrazic, "Measurement of Excitation Winding Temperature on Synchronous Generator in Rotation Using Infrared Thermography," IEEE Transactions on Industrial Electronics, vol. 59, no. 5, pp. 2288–2298, 2012.

[124] C. Mejuto, M. Mueller, M. Shanel, A. Mebarki, M. Reekie, and D. Staton, "Improved Synchronous Machine Thermal Modelling," in 8th International Conference on Electrical Machines, 2008, pp. 1–6.

[125] Kirchgässner, W., Wallscheid, O., & Böcker, J, "Data-Driven Permanent Magnet Temperature Estimation in Synchronous Motors with Supervised Machine Learning," ArXiv, 8969(c), 1–9, 2020.

[126] A. Choudhary, D. Goyal and S. S. Letha, "Infrared Thermography-Based Fault Diagnosis of Induction Motor Bearings Using Machine Learning," in IEEE Sensors Journal, vol. 21, no. 2, pp. 1727-1734, 15 Jan.15, 2021.

[127] W. Xu, Y. Wan, T. Zuo and X. Sha, "Transfer Learning Based Data Feature Transfer for Fault Diagnosis," IEEE Access, pp. 8, 76120–76129, 2020.

[128] 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 Transactions on Industrial Electronics, 60(9), 4034–4042, 2013.

[129] A. Abid, M. T. Khan and M. S. Khan, "Multidomain Features-Based GA Optimized Artificial Immune System for Bearing Fault Detection," IEEE Transactions on Systems, Man, and Cybernetics: Systems, pp. 50(1), 348–359, 2020.

[130] 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.

[131] 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, 2006, pp. 133–137.

[132] 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, 2010.

[133] 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.

[134] A. Soualhi, G. Clerc and H. Razik, "Detection and diagnosis of faults in induction motor using an improved artificial ant clustering technique," IEEE Transactions on Industrial Electronics, 60(9), 4053–4062, 2013.

[135] Z. Zhang et al., "Faster Multiscale Dictionary Learning Method with Adaptive Parameter Estimation for Fault Diagnosis of Traction Motor Bearings," IEEE Transactions on Instrumentation and Measurement, 70, 2021.

[136] 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, 7, 157796–157806, 2019.

Wangjie Lang was born in Zhejiang province in P.R. China, on December 11, 1998. He received the B.Eng. degree in electrical engineering from the University of Strathclyde and Lanzhou University of Technology. Currently, he is a PHD student with the major of Electrical Engineering in the Department of Electronic Engineering of University of York, Heslington, York, the UK.

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

Yihua Hu (M'13-SM'15) received the B.S. degree in electrical motor drives in 2003, and the Ph.D. degree in power electronics and drives from China University of Mining and Technology, in 2003 and 2011, respectively.

Between 2011 and 2013, he was with the College of Electrical Engineering, Zhejiang University as a Postdoctoral Fellow. Between 2013 and 2015, he worked as a Research Associate at the power electronics and motor drive group, the University of Strathclyde. Currently, he is a Reader at the Department of Electronic Engineering, University of York (UoY). His research interests include renewable generation, power electronics converters & control, electric vehicle, more electric ship/aircraft, smart energy system and non-destructive test technology. He is the associate editor of IET Renewable Power Generation, IET Intelligent Transport Systems and Power Electronics and Drives.

Chao Gong (S'19) was born in Shandong province in P.R. China, on February 22, 1991. He received the B.Eng. and the M.Eng degree in electrical engineering from Northwestern Polytechnical University, Xi'an, China, in 2014 and 2016, respectively. Currently, he is a PHD student with the major of Electrical Engineering in the Department of Electronic Engineering of University of York, Heslington, York, the UK.

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

Xiaotian Zhang was born in Shandong province in P.R. China, on Auegst 24, 1996. He received the BSc degree in applied physics from Hohai University, Nanjing, China, in 2018, and the MSc degree in electronic engineering from Kings College London, London, the UK, in 2020. Currently, he is a PHD student with the major of Electrical Engineering in the Department of Electronic Engineering of University of York, Heslington, York, the UK. His research interests include advanced AI condition monitoring, fault recognization, and engineering application of AI.

Dr Hui Xu received the B.S. degree, MEng and PhD from the China University of Ming and Technology, China, in 2005, 2009 and 2012 respectively. During 2012-2013, she was a postdoc researcher at Zhejiang University. She is currently a research assistant in artificial intelligence and energy with Leeds Beckett University.

Jiamei Deng received the B.S. degree from the Huazhong University of Science and Technology, Wuhan, China, in 1988, the M.S. degree from the Shanghai Institute of Mechanical Engineering, Shanghai, China, and the Ph.D. degree from Reading University, U.K., in 2005. She is currently a Professor in artificial intelligence and energy with Leeds Beckett University with emphasis on control systems design, predictive models, data-driven and power plant safety monitoring.