Applications of machine learning to machine fault diagnosis: A review and roadmap

Yaguo Lei [a,*], Bin Yang [a], Xinwei Jiang [a], Feng Jia [a], Naipeng Li [a], Asoke K. Nandi [b]

[a] Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an 710049, China

[b] Department of Electronic and Computer Engineering, Brunel University London, Uxbridge UB8 3PH, United Kingdom

Abstract

Intelligent fault diagnosis (IFD) refers to applications of machine learning theories to machine fault diagnosis. This is a promising way to release the contribution from human labor and automatically recognize the health states of machines, thus it has attracted much attention in the last two or three decades. Although IFD has achieved a considerable number of successes, a review still leaves a blank space to systematically cover the development of IFD from the cradle to the bloom, and rarely provides potential guidelines for the future development. To bridge the gap, this paper presents a review and roadmap to systematically cover the development of IFD following the progress of machine learning theories and offer a future perspective. In the past, traditional machine learning theories began to weak the contribution of human labor and brought the era of artificial intelligence to machine fault diagnosis. Over the recent years, the advent of deep learning theories has reformed IFD in further releasing the artificial assistance since the 2010s, which encourages to construct an end-to-end diagnosis process. It means to directly bridge the relationship between the increasingly-grown monitoring data and the health states of machines. In the future, transfer learning theories attempt to use the diagnosis knowledge from one or multiple diagnosis tasks to other related ones, which prospectively overcomes the obstacles in applications of IFD to engineering scenarios. Finally, the roadmap of IFD is

* Corresponding author
E-mail: yaguolei@mail.xjtu.edu.cn (Yaguo Lei)

pictured to show potential research trends when combined with the challenges in this field.

***Keywords*:** Machines, intelligent fault diagnosis, machine learning, deep learning, transfer learning, review and roadmap.

Contents

1. Introduction 3
2. Past: IFD using traditional machine learning theories 9
    2.1. Overview 9
    2.2. Step 1: Data collection 10
    2.3. Step 2: Artificial feature extraction 11
        2.3.1. Feature extraction 11
        2.3.2. Feature selection 12
    2.4. Step 3: Health state recognition 13
        2.4.1. Expert system-based approaches 14
        2.4.2. ANN-based approaches 17
        2.4.3. SVM-based approaches 21
        2.4.4. Other approaches 26
    2.5. Epilog 30
3. Present: IFD using deep learning theories 30
    3.1. Overview 31
    3.2. Step 1: Big data collection 31
    3.3. Step 2: Deep learning-based diagnosis 33
        3.3.1. Stacked AE-based approaches 34

3.3.2. DBN-based approaches 37
3.3.3. CNN-based approaches 40
3.3.4. ResNet-based approaches 44
3.4. Epilog 46
4. Future: IFD using transfer learning theories 47
4.1. Motivation of applying transfer learning to IFD in engineering scenarios 47
4.2. Definitions of transfer learning in IFD 48
4.2.1. Transfer problems in IFD 48
4.2.2. Transfer scenarios in IFD 49
4.3. Categories of transfer learning-based approaches in IFD 50
4.3.1. Feature-based approaches 51
4.3.2. GAN-based approaches 57
4.3.3. Instance-based approaches 58
4.3.4. Parameter-based approaches 60
4.4. Epilog 61
5. Discussions: Future challenges and a roadmap in IFD 62
6. Conclusions 67
Acknowledgements 68
References 68

1. Introduction

Fault diagnosis serves an important role in pursuing the relationship between the monitoring data and the health states of machines [1, 2], which has been a widely concerned issue in machine health management. Traditionally, the relationship is caught by abundant experience and huge expert knowledge of engineers. For example, an experienced engineer is able to diagnose the faults of engines depending on the abnormal sound or locate the bearing faults by using advanced signal processing methods to analyze the vibration signals. In engineering scenarios, however, the machine users would like an automatic method to shorten the maintenance cycle and improve the diagnosis accuracy. In particular, with the help of artificial intelligence, the procedure of fault diagnosis is expected to be intelligent enough to automatically detect and recognize the health states of machines [2-5].

Intelligent fault diagnosis (IFD) refers to applications of machine learning theories, such as artificial neural networks (ANN), support vector machine (SVM), and deep neural networks (DNN), to machine fault diagnosis [6, 7], which is promising to achieve the above purpose. This kind of methods uses machine learning theories to adaptively learn the diagnosis knowledge of machines from the collected data instead of utilizing the experience and knowledge of engineers. Specifically, IFD aims to construct diagnosis models that are able to automatically bridge the relationship between the collected data and the health states of machines.

In recent years, IFD has attracted much attention from academic researchers and industrial engineers, which deeply relates to the development of machine learning [6-9]. We count the number of publications about IFD based on the search results from the Web of Science, which is shown in Fig. 1. According to the results on the topic of machine fault diagnosis by using machine learning, the number of publications has rapidly increased since the 2010s. Therefore, we roughly divide the research of IFD into three periods as follows.

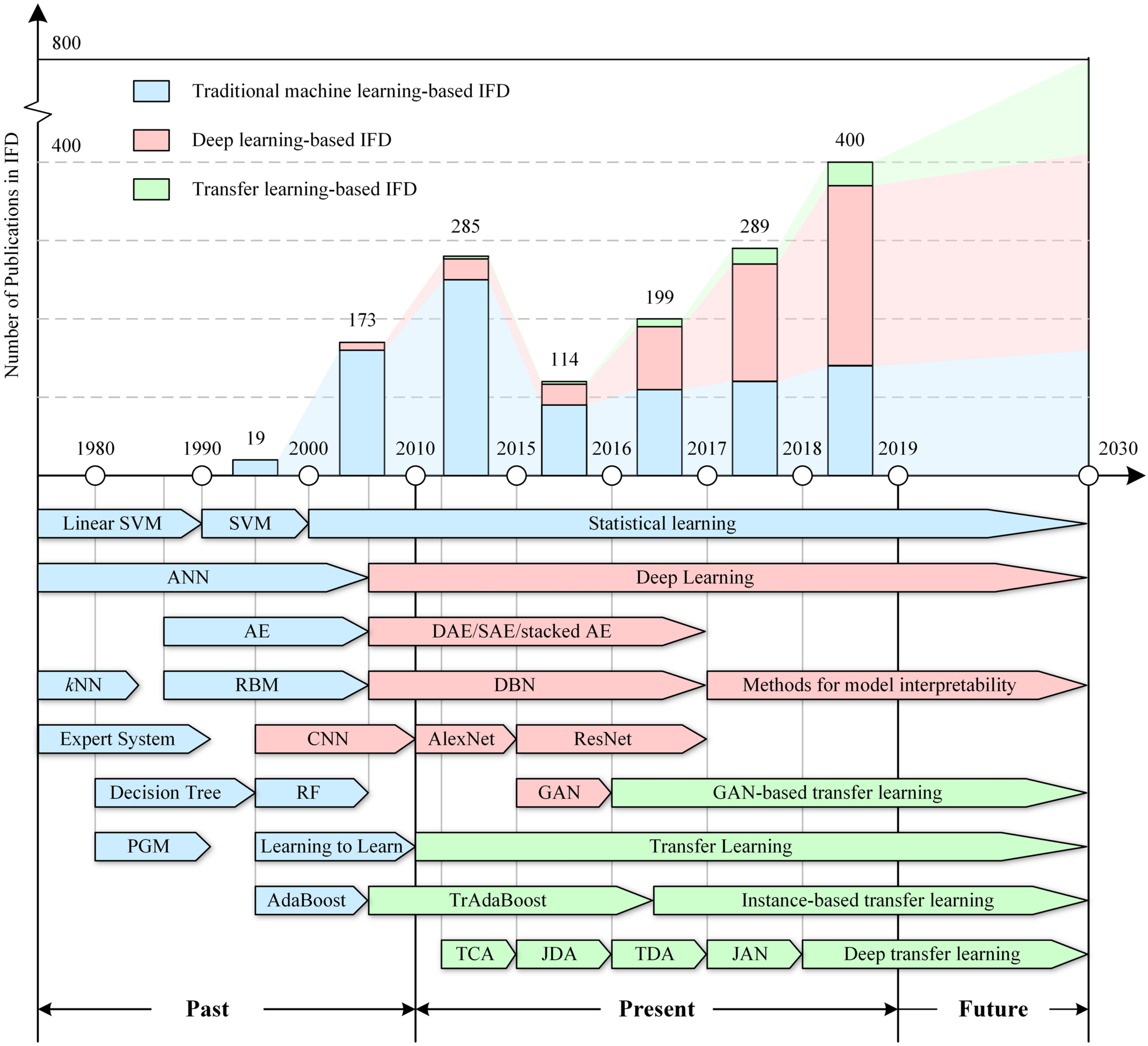


Fig. 1. Development and milestones of IFD using machine learning.

1) In the past, traditional machine learning theories were popularly conducted in IFD from the debut to the 2010s. The early research of machine learning derived back to the 1950s, and boosted to be an important interest of artificial intelligence in the 1980s [10]. A number of traditional theories were invented during this period such as ANN [11], SVM [12], *k*-Nearest Neighbor (*k*NN) [13], and probabilistic graphical model (PGM) [14]. These theories promoted

the emergence of IFD including expert system-based approaches [15], ANN-based approaches [16], SVM-based approaches [17], and other intelligent approaches [7]. In these approaches, the features were artificially extracted from the collected data. After that, the sensitive features were selected to train diagnosis models that could automatically recognize the health states of machines. With the help of traditional machine learning, the diagnosis models began to establish the relationship between the selected features and the health states of machines, which weakened the contribution of human labor in machine fault diagnosis and pushed it into the era of artificial intelligence.

2) In the present, deep learning theories have come to reform IFD since the 2010s [18]. Although IFD in the past is able to recognize the health sates of machines instead of the fault inspection by humans, the artificial feature extraction still mostly depends on the human labor. Furthermore, traditional machine learning theories are not applicable to the increasingly-grown data because of the low generalization performance, which reduces the diagnosis accuracy. Deep learning is a new topic in the field of machine learning, which could date back to the research of neural networks in the 1980s [19, 20], such as autoencoders (AE) [21] and the restricted Boltzmann machine (RBM) [22]. This topic has been widely concerned since 2006 when Hinton [23] used the greedy layer-wise pre-training strategy to train the deep belief network (DBN). In addition, convolutional neural network (CNN) also made a series of breakthroughs, such as AlexNet [24] and ResNet [25]. These theories further inspired the development of IFD and induced a number of achievements [4, 6-9] including stacked AE-based approaches, DBN-based approaches, CNN-based approaches, and ResNet-based approaches. In the approaches, deep learning helps automatically learn fault features from the collected data instead of the artificial feature extraction in the past period of IFD. They attempt to provide end-to-end diagnosis

models when handling the increasingly-grown data. The models are expected to directly connect the raw monitoring data to their corresponding health states of machines, which further releases the contribution of human labor in IFD.

3) In the future, transfer learning theories will promote the research of IFD in engineering scenarios. Although deep learning has achieved significant successes in machine fault diagnosis currently, the successes are subject to a common assumption that there are sufficient labeled data to train diagnosis models. However, such assumption is unpractical in engineering scenarios due to two main reasons [26]. First, machines usually work with the healthy condition, and faults seldom happen. As a result, a considerable number of healthy data are collected, while faulty data are insufficient. Second, it takes huge cost to acquire the machine health states corresponding to the collected data, i.e., labeling the data. Consequently, majority of collected data are unlabeled in engineering scenarios. For the above reasons, the collected data are insufficient to train reliable diagnosis models. Fortunately, transfer learning is concerned to overcome such weaknesses by applying the knowledge learned from one or multiple tasks to other related but novel ones [27]. This theory derives back to 1995 in a different name of the learning to learn [28], and has got some achievements since the 2010s, such as transfer component analysis (TCA) [29], joint distribution adaptation (JDA) [30], and TrAdaboost [31]. After that, it has been developed with the help of deep learning theories since 2015 and yielded some achievements in the field of computer vision [32], such as transfer denoising autoencoder (TDA) [33] and joint adaptation network (JAN) [34]. In the field of IFD, some researchers have begun to develop a few studies generally including feature-based approaches [26], generative adversarial network (GAN) based approaches [35], instance-based approaches [36], and parameter-based approaches [37]. These approaches are expected to provide diagnosis models

that could transfer the diagnosis knowledge learned from one or multiple diagnosis tasks to realize other related but different diagnosis ones. Therefore, transfer-learning theories are expected to overcome the problems of lacking labeled samples and finally enlarge the applications of IFD in engineering scenarios.

To summarize the research of IFD, Liu et al. [7] briefly reviewed the applications of artificial intelligence for fault diagnosis of rotating machines, and mainly focused on applications of traditional machine learning. Khan et al [8] presented categories of the artificial intelligence-based methods in system health management, and further reviewed the applications of deep learning. Duan et al. [6] and Zhao et al. [9] reviewed the commonly-used deep learning theories and the applications to machine fault diagnosis. Hoang et al. [4] provided a review about applications of deep learning to bearing fault diagnosis. However, these reviews have three shortcomings. 1) The previous reviews just concerned IFD in a certain period like using traditional machine learning or using deep learning. For example, Ref. [7] mainly focused on the applications of traditional machine learning, and Ref. [4, 6, 8, 9] just reviewed applications of deep learning to machine fault diagnosis. As a result, a review to systematically cover the development of IFD from the past to the future is still left blank. 2) These reviews have not reported a roadmap on IFD to predict future trends yet. But for review papers, the readers prefer to be interested in potential trends of IFD in the next five or ten years. 3) They just cover the publications of IFD before 2017. In recent years, however, the number of publications has rapidly increased every year. As shown in Fig. 1, the number of publications from 2017 to 2019 almost approaches to the total of that before 2017. Therefore, it needs a new review to summary the current research progress of IFD.

In order to overcome aforementioned shortcomings, this paper systematically reviews the development of IFD and pictures a roadmap for this field. The contributions of this review are refined as follows. 1) The development of IFD is summarized into three periods. In the past, traditional machine learning was expected to lead machine fault diagnosis into the era of artificial intelligence. In the present, deep learning focuses on further enhanced benefits in IFD. In the future, transfer learning is viewed as a future prospect to promote the applications of IFD to engineering scenarios, and we try to cautiously detail them following the issues of "why to transfer", "what is transfer", and "how to transfer". 2) A roadmap of IFD is pictured in this review. The roadmap includes potential research trends and provides valuable guideline for researchers over the future works.

The rest of this review is organized as follows. In Section 2, we focus on the development of IFD in the past including applications of traditional machine learning theories. Section 3 reviews the applications of deep learning theories, which are considered as the present period in the development of IFD. Section 4 argues applications of transfer learning to IFD including the motivation, the definitions, and some exploratory works. In Section 5, we further display a roadmap when combined with the challenges of IFD. Conclusions are enclosed in Section 6.

2. Past: IFD using traditional machine learning theories

This section includes the motivation about applications of traditional machine learning theories, and further reviews IFD in the past according to a commonly-implemented diagnosis process including data collection, artificial feature extraction, and health state recognition.

2.1. Overview

Traditionally, the process of fault diagnosis is mostly developed by manually inspecting the health states of machines, which increases the labor intensity and reduces the diagnosis accuracy. Advanced signal processing methods [38-40] enable to help ensure which types of faults or where the faults happened in machines. However, these methods greatly rely on the specialized knowledge that maintainers mostly lack in engineering scenarios. Furthermore, the diagnosis results by signal processing methods are too specialized to be understood by the machine users. Therefore, modern industrial applications prefer the fault diagnosis methods that could automatically recognize the health states of machines.

With the help of machine learning theories, IFD is expected to achieve the above purpose [1, 5]. In the past period of IFD, some traditional machine learning theories, such as ANN and SVM, are applied to machine fault diagnosis. The diagnosis process includes three steps [1], i.e., data collection, artificial feature extraction, and health state recognition, as shown in Fig. 2. Each step will be detailed in the following subsections.

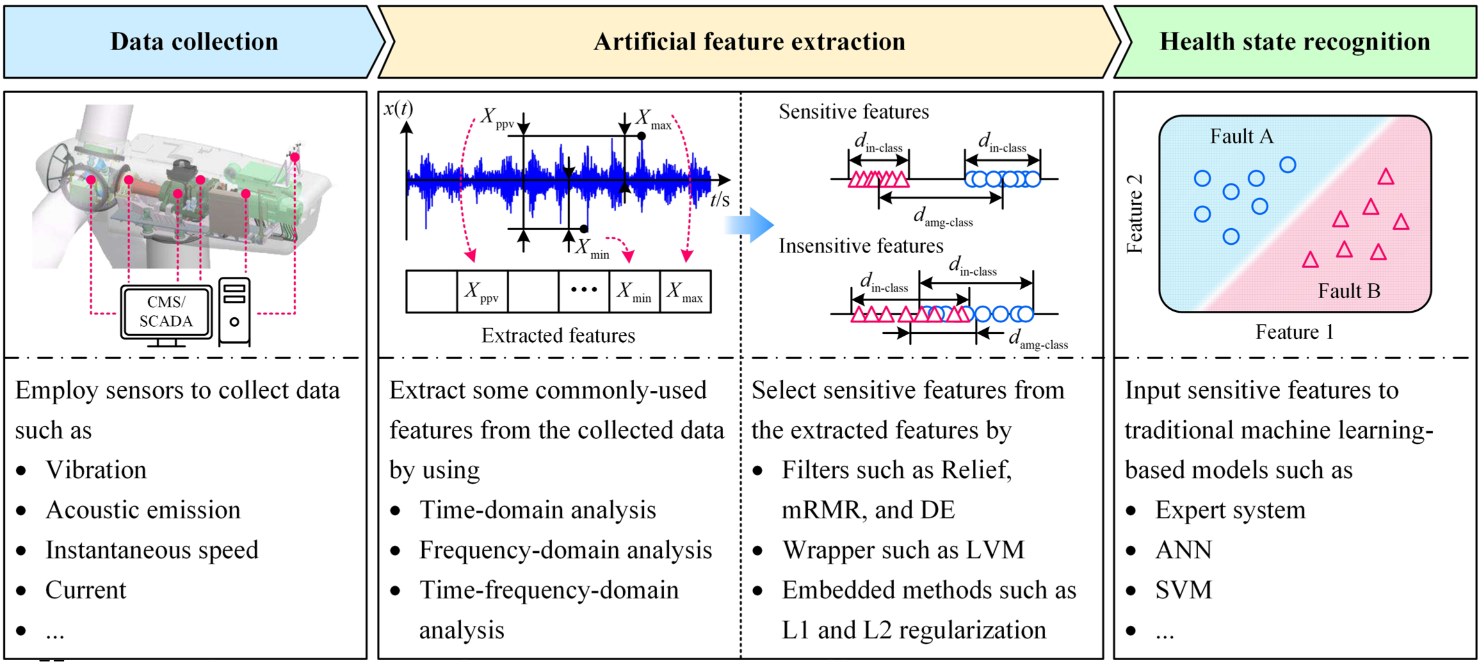


Fig. 2. Diagnosis process of IFD using traditional machine learning theories.

2.2. Step 1: Data collection

In the step of data collection, the sensors are mounted on machines to constantly collect data. Different sensors are usually employed, such as vibration, acoustic emission, temperature, and current transformer. Among them, vibration data are widely used in fault diagnosis of bearings [4, 41] and gearboxes [42, 43]. Acoustic emission data are potential to detect the incipient faults and the deformation of bearings [44, 45] and gears [46-48], especially under the low-speed operation conditions and low-frequency-noise environment. Instantaneous speed data are commonly used in fault diagnosis of engines [49-51], which are strongly anti-interference. Current data play an important role in fault diagnosis of electric-driven machines [52-54]. This kind of data is easily collected just by using the current transformer, and does not include into the running of machines. In addition, researchers discovered that the data from multi-source sensors have complementary information, which could be fused to achieve higher diagnosis accuracy than just using data from the individual sensor [55].

2.3. Step 2: Artificial feature extraction

Artificial feature extraction includes two steps. First, some commonly-used features, such as time-domain features, frequency-domain features, and time-frequency-domain features, are extracted from the collected data. These features contain the health information reflecting the health states of machines. Second, feature selection methods, such as filters, wrappers, and embedded methods, are used to select sensitive features to health states of machines from the extracted features. It is beneficial to getting rid of the redundant information and further improving the diagnosis results. These two steps are detailed as follows.

2.3.1. Feature extraction

The commonly-used features can be extracted from the time domain, the frequency domain analysis, or the time-frequency domain. 1) The time-domain features can be divided into the dimensional ones and the dimensionless ones. The former includes mean, standard deviation, root amplitude, root mean square, peak value, etc., which are affected by the speed and the load of machines. The later mostly includes shape indicator, skewness, kurtosis, crest indicator, clearance indicator, impulse indicator, etc., which are robust to the operation conditions of machines [56, 57]. 2) The frequency-domain features are extracted from the frequency spectrum, such as mean frequency, frequency center, root mean square frequency, standard deviation frequency, etc., which are introduced in Ref. [56, 57]. They contain the information that cannot found in the time-domain features. 3) The time-frequency-domain features, such as energy entropy [56, 57], are usually extracted by wavelet transform (WT), wavelet package transform (WPT) or empirical model decompose (EMD). These features are able to reflect health states of machines under non-stationary operation conditions.

2.3.2. Feature selection

The extracted features from the time domain, the frequency domain, and the time-frequency domain contain the redundant information. They may aggravate the computation cost, and even result in the curse of dimensionality. To weaken this problem, some publications [58-65] select sensitive features to the health states of machines from the collected features. They can be divided into three categories, i.e., filters, wrappers, and embedded methods.

(1) Filter-based methods

Filters directly preprocess the collected features, which are independent to the training of the classifier [58]. Some filters are briefly introduced as follows. 1) Relief [66] and Relief-F [67] construct a relevant indicator to determine the sensitivity of features to the health states of machines. 2) Information gain and gain ratio [68], from the information theory, are also two commonly-used methods for feature selection. The features with the greater information gain and the higher gain ratio would be selected to train diagnosis models and improve their diagnosis results. 3) Minimum Redundancy Maximum Relevance (mRMR) [69] attempts to select features subject to the maximal dissimilarity with each other. 4) Fisher score [70] is regarded as a distance metric to feature selection, which aims to select a feature that is able to maximize the among-class distance, but minimize the in-class distance. 5) Distance evaluation (DE) [71, 72] is used to select a feature set by the distance metric, in which the sensitive features are subject to the small in-class distance and the large among-class distance.

(2) Wrapper-based methods

Different from filter-based methods, wrappers focus on the interaction of feature selection with training classifiers [58]. In other words, the performance of classifiers is used to assess the selected feature set. If the selected feature subset cannot produce the optimal classification accuracy, another subset is reselected in the next iteration until the selected features enforce the classifiers at the most favorable performance. Las Vegas wrapper (LVW) [73] is widely used to select the features, in which the Las Vegas algorithm is employed to search for the feature subset, and the error of the classifiers is considered as the metric to feature assessment.

(3) Embedded methods

Embedded methods integrate the feature selection into training the classifiers. Generally, they impose the regularization terms on the optimization objects of the classifiers, and automatically select the features once the training of classifiers is done [58]. Two regularization terms are commonly considered. One is the L1 regularization [74], and the other one is the L2 regularization [75]. Both them can alleviate the over fitting that occurs in training with a small amount of training samples. In contrast, the L1 term prefers to obtain the sparse parameters, which is able to abandon redundant features in classification and further enforce the classifier to achieve the high classification performance.

2.4. Step 3: Health state recognition

Health state recognition uses machine learning-based diagnosis models to establish the relationship between the selected features and the health states of machines. To achieve the purpose, the diagnosis models are first trained with labeled samples. After that, the models are able to recognize the health states of machines when the input samples are unlabeled. According to the research popularity, we will briefly introduce four IFD approaches using traditional machine learning in the following subsections.

2.4.1. Expert system-based approaches

(1) A brief introduction to expert system

The expert system is viewed as a method that can provide the expert-level diagnosis knowledge to solve the diagnosis tasks of machines instead of huge human labor. The expert system-based diagnosis models, as shown in Fig. 3, consist of five parts, i.e., the knowledge base, the database, the inference engine, the user interface, and the explanation system [76]. Each part is briefly described as follows.

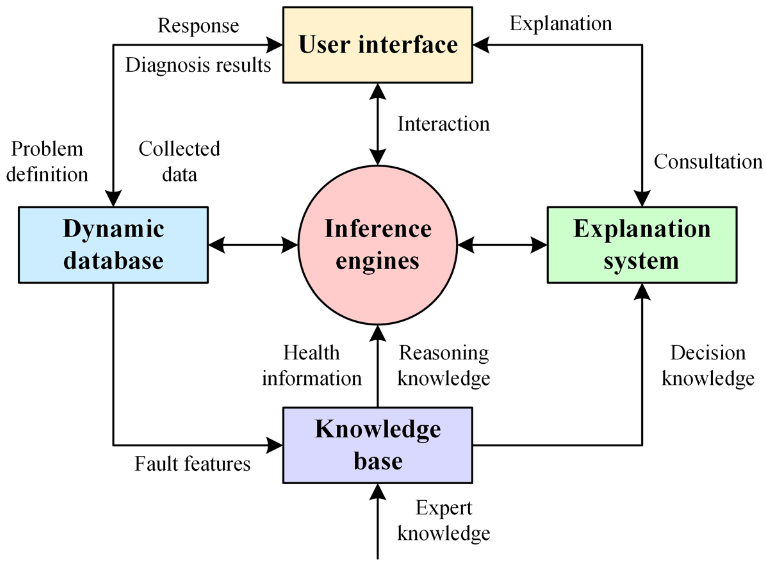


Fig. 3. Architecture of expert system-based diagnosis models.

- A dynamic dataset collects the data that are generated in each parts of solving the diagnosis tasks, and serves as the memory for the operation of inference engines.
- A knowledge base contains the expert knowledge about the diagnosis task. Moreover, it

further includes the fault features to provide the health information for inference engines.

- An inference engine uses the input health information and the reasoning knowledge (the designed rules and strategies) from the knowledge base to interact with the dynamic dataset and the explanation system, and further infers the diagnosis results.
- A user interface is a function-integrated interface, in which the users can interact with the system about the data transmission, the parameter configuration, the result acquisition, and the problem definition and consultation.
- An explanation system makes response to the consultation of the user inference about the inference process, and further explains the reason why the expert system makes a given diagnosis decision.

(2) Applications of expert system to machine fault diagnosis

According to different inference engines, expert system-based diagnosis models can be divided into four categories, i.e. rule-based reasoning, fuzzy logic-based reasoning, neural network-based reasoning, and case-based reasoning. Each part is reviewed as follows. 1) The rule-based reasoning is used to manipulate diagnosis knowledge and then make decision by designed rules [15, 77]. In the field of IFD, Krishnamurthi et al. [78] designed a rule-based reasoning expert system for a Cincinnati Milacron 786 robot, which was one of the earliest research in this field. The designed diagnosis shell really reduced the development time and effort of diagnosis models in knowledge acquisition, application system generation, learning, explanation. Gelgele et al. [79] employed IF-THEN rule to construct an expert system-based diagnosis model for automotive engines. Furthermore, the rule-based reasoning was used for fault diagnosis of hydraulic systems [80], rolling element bearings [81], and centrifugal pumps [82]. Although the rule-based reasoning can establish the nonlinear mapping from the elected features to the health states, the reasoning efficiency drops with the increasing number of designed rules for sophisticated machines. 2) The fuzzy logic-based reasoning introduces fuzzy set theory into inference engine to describe the imprecise and non-numerical information [15, 77]. In IFD, Lee et al. [83] designed a fuzzy reasoning system for power systems, which was one of the earliest work about the applications of fuzzy logic-based reasoning. The proposed system included four parts, i.e., Meta inference system, expert system for hybrid diagnosis, expert system for the diagnosis of substations, and expert system for the diagnosis of transmission network, which improved efficiency and reliability of the fault diagnosis process. Liu et al. [84] used fuzzy multi-attribute group decision making method to construct expert system-based diagnosis models. Wu et al. [85] employed the fuzzy-logic inference to recognize the health states of scooter engines. Berredjem et al. [86] applied fuzzy expert system to fault diagnosis of bearings and achieved high diagnosis accuracy. The performance of fuzzy logic-based reasoning is related to the fuzzy dataset, but it is difficult to be captured. Therefore, such reasoning commonly has low learning ability, which may reduce the diagnosis accuracy. 3) The neural network-based reasoning inherits the capabilities of learning, association, and memory of neural networks [15, 77]. Wu et al. [87, 88] respectively used the probability neural network and the generalized regression neural network to construct expert system-based diagnosis models for internal combustion engines. Hajnayeb et al. [89] employed multi-layer perceptron neural network to construct the inference engine and infer the relationship between the collected data and the health states of bearings. Jayaswal et al. [90] combined ANN and fuzzy rules to construct expert system-base diagnosis models for bearings. The neural network-based reasoning needs to acquire diagnosis knowledge from sufficient training data, which is difficult to be met in engineering scenarios. In addition, such reasoning cannot clearly explain the reasoning process and the physical meaning of the saved knowledge due to the black box of the neural network. 4) The case-based reasoning attempts to solve specialized problems according to the solutions of similar existing problems [15]. Vingerhoeds et al. [91] used the case-based reasoning to incorporate the knowledge and experience from both train manufacturers and railway companies for on-line fault diagnosis. Varma et al. [92] presented a case-based reasoning system for fault diagnosis of locomotive by using on-board fault messages. Wu et al. [93] developed an expert system for fault diagnosis of modern commercial aircraft, which was designed by the case-based reasoning and the fuzzy logic. Vong et al. [94] constructed a computer-aided diagnosis system based on the case-based reasoning and the kernel *k*-means for the automotive engine ignition system.

The expert system-based diagnosis models represent the diagnosis knowledge from experts as the inference algorithm to automatically recognize the health states of machines. However, the performance of the diagnosis models greatly relies on the expert knowledge, which is difficult to be acquired and expressed. The incorrect and incomplete knowledge may reduce the diagnosis accuracy. Furthermore, the expert system lacks the self-learning capability so that the diagnosis knowledge base is difficult to be expanded and corrected.

2.4.2. ANN-based approaches

ANN imitates the activities of human brains in information processing, which is an effective way to establish the diagnosis models. This section reviews the applications of ANN to fault diagnosis of machines.

(1) A brief introduction to ANN

Back propagation neural network (BPNN) is a multilayer perceptron by supervised learning, which consists of the forward propagation and the back propagation. In the forward propagation, as shown in Fig. 4, the input samples are processed by multi-hidden layers, and they are finally mapped into the target class in the output layer. Given the training dataset $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^m$ with m samples, where $\boldsymbol{x}_i \in \Re^d$ includes d features and $\boldsymbol{y}_i \in \Re^l$ includes l health states, the output of the hth hidden layer is expressed as

$$\left(\boldsymbol{x}_i^h\right)_j = \sigma^h\left(\sum_{i=1}^{n_{h-1}} \boldsymbol{\omega}_j^h \cdot \boldsymbol{x}_i^{h-1} + b_j^h\right), \qquad j = 1,2,\cdots,n_h, \qquad h = 1,2,\cdots,H, \tag{1}$$

where $\left(\boldsymbol{x}_i^h\right)_j$ is the output of the jth neuron in the hth hidden layer, and $\boldsymbol{x}_i^0 = \boldsymbol{x}_i$, n_h is the number of neurons in the hth hidden layer, σ^h represents the activation function of the hth hidden layer, n_{h-1} is the number of neurons in the $(h-1)$th hidden layer, $\boldsymbol{\omega}_j^h$ is the weights between the neurons in the previous layer and the jth neuron in the hth hidden layer, and b_j^h is the bias of the hth hidden layer. The predicted output of BPNN is

$$(\hat{\boldsymbol{y}}_i)_k = \sigma^{\text{out}}\left(\sum_{i=1}^{n_H} \boldsymbol{\omega}_j^{\text{out}} \cdot \boldsymbol{x}_i^H + b_j^{\text{out}}\right), \qquad k = 1,2,\cdots,l, \tag{2}$$

where $(\hat{\boldsymbol{y}}_i)_k$ is the predicted output of the kth neuron in the output layer, σ^{out} is the activation function of the output layer, $\boldsymbol{\omega}_j^{\text{out}}$ and b_j^{out} are respectively the weights and bias of the output layer. When given a certain training sample $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}$, the optimization objective of BPNN aims to minimize the error between the predicted output and the target one by

$$\min_{\boldsymbol{\omega},\boldsymbol{b}} \qquad E_i = \frac{1}{2}\sum_{k=1}^{l}[(\boldsymbol{y}_i)_k - (\hat{\boldsymbol{y}})_k]^2 \tag{3}$$

In order to solve this problem, the training parameters $\boldsymbol{\omega}$ and $\boldsymbol{b}$ are updated by the gradient descent as follows.

$$\boldsymbol{\omega} \leftarrow \boldsymbol{\omega} - \eta \cdot \frac{\partial E_i}{\partial \boldsymbol{\omega}}, \qquad \boldsymbol{b} \leftarrow \boldsymbol{b} - \eta \cdot \frac{\partial E_i}{\partial \boldsymbol{b}} \tag{4}$$

where η is the learning rate. The error gradient propagates backward from the output layer to the input layer, and updates the training parameters layer by layer. Ref. [21] provides more detail about the back propagation algorithm.

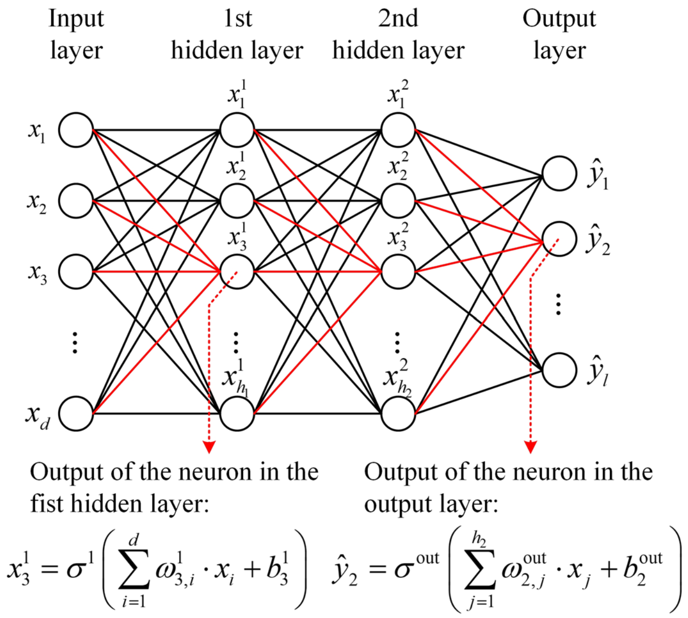


Fig. 4. Architecture of BPNN with two hidden layers.

(2) Applications of ANN to machine fault diagnosis

The publications about applications of ANN to fault diagnosis are listed in Table 1, which are divided into five categories in terms of diagnosis objects like rolling element bearings, gears, motors, engines. In terms of methodologies, BPNN, the radial basis function network (RBFN), and the wavelet neural network (WNN) are widely used to complete the diagnosis tasks. Some researchers further investigated the varieties of ANN for fault diagnosis of machines. Merainani et al [95] used the self-organizing feature map neural network to identify the health states of automatic gearboxes under different operation modes. Wong et al [96] proposed the modified self-organizing map for fault diagnosis of bearings. Yang et al [71] constructed a diagnosis model using the Kohonen neural network with adaptive resonance theory for the rotor system, which obtained higher diagnosis accuracy than the conventional RBFN. Chen et al. [97] employed the probabilistic neural network for efficiently fault diagnosis of hydraulic generator units. Zhong et al. [98] proposed a hierarchical ANN for fault diagnosis of the rotor system, which divided the label space into several subspaces and recognized multiple faults. Barakat et al. [99, 100] introduced the growing neural network to construct a diagnosis model for motor bearings, which obtained the higher diagnosis accuracy for a large number of data when compared with the conventional RBFN and the probabilistic neural network.

Thanks to the high self-learning capability, ANN-based diagnosis models could automatically learn diagnosis knowledge from the input data by minimizing the empirical risk. Furthermore, they can easily recognize multiple states of machines. However, there are two disadvantages. First, the complexity of the diagnosis models would greatly enhance with the increase of input monitoring data. The increasing model parameters lower the training efficiency and further result in the over fitting, which reduces the diagnosis accuracy of the diagnosis models. Second, the ANN-based diagnosis models are black-boxed due to the lack of rigorous theoretical supports. As a result, they are subject to the low interpretability.

Table 1 Summary of applications of ANN to machine fault diagnosis.

Objects	References	Methodologies
Bearings	Yang et al. [101], Samanta et al. [102], Yu et al. [103], Castejon et al. [104], Muruganatham et al. [105], Unal et al. [106], Zarei et al. [107], Almeida et al. [108], and Ahmed et al. [109]	BPNN
	Wang et al. [110], Lei et al. [111], Vijay et al. [112], Jiang et al. [113], and Tang et al. [114]	RBFN
	Lei et al. [115], Wu et al. [116]	WNN
Gears	Abu-Mahfouz et al. [117], Rafiee et al. [118], Hajnayeb et al. [119], Cerrada et al. [120], Kane et al. [121], Waqar et al. [122], and Tyagi et al. [123]	BPNN
	Lai et al. [124], Li et al. [125], and Liu et al. [126]	RBFN
	Chen et al. [127]	WNN

Motors	Ayhan et al. [128], Sadeghian et al. [129], Arabaci et al. [130], Cabal-Yepez et al. [131], Hernandez-Vargas et al. [132], and Moosavi et al. [133]	BPNN
	Ghate et al. [134], and Palacios et al. [135]	RBFN
	Boukra et al. [136]	WNN
Engines	Sharkey et al. [137], Lu et al. [138], Chen et al. [139, 140], Khazaee et al. [141, 142], and Zabihi-Hersari et al. [143]	BPNN
	Wu et al. [144, 145]	RBFN
	Shen et al. [146], Zhang et al. [147]	WNN
Others	Kuo et al. [148], Ilott et al. [149], Wu et al. [150], Mohammed et al. [151], Walker et al. [152], Malik et al. [153], and McCormick et al. [154, 155]	BPNN
	Wu et al. [156], and Villanueva et al. [157]	RBFN
	Liu et al. [158], Chen et al. [159], Guo et al. [160], Xiao et al. [161], and Jin et al. [162]	WNN

2.4.3. SVM-based approaches

SVM is a supervised learning method, which is widely concerned in classification tasks. We briefly review the theory of SVM and summarize its applications to machine fault diagnosis in this section.

(1) A brief introduction to SVM

Given the dataset $\{\boldsymbol{x}_i, y_i\}_{i=1}^{m}$ with m samples and $y_i \in \{-1, 1\}$, a hyperplane $f(x) = 0$ is expected to be found to separate the given datasets into two classes, which is shown as

$$f(\boldsymbol{x}) = \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x} + b = \sum_{i=1}^{m} \boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b = 0 \tag{5}$$

where $\boldsymbol{\omega}$ and b are parameters to determine the hyperplane. In order to separate the samples into the positive class and the negative class, the created hyperplane is subject to

$$y_i f(\boldsymbol{x}_i) = y_i(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1, \qquad i = 1,2,\cdots,m. \tag{6}$$

As shown in Fig. 5, the support vectors H_1 and H_2 can satisfy the constraints in Eq. (6). The linear SVM is expected to place a hyperplane H^* between the positive and negative datasets, which is orientated by maximizing the margin $\gamma = 2/\|\boldsymbol{\omega}\|$. Therefore, the optimization objective of the linear SVM is shown as follows [12].

$$\begin{aligned} &\min_{\boldsymbol{\omega},b} \quad \frac{1}{2}\|\boldsymbol{\omega}\|^2 \\ &\text{s.t.} \quad y_i(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1, \qquad i = 1,2,\cdots,m. \end{aligned} \tag{7}$$

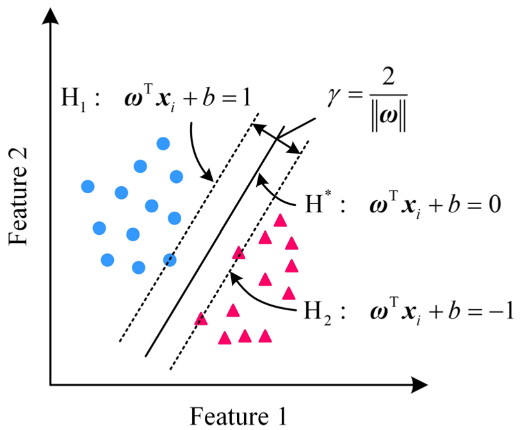


Fig. 5. Classification by the linear SVM.

(2) Applications of SVM to machine fault diagnosis

The applications of SVM to IFD are carefully summarized in Table 2. According to the results, SVM serves as a widely-used machine learning method in health state recognition, especially for fault diagnosis of rolling element bearings, gears, motors, engines, rotor systems [163-168], and hydraulic equipment [169-171]. For these diagnosis objects, SVM-based diagnosis models are expected to recognize multiple states but not just the binary states of health and faults. Thus, the one-against-all strategy (OAA) and one-against-one strategy (OAO) are mainly concerned [17]. Platt et al. [172] and Hsu et al. [173] compared the performance of OAA with OAO, and provided valuable suggestion in selecting prior strategy to obtain better diagnosis accuracy. After that, some publications further introduced some advanced multi-class strategies in applications of SVM, such as the direct acyclic graph [174-176] and the binary tree [164-166, 177-185], which effectively overcame the weaknesses of OAA and OAO. In order to improve the diagnosis accuracy of SVM-based models, researchers mainly focused on two branches, i.e., the modified SVM and the algorithm optimization. For the former, they applied the modified SVM to machine fault diagnosis, such as the least square SVM [63, 165, 186-191], the proximal SVM [192-194], the one-class SVM [195], the hyper-sphere-structured SVM [196], the wavelet SVM [182, 189, 197-200], the ensemble SVM [201, 202], the fuzzy SVM [203, 204], the multi-kernel SVM [205, 206], and the relevance vector machine [207], which achieved better diagnosis performance than the conventionally SVM-based approaches. In addition, the algorithm optimization is concerned to improve the complex solution and simplify the parameter selection of SVM. To achieve this purpose, some researchers introduced the optimization algorithms, such as the kernel Adatron algorithm [208, 209], the sequential minimal optimization [210-214], the genetic algorithm [209, 215-218], the particle swarm optimization [167, 205, 206, 219-222], and the ant colony optimization [168, 223].

Different from the ANN, SVM-based diagnosis models are trained by minimizing the structural risk, which is beneficial to improving the interpretability of the models due to the rigorous theories. The optimization objective solution of SVM refers to the convex quadratic optimization so that the diagnosis models could easily obtain the global optimal solution and further get the high diagnosis accuracy. Three disadvantages of SVM-based diagnosis models need to be considered. First, such diagnosis models are effective to handle the small number of monitoring data. However, they are difficult to fit the massive data, which may results in the curse of computation. Second, the performance of SVM-based diagnosis models is sensitive to the kernel parameters. The inappropriate kernel parameter even cannot induce the reliable diagnosis result. Third, the SVM algorithm is originally used to solve binary classification tasks. In terms of multi-class classification tasks in IFD, it always needs to use complicated architectures, such as OAA and OAO, to integrate results from multiple SVM-based models.

Table 2 Summary of applications of SVM to machine fault diagnosis.

Objects	References	Methodologies
Bearings	Abbasion et al. [224], Yang et al. [225], Xian et al. [226], Hao et al. [227], Gryllias et al. [228], Islam et al. [229], Sugumaran et al. [192], Zheng et al. [202], Jack et al. [208], Rojas et al. [210], HungLinh et al. [230], Kang et al. [216], and Li et al. [222]	OAA-based SVM
	Yang et al. [231], Yang et al. [232], Wu et al. [233, 234], Saidi et al. [235], Zhu et al. [236], Ziani et al. [237], Islam et al. [238], Widodo et al. [211], Zhang et al. [239], and Zhu et al. [219]	OAO-based SVM
	Sugumaran et al. [192, 195], Wang et al. [196], Dong et al. [197], Zhang et al. [201], Li et al. [177-179], Xu et al. [186], Zheng et al. [202], and Chen et al. [206]	Varieties of SVM
	Jack et al. [208, 209], Samanta et al. [215], Rojas et al. [210], Widodo et al. [211], Li et al. [223], Zhang et al. [239], Chen et al. [206], Zhu et al. [219], Dong et al. [221], HungLinh et al. [230], Kang et al. [216], Su et al. [220], Zhu et al. [217], and Li et al. [222]	SVM with optimization algorithm
Gears	Liu et al. [240], and Li et al. [241]	OAA-based SVM
	Lu et al. [242], Cheng et al. [243], Xing et al. [244], Liu et al. [245], Shen et al. [246], Jiang et al. [188], Heidari et al. [198], and Bordoloi et al. [247]	OAO-based SVM
	Saravanan et al. [193], Shen et al. [246], Tang et al. [182], Jiang et al. [188],	Varieties of SVM

	Yang et al. [180], Heidari et al. [189, 198], Jiang et al. [190], Li et al. [63, 181, 187], and Chen et al. [200]	
	Samanta et al. [218], Li et al. [212], Chen et al. [200], Bordoloi et al. [247], and Zhang et al. [248]	SVM with optimization algorithm
Motors	Widodo et al. [213, 249], Ebrahimi et al. [250], Shahriar et al. [251], Kang et al. [252], Singh et al. [64], and Ebrahimi et al. [203]	OAA-based SVM
	Kurek et al. [253], Gangsar et al. [254-256], Sun et al. [257], Martinez-Morales et al. [258], Keskes et al. [199], and Widodo et al. [214]	OAO-based SVM
	Tsoumas et al. [259], Bacha et al. [174], Salem et al. [184], Kang et al. [183], Keskes et al. [176, 199], Ebrahimi et al. [203], and Zgarni et al. [175]	Varieties of SVM
	Widodo et al. [213, 214]	Optimization
Engines	Li et al. [260], and Zhang et al. [51]	OAA-based SVM
	Lee et al. [261], Wang et al. [262], Liu et al. [207], and Jafarian et al. [263]	OAO-based SVM
	Vong et al. [191], Jena et al. [264], Liu et al. [207], Cai et al. [185], and Li et al. [205]	Varieties of SVM
	Li et al. [205]	SVM with Optimization algorithm
Others	Namdari et al. [265], and Jegadeeshwaran et al. [169]	OAA-based SVM
	Rapur et al. [170, 171], Pang et al. [163], Hang et al. [204], Tang et al. [167], and Zhang et al. [168]	OAO-based SVM
	Chiang et al. [194], Yuan et al. [166], and Jin et al. [165], Tang et al. [182],	Varieties of SVM

Wang et al. [164], Hang et al. [204]	
Tang et al. [167], and Zhang et al. [168]	SVM with optimization algorithm

2.4.4. Other approaches

In addition to the approaches in the aforementioned sections, other methods are also widely concerned in IFD, such as, *k*NN, PGM, and the decision tree. We will review them in this section.

(1) *k*NN

The *k*NN is a commonly-used supervised learning model to complete the classification tasks [13]. In this method, a distance metric is used to search for k samples near a given unlabeled sample. As shown in Fig. 6, the majority label of the k samples will be assign to the unlabeled sample as its predicted result. The *k*NN has been concerned in the research of IFD, especially for fault diagnosis of rolling element bearings [266-273], gears [274-277], and motors [278]. However, the performance of *k*NN is subject to some problems, such as the indistinguishable neighborhood boundary and the difficulty in selecting the optimal neighborhood parameter. Some researchers investigated the modified *k*NN and used them for machine fault diagnosis. Lei et al. [279] integrated a set of weighted *k*NN for fault diagnosis of bearings, in which the extracted features were weighted in training the *k*NN-based diagnosis models according to the sensitivity of features to the health states of machines. Similarly, Zhao et al. [280] proposed the Euclidean weighted *k*NN classifiers for fault diagnosis of bearings, and the Euclidean distance was used to weight the extracted features to highlight the sensitivity of them to classification. Li et al. [281] introduced the optimized evidence-theoretic *k*NN classifier for fault diagnosis of bearings, which improved the diagnosis accuracy and robustness of the original version. Dong et al. [282, 283] optimized the *k*NN by the particle swarm optimization algorithm, and obtained higher diagnosis accuracy for bearings than other methods without optimization. Pandya et al. [45] proposed a modified *k*NN algorithm based on asymmetric proximity function to improve the diagnosis accuracy of bearings.

The *k*NN-based diagnosis models are easily-implemented. However, it takes much computation cost to handle the large-volume dataset. In particular, the imbalanced distribution of the collected data would reduce the diagnosis accuracy of this kind of diagnosis models. Furthermore, the parameter k is difficult to be determined, which greatly affects the performance of the diagnosis models.

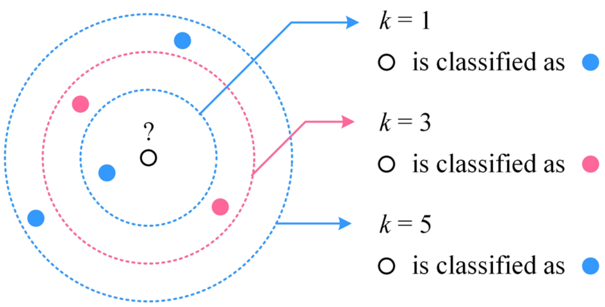


Fig. 6. Illustration of the kNN algorithm.

(2) PGM

PGM serves as a probabilistic model to express the relationship between the variables by graphic architectures. In the models, the notes represent a set of random variables, and the links between the notes mean the probabilistic relationship among the variables, as shown in Fig. 7. PGM can be divided into two classes, i.e. Bayesian classifiers and Markov models. In terms of the former, Yuan et al. [284] and Yu et al. [285] used the naive Bayesian classifier to respectively recognize the health states of rolling element bearings and gears. In order to obtain higher diagnosis accuracy compared with the conventional naive Bayesian classifier, Yu et al.

[286, 287] further employed the normal naive Bayesian classifier and the flexible naive Bayesian classifier for fault diagnosis of gears. It is noted that the applications of naive Bayesian classifier are subject to an assumption of independence among features. The non-naive Bayesian classifier [288] is further developed to release the assumption, and it has also been used for fault diagnosis of bearings [289] and gears [290]. Furthermore, the hidden Markov models are considered as classifiers in the health state recognition of bearings [291-295], the synchronous motors [296], and the hydraulic pumps [297]. Some publications further improved the hidden Markov models. Xiao et al. [298] presented a diagnosis model based on the coupled hidden Markov models for bearing fault diagnosis, which was beneficial to fusing multichannel information by using the multiple state sequences and observation sequences. Huang et al. [299] used predictive neural network and intuitionistic fuzzy sets to determine the observation matrix of hidden Markov models, which improved the diagnosis accuracy of the motor drive system.

It is easy for PGM-based diagnosis models to achieve fault diagnosis with multiple health states of machines. They can be used to analyze the among-class discrepancy in convenience. However, this kind of diagnosis models is difficult to represent the complicated function relationship due to the low ability of data fitting. Furthermore, if the probabilistic relationship among the variables is not clear, it would be difficult to construct the diagnosis models.

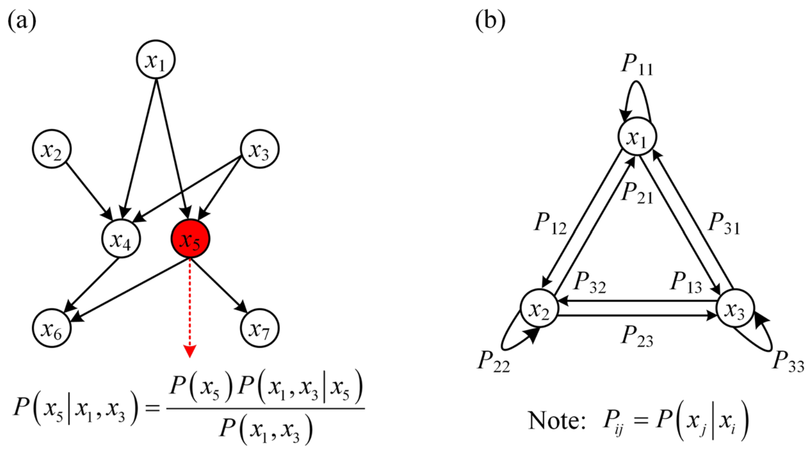


Fig. 7. Illustration of PGM: (a) Bayesian classifier, and (b) hidden Markov model.

(3) Decision tree

Decision tree is also a commonly-used supervised method in classification, which establishes the relationship between the class and the attributes by the tree-shaped architectures, as shown in Fig. 8. Among the proposed methods, the algorithm of C4.5 is widely used to induce a decision tree for classification and obtain satisfactory accuracy and easily-understood classification rules [300]. The C4.5-induced decision tree has been introduced into the fault diagnosis of rolling element bearings [82, 301, 302], gearboxes [303], rotor systems [304], and centrifugal pumps [305]. In order to improve the generalization performance of decision tree, the random forest [306] is further investigated by integrating decisions from multiply tree-based classifiers. In IFD, the decision tree and the extended random forest have been applied for decades, and gain some achievements. Yang et al. [307] discussed the applications of random forest classifier to fault diagnosis of induction motors. Li et al. [46] employed random forest to fuse the diagnosis results of multiple classifiers on the gearboxes, which obtained higher diagnosis accuracy than other fusion tools. Wang et al. [308] proposed a diagnosis model based on the random forest classifier for fault diagnosis of rolling element bearings. Tang et al. [309] used particle warm optimization algorithm to select the optimal parameters of random forest, and improved the diagnosis performance for bearings.

The decision tree-based diagnosis models are naturally interpreted, which may not relay on the explanation of experts and can be easily converted into diagnosis rules. Furthermore, they could achieve the diagnosis tasks with missing data. However, this kind of diagnosis models is easily caught by the over fitting and the low generalization performance, which would reduce the diagnosis performance of the models on the diagnosis tasks. In addition, the tree-type models are mostly constructed according to the expert knowledge.

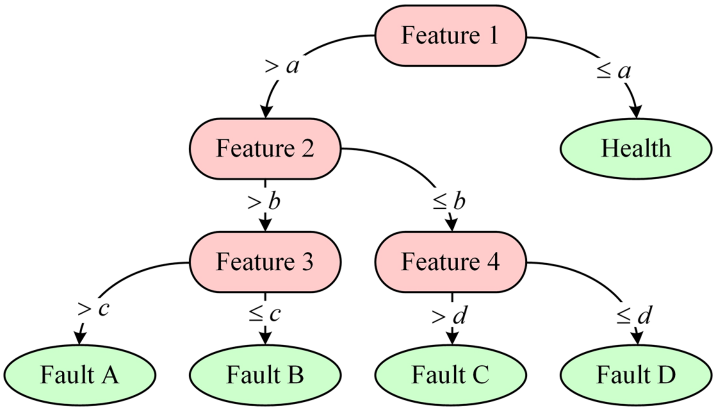


Fig. 8. Illustration of the decision tree.

2.5. Epilog

This section reviews the traditional IFD, and divides the diagnosis architecture into three steps, i.e., data collection, artificial feature extraction, and health state recognition. There are two weaknesses for such diagnosis architecture [310]. First, the step of artificial feature extraction depends on the human labor, in which the engineers need to design powerful algorithms to extract sensitive features to health states of machines. However, it is still unrealistic for engineers to extract specialized features from the large-volume monitoring data by expert experience because of the huge labor cost. Second, the generalization performance and the self-learning capability of traditional diagnosis models are short to bridge the relationship between massive collected data and their corresponding health states, which reduces the diagnosis accuracy. Therefore, it is urgent to investigate diagnosis models that are able to simultaneously extract features from raw collected data and automatically recognize health states of machines.

3. Present: IFD using deep learning theories

This section summarizes the applications of deep learning to fault diagnosis of machines. First, we introduce the diagnosis process of IFD using deep learning theories. Second, the achievements are reviewed about the deep learning-based approaches.

3.1. Overview

With the rapid development of internet technologies and internet of things (IoT), the volume of collected data is dramatically gathered than ever before. The increasingly-grown data bring more sufficient information to machine fault diagnosis so that it is more possible to provide accurate diagnosis results. Unfortunately, the fault diagnosis based on traditional machine learning theories in the past is not appropriate for such big data scenarios. It is necessary to develop some advanced IFD methods.

Deep learning, derived from the research of neural networks [18, 23], employs deep hierarchical architectures to represent the abstract features automatically, and further establish the relationship between the learned features and the target output directly. The deep learning-based diagnosis process, as shown in Fig. 9, consists of two steps, i.e., big data collection and deep learning-based diagnosis [311]. Each step is detailed in the following subsections.

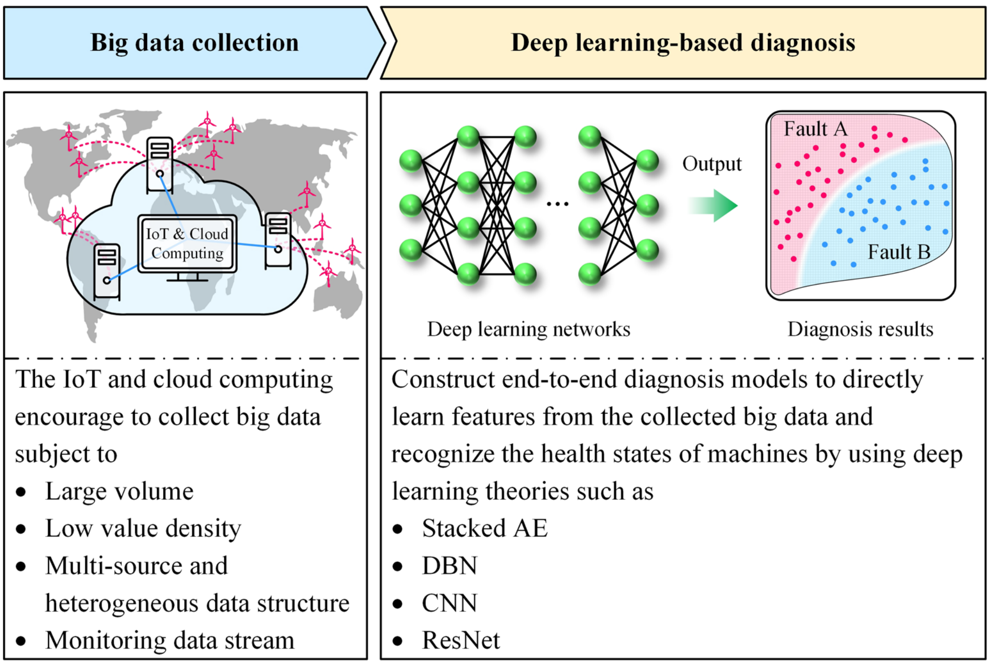


Fig. 9. Diagnosis process of IFD using deep learning theories.

3.2. Step 1: Big data collection

Big data has been a popular term in the modern industry and other application scenarios.

Generally, big data includes four characteristics, i.e., volume, velocity, variety, and veracity [312]. In contrast, monitoring big data of machines holds these characteristics, and further extends to the much-specialized ones. The characteristics are summarized as follows.

- Large volume. The volume of the collected data sustainably grows during the long-term operation of machines, especially for the large-group machines, such as wind turbines in wind sites.
- Low value density. There is incomplete health information in the collected big data. Furthermore, a proportion of poor-quality data is mingled in the massive data [313].
- Multi-source and heterogeneous data structure. Multi-source data will be collected by different kinds of sensors. Furthermore, the data are heterogeneous because of the different storage structures.
- Monitoring data stream. The high-speed transmission channels are able to collect the data from machines immediately.

Such characteristics are contributed by the following conditions [314-316]. 1) In the modern industry, most production activities are achieved by a group of machines. Thus, fault diagnosis tends to focus on machine groups. For example, the monitoring system of a wind site needs to monitor hundreds of wind turbines. During the long-term operation of these machines, the monitoring system constantly acquires data. In particular, it is necessary to collect data with the high sampling frequency, such as the vibration data from gearboxes, because the health information is mostly hidden in the high-frequency band. As a result, the volume of the collected data tends to increase. 2) Although the monitoring system could acquire massive data, only minority of them is valuable. First, the healthy condition accounts for the majority of the long-term operation for machines, while faults seldom happen. Consequently, it is more easily to collect healthy data than faulty data. Second, the quality of the collected data is not always satisfactory because some of them may suffer from the emergencies, such as the transmission interruption and the anomaly of measurement devices. 3) In order to collect sufficient health information of machines, there are many monitoring points on machines. Furthermore, multi-source sensors are used to collect different kinds of data. The collected data are stored with various data structures. For example, the monitoring data of a wind turbine include not only the vibration and speed data from the condition monitoring system (CMS) but also the control parameters from the supervisory control and data acquisition system (SCADA). 4) The prosperous development of sensor technologies and data transmission, especially with the advent of IoT and high-speed internet, promotes to collect a large number of data that contain the real-time information. Furthermore, the advent of state-of-art technologies, such as the edge computing and the applications of GPU, helps cope with monitoring data stream effectively.

3.3. Step 2: Deep learning-based diagnosis

Deep learning-based diagnosis models automatically learn features from the input monitoring data and simultaneously recognize the health states of machines according to the learned features. They mostly include the feature extraction layers and the classification layer. The models first employ the hierarchical networks, such as stacked AE, DBN, CNN, and ResNet, to learn abstracted features layer by layer. Furthermore, the output layer is placed after the last extraction layer for health state recognition, generally with an ANN-based classifier because of the high capability in multi-class classification. During the training process, the error between the actual output and the target is minimized by using BP algorithm to update the training parameters of the diagnosis models. This section reviews four typical deep learning methods and their applications in machine fault diagnosis.

3.3.1. Stacked AE-based approaches

(1) A brief introduction to AE and Stacked AE

As depicted in Fig. 10, AE consists of the encoder network and the decoder network [21]. Given the dataset $\{\boldsymbol{x}_i, \boldsymbol{y}_i\}_{i=1}^{m}$ with m samples, the represented features $\boldsymbol{h}_i$ are defined as

$$\boldsymbol{h}_i = f_{\boldsymbol{\theta}}(x_i) = \sigma_f(\boldsymbol{\omega}^{\mathrm{T}} \cdot \boldsymbol{x}_i + \boldsymbol{b}), \tag{8}$$

where σ_f is the activation function of the encoder network, and $\boldsymbol{\theta} = \{\boldsymbol{\omega}, \boldsymbol{b}\}$ is the training parameters of the encoder network. The reconstructed sample $\hat{\boldsymbol{x}}_i$ can be obtained by the decoder network, which is expressed as follows.

$$\hat{\boldsymbol{x}}_i = g_{\boldsymbol{\theta}'}(\boldsymbol{h}_i) = \sigma_{\boldsymbol{g}}(\boldsymbol{\omega}'^{\mathrm{T}} \cdot \boldsymbol{h}_i + \boldsymbol{b}'), \tag{9}$$

where σ_g is the activation function of the decoder network, and $\boldsymbol{\theta}' = \{\boldsymbol{\omega}', \boldsymbol{b}'\}$ represents the training parameters of the decoder network. In order to reconstruct the original input as well as possible, the optimization objective of AE focuses on minimizing the error between the input samples and the reconstructed ones by

$$\min_{\boldsymbol{\theta}, \boldsymbol{\theta}'} \quad L(\boldsymbol{x}_i, \hat{\boldsymbol{x}}_i) = \frac{1}{2m} \sum_{i=1}^{m} \|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i\|^2. \tag{10}$$

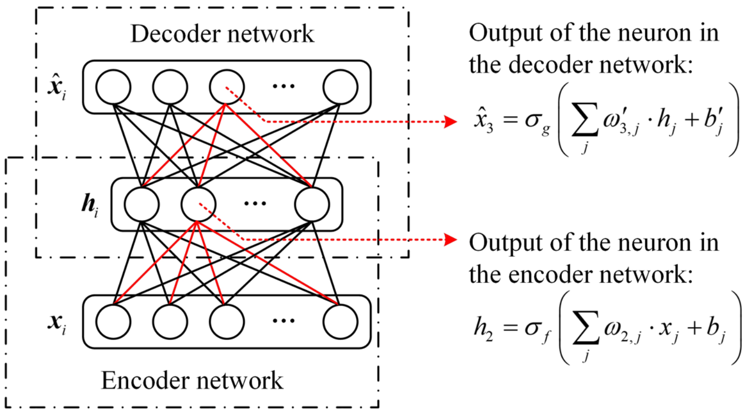


Fig. 10. Architecture of AE.

Multiple AE can be stacked to represent the information contained in the input data deeply, and further obtain the features in deep layers, as shown in Fig. 11. The represented features of the lth AE can be calculated as

$$\boldsymbol{h}_i^l = f_{\boldsymbol{\theta}^l}\left(\boldsymbol{h}_i^{l-1}\right), \qquad l = 2,3,\cdots,L, \tag{11}$$

where $\boldsymbol{\theta}^l$ is the training parameters of the lth AE, and $\boldsymbol{h}_i^l$ is the represented features by the first AE from the given sample $\boldsymbol{x}_i$. After the Lth AE is pre-trained, we can obtain the deep features $\boldsymbol{h}_i^L$. The features can be mapped into the target classes according to an individual classification layer, and the output is $\hat{\boldsymbol{y}}_i = f_{\boldsymbol{\theta}^{L+1}}(\boldsymbol{h}_i^L)$.

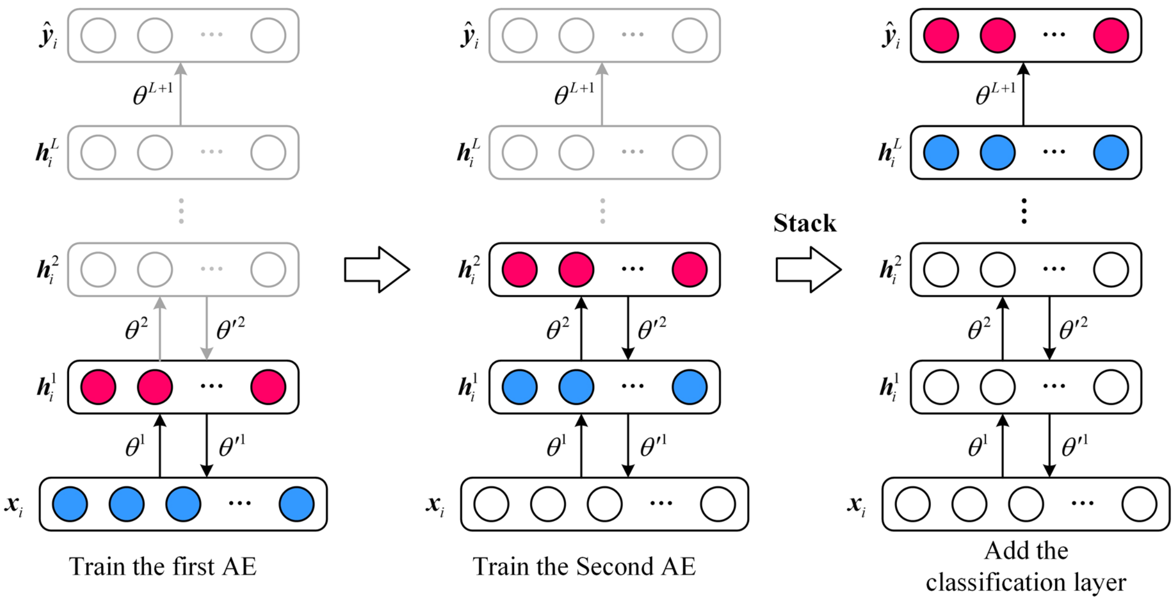


Fig. 11. Greedy layer-wise pre-training for stacked AE.

(2) Applications of AE to machine fault diagnosis

Some publications [310, 317-330] have introduced AE and its common varieties into machine fault diagnosis. Among them, Jia et al. [310] used the stacked AE to automatically learn features from the frequency-domain data and subsequently complete the diagnosis tasks of rolling element bearings and gears, which was one of the earliest studies in applications of stacked AE. The constructed diagnosis model included three stacked AE, which helped automatically separate the useless health information and compress the helpful information rather than manually extract statistic features as the traditional IFD did. From the results, the proposed method was expected to handle massive monitoring data and obtain high diagnosis accuracy. In addition, Liu et al [317] and Lu et al [319] respectively employed the stacked sparse AE and the stacked denoising AE for fault diagnosis of bearings, and the research results presented higher diagnosis accuracy than other methods such as SVM and ANN.

In order to improve the performance of AE-based diagnosis models, researchers further investigated the optimization algorithms of AE. They mainly concerned special varieties of AE on the basic of the common ones. For example, Jia et al [331] proposed a normalized sparse AE to automatically learn meaningful and dissimilar features from the input vibration data, which was one of the earliest work to construct end-to-end diagnosis models. Aiming at the shift-variant properties of raw vibration data, they further proposed the locally connected networks by normalized sparse AE to construct end-to-end networks that encouraged to directly bridge the relationship from the raw monitoring data to the health states of machines. Ref. [332] presented an AE-based diagnosis model for machine fault diagnosis, in which the optimization object of AE was redesigned by the maximum correntropy, and the artificial fish warm algorithm was used to optimize the parameters of AE. Liu et al. [333] used AE to construct a recurrent neural network for fault diagnosis of motor bearings. Ma et al. [334] published a deep coupling AE that could fuse the learned features of multi-source data in high level. Shao et al. [335] discussed the effects of activation functions on the diagnosis performance of AE-based models, and used Gaussian wavelet function as the activation function to design the wavelet AE. They further integrated diagnosis results from a set of stacked AE that were constructed with different activation functions [336]. In Ref. [337, 338], the contractive AE and the convolutional AE were respectively introduced to construct diagnosis models for fault diagnosis of rotating machines. Ref. [339] presented a joint multiple reconstructions AE to jointly learn discriminative and robust features from the multi-scale signals. In addition, researchers aimed to develop the hybrid diagnosis models combined with AE and other methods. For example, the extreme learning machine is used to construct AE-based diagnosis models, in which the parameters are randomly determined rather than adopting the BP algorithm. The training strategy improves the generalization performance and the rate of convergence of conventional AE-based models, and the hybrid diagnosis models have been successfully used for fault diagnosis of motor bearings [340, 341] and wind turbines [342]. DBN is the other method to construct hybrid diagnosis models with stacked AE [343, 344]. In the diagnosis models, stacked AE is considered to learn features from the input monitoring data, while DBN is regard to recognize the health states of machine according to the learned features. In Ref. [345], the authors introduced the batch normalization layer into the stacked AE, which solved the problem of internal covariate shift in training a multi-layer network and accelerated the convergence. Saufi et al. [346] optimized the performance of AE-based diagnosis models by the algorithms of the RProp and the differential evolution.

Stacked AE-based diagnosis models are able to automatically represent the health information from the input monitoring data, which does not rely on much expert knowledge in feature extraction. As an unsupervised learning method, the stacked AE cannot be directly used to recognize the health states of machines. Therefore, a classification layer is usually added at the top of the architecture of the model, and the constructed diagnosis models need to be trained with sufficient labeled samples.

3.3.2. DBN-based approaches

(1) A brief introduction to RBM and DBN

As shown in Fig. 12, RBM is a special type of generative stochastic neural network including visible units $\boldsymbol{v} = \{v_1, v_2, \cdots, v_m\}$ and hidden units $\boldsymbol{h} = \{h_1, h_2, \cdots, h_n\}$ [22]. It is noted that all the units are binary, i.e., $\boldsymbol{v}, \boldsymbol{h} \in \{0,1\}$. As an energy-based model, the variables $\boldsymbol{v}$

and $\boldsymbol{h}$ are subject to the joint configuration as follows.

$$E(\boldsymbol{v},\boldsymbol{h},\boldsymbol{\theta}) = -\sum_{i=1}^{m}\sum_{j=1}^{n}\omega_{i,j}v_i h_j - \sum_{i=1}^{m} b_i v_i - \sum_{j=1}^{n} a_j h_j, \tag{12}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\omega},\boldsymbol{a},\boldsymbol{b}\}$ represents the parameters of RBM. After that, the marginal distribution of the visual units can be calculated as

$$P(\boldsymbol{v}|\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})}\sum_{\boldsymbol{h}}\exp[-E(\boldsymbol{v},\boldsymbol{h},\boldsymbol{\theta})], \tag{13}$$

where $Z(\boldsymbol{\theta}) = \sum_{\boldsymbol{v},\boldsymbol{h}}\exp[-E(\boldsymbol{v},\boldsymbol{h},\boldsymbol{\theta})]$ is the partition function. The activation conditions for visible units and hidden units are defined as follows.

$$P(v_i = 1|\boldsymbol{h}) = \sigma_s\left(b_i + \sum_{j=1}^{m}\omega_{i,j}\cdot h_j\right) \quad \text{and} \quad P(h_i = 1|\boldsymbol{v}) = \sigma_s\left(a_i + \sum_{j=1}^{n}\omega_{i,j}\cdot v_j\right), \tag{14}$$

where σ_s is the activation function of Sigmoid. The maximum likelihood estimation is employed to obtain the parameters of RBM, which is calculated by

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \quad \ln[P(\boldsymbol{\theta}|\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_k)] = \frac{1}{k}\sum_{i=1}^{k} ln[P(\boldsymbol{x}_i|\boldsymbol{\theta})], \tag{15}$$

where $\{\boldsymbol{x}_i\}_{i=1}^{k}$ represents the input dataset with k samples. In order to simplify the solution shown in Eq. (15), the contrastive divergence algorithm [22] is employed to accelerate the computation and further obtain the estimated parameters.

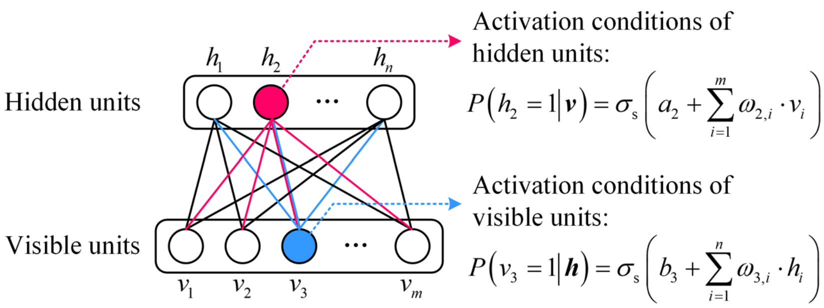


Fig. 12. Architecture of RBM.

Similar to the pre-training in stacked AE, DBN is constructed by stacking a set of RBMs, in which the hidden units of the previous layer in DBN are also viewed as the visible units of the next layer [23]. After the Lth RBM is trained, the represented features in the deep layer can be mapped into the target class by mostly adding the Softmax classification layer as

$$\widehat{\boldsymbol{y}}_i = [P(y_i = 1|\boldsymbol{h}_i^L, \boldsymbol{\theta}^C), \cdots, P(y_i = q|\boldsymbol{h}_i^L, \boldsymbol{\theta}^C), \cdots, P(y_i = k|\boldsymbol{h}_i^L, \boldsymbol{\theta}^C)] \ , \tag{16}$$

where $P(y_i = q|\boldsymbol{h}_i^L, \boldsymbol{\theta}^C) = \exp\left(\boldsymbol{\omega}_q^C \cdot \boldsymbol{h}_i^L + \boldsymbol{b}_q^C\right) / \sum_{q=1}^{k} \exp(\boldsymbol{\omega}_q^C \cdot \boldsymbol{h}_i^L + \boldsymbol{b}_q^C)$, $\boldsymbol{h}_i^L$ is the represented features in the Lth layer from the ith sample, $\widehat{\boldsymbol{y}}_i$ represents the predicted one-hot class, and $\boldsymbol{\theta}^C = \{\boldsymbol{\omega}^C, \boldsymbol{b}^C\}$ is the training parameters of the classification layer.

(2) Applications of DBN to machine fault diagnosis

DBN has been an effective way in the research of IFD. For fault diagnosis of rolling element bearings, Ref. [347] presented a diagnosis model based on the single Gaussian RBM. Jiang et al [348] and Han et al. [349] stacked multiple RBMs to construct the DBN-based diagnosis models, which presented higher diagnosis accuracy than the traditional ones. In order to improve the diagnosis performance, researchers further investigated the optimization algorithm for the DBN-based models. In Ref. [350, 351], the Nesterov momentum was used to adaptively optimize the training of DBN-based diagnosis models, and the diagnosis accuracy was higher than the standard DBN. Shao et al. [352] constructed an adaptive DBN that was trained with the algorithm of adaptive learning rate and momentum. They also tried to adaptively determine the structure of DBN-based diagnosis models by using the particle warm [353]. In Ref. [354, 355], Shao et al. further presented a convolutional DBN for fault diagnosis of bearings, and the exponential moving average technique was used to improve the performance of the diagnosis models. Furthermore, DBN has been used for fault diagnosis of other objects. For example, Tamilselvan et al [356] used DBN for fault diagnosis of aircraft engines, which was one of the earliest research in this field. Sun et al. [357] proposed a fault diagnosis model named Tilear for the electromotor, and the model was constructed with DBN. Tran et al. [358] presented a DBN-based diagnosis model for reciprocating compressor valves, in which the Gaussian-Bernoulli RBM was considered to stack the hierarchical structure. In Ref. [359], Qiu et al. constructed a diagnosis model based on DBN and the hidden Markov model for the early-warning of compressor unit. Similarly, Gao et al. [360] added a quantum inspired neural network to the top layer of DBN, which was applied for fault diagnosis of aircraft fuel system. In Ref. [361], DBN was employed for automated diagnosis of vehicle on-board equipment of high speed trains, which presented better diagnosis performance than *k*NN and ANN. He et al [362] used DBN for fault diagnosis of a gear transmission chain, and the genetic algorithm was further used to optimize the structure of DBN. For fault diagnosis of rotor systems [363] and hydraulic equipment [364], DBN was considered to construct diagnosis models with higher diagnosis accuracy than the traditional methods. For air-conditioning system, Guo et al. [365] constructed a diagnosis model with the help of DBN to recognize the faults of the four-way reversing valve, the outdoor unit, and the refrigerant charge. In addition, Yu et al [366] proposed a data-driven fault diagnosis model for wind turbines, which was also implemented by DBN.

Different from the stacked AE, DBN-based diagnosis models could automatically learn features from the input data by pre-training a set of stacked RBMs, which solves the problem of vanishing gradient in using BP algorithm to fine-tune the deep-layer networks. In order to recognize the health states of machines, DBN maps the learned features into the label space by adding the classification layer. It is necessary to use sufficient labeled data to train the constructed diagnosis models so as to obtain the convinced diagnosis results.

3.3.3. CNN-based approaches

(1) A brief introduction to CNN

CNN, as a supervised deep learning method, has completed several superior achievements in speech recognition, image identification, and target tracking [367]. Generally, CNN consists of convolutional layers, pooling layers, and full-connected layers [24]. The basic principles of convolution and pooling are detailed in Fig. 13. In convolutional layers, the filter kernels $k^c \in \Re^{H\times L\times D}$ are used to convolve the input vectors $\boldsymbol{x}^{c-1} \in \Re^{M\times N}$ from the previous $(c-1)$th layer, where H is the height of the kernels, L and D are respectively the length and the depth of the kernels. The output feature map of the cth layer is obtained as follows.

$$\begin{gathered}\boldsymbol{x}_i^c = \sigma_{\mathrm{r}}(\boldsymbol{x}_i^{c-1} * \boldsymbol{k}^c + \boldsymbol{b}^c) \in \Re^{(M-H+1)\times(N-L+1)\times D}, \qquad c = 2,3,\cdots, \\ (\boldsymbol{x}_i^{c-1} * \boldsymbol{k}^c)_{j,k,d} = \sum_{m=1}^{M}\sum_{h=1}^{H}\sum_{l=1}^{L} x_{(i),j+h-1,k+l-1,m}^{c-1} \cdot k_{h,l,d}^{c}\end{gathered} \tag{17}$$

where σ_{r} represents the rectified linear unit (ReLU) [368]. In pooling layers, the down-sampling processing is used to reduce the number of the training parameters and overcome the over fitting effectively, as shown in Fig. 13(b). The commonly-used down-sampling forms include max pooling and mean pooling. The pooled feature map is further expressed as follows.

$$\begin{gathered}\left(\boldsymbol{v}_i^{\mathrm{p}}\right)_{m,n,d} = \mathrm{down}\left\{x_{(i),j,k,d}^{p-1} \middle| \forall x_{(i),j,k,d}^{p-1} \in \boldsymbol{x}_i^{p-1}, j,k \in N^+, \boldsymbol{s}^{\mathrm{r}\times\mathrm{t}}\right\}, \\ \mathrm{s.t.} \qquad s^{\mathrm{r}} \le j \le s^{\mathrm{r}} m, \qquad s^{\mathrm{t}}(n-1) \le j \le s^{\mathrm{t}} n\end{gathered} \tag{18}$$

where $\mathrm{down}(\cdot)$ is the down-sampling functions respectively including $\max(\cdot)$ and $\mathrm{mean}(\cdot)$, and $\boldsymbol{s}^{\mathrm{r}\times\mathrm{t}}$ is the filters in the pooling layers. According to stacking the convolutional layers and pooling layers, CNN is able to learn the deep-layer features from the input data. These features are then flattened into a 1D vector as the input of the full-connected layers. Through the multi-layer neural networks, they are further mapped into the class target. The outputs of full-connected layers are represents as

$$\boldsymbol{x}_i^f = \sigma_r(\boldsymbol{\omega}^f \cdot \boldsymbol{x}_i^{f-1} + \boldsymbol{b}^f), \qquad f = 2,3,\cdots, \tag{19}$$

where $\boldsymbol{x}_i^1 = \text{flatten}(\boldsymbol{v}_i^p)$ is the input of full-connected layers, and $\boldsymbol{\theta}^f = \{\boldsymbol{\omega}^f, \boldsymbol{b}^f\}$ represents the training parameters of the full-connected layers.

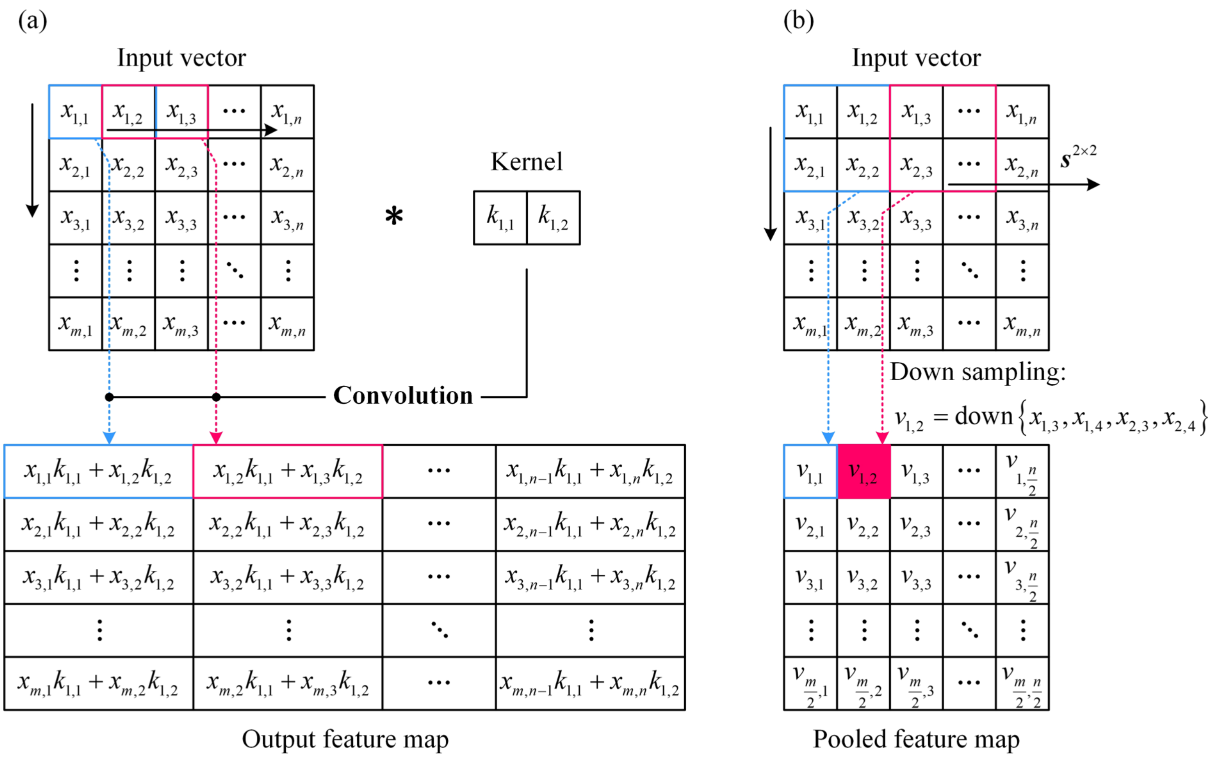


Fig. 13. Convolution and pooling process: (a) Convolution process by using the kernel $\boldsymbol{k}^c \in \Re^{2\times1\times1}$, and (b) pooling process by using the filter $\boldsymbol{s}^{2\times2}$.

(2) Applications of CNN to machine fault diagnosis

The applications of CNN to machine fault diagnosis are categorized in Table 3. According to the architectures of CNN, they can be divided into the 2-dimensional (2D) CNN-based diagnosis models and the 1D CNN-based diagnosis models. Originally, the 2D CNN serves as the standard version for image identification, in which the input images are 2D data. For fault diagnosis of machines, however, the 2D CNN is unable to handle the 1D signals, such as vibration data. In order to construct a CNN-based diagnosis model and obtain high performance, researchers adopted some effective approaches, which are summarized into three types. First, the signal processing methods, such as the wavelet packet [369-371], the continue wavelet transform [372, 373], the dual-tree complex wavelet transform [374, 375], and the synchrosqueezing transform [376], are employed to preprocess the input 1D data, which are expected to convert the signals from the time domain to the time-frequency domain. After that, CNN is able to handle the monitoring data with the 2D time-frequency spectrum. Second, some publications [377-382] manually reshaped the dimensions of the input data to make them suitable for the CNN-based diagnosis models. In these achievements, the data preprocessing mostly derives from some simple and ingenious operation rather than the advanced signal processing, which almost gets rid of the help of the expert knowledge. Third, the image data, such as grey scale image [383, 384] and infrared thermal image [385, 386], are further developed for fault diagnosis of machines. Consequently, the CNN-based diagnosis models are able to directly work well, just like the classification tasks of image identification. In addition, the 1D CNN begin to cope with the vibration data that are subject to the shift-variant characteristics, and successfully helped construct the end-to-end diagnosis models for rolling element bearings [387-391], gears [392-398], motors [399], and hydraulic pumps [400], in which the input of the diagnosis models is the raw data without preprocessing.

Compared with the stacked AE and DBN, CNN-based diagnosis models are able to directly learn features from the raw monitoring data without preprocessing such as the frequency-domain transformation because CNN is able to capture the shift-variant properties of input data. Furthermore, the number of training parameters in diagnosis models is reduced by sharing weights, which could accelerate the convergence and restrain the over fitting. Similar to other deep learning-based models, the diagnosis performance of CNN-based diagnosis models is subject to training with sufficient labeled samples.

Table 3 Summary of the applications of CNN to machine fault diagnosis

Architectures	Input data	References
2D CNN	Time-frequency spectrum	Ding et al. [370], Sun et al. [374], Verstraete et al. [401], Guo et al. [402], Han et al. [371], Xin et al. [403], Guo et al. [372, 373], Cao et al. [375], Chen et al. [404], Han et al. [405], Islam et al. [369], Zhao et al. [376], and Zhu et al. [406]
	Reshaped matrix	Jiang et al. [407], Li et al. [377], Liu et al. [378], Lu et al. [379], Wang et al. [381], Xia et al. [382], and Wang et al. [380]
	Images	Janssens et al. [386], Wen et al. [383], Yuan et al. [385], Zhou et al. [384], and Suh et al. [408]
1D CNN	Raw data or frequency spectrum	Ince et al. [399], Yan et al. [400], Eren et al. [387, 389], Jing et al. [392, 394], Appana et al. [391], Chen et al. [388], Jia et al. [390], Jiao et al. [393], Yao et al. [396], Han et al. [395], Huang et al. [409], Jiang et al. [398], and Li et al. [397]

3.3.4. ResNet-based approaches

(1) A brief introduction to ResNet

ResNet plays an ingenious and successful role in the research of neural networks, which constructs deep-learning model by stacking several residual blocks [25]. The residual blocks consist of the forward channel and the shortcut connection, as shown in Fig. 14. The forward channels of the residual blocks are originally constructed by stacking some convolutional layers. For example, two convolutional layers are used to handle the input features by

$$g\left(\boldsymbol{x}_i^{l-1}\middle|\boldsymbol{\theta}^l\right) = \sigma_{\mathrm{r}}\left(\boldsymbol{x}_i^{l-1} * \boldsymbol{k}^{l_1} + \boldsymbol{b}^{l_1}\right) * \boldsymbol{k}^{l_2} + \boldsymbol{b}^{l_2}, \tag{20}$$

where $\boldsymbol{\theta}^l = \{\boldsymbol{k}^{l_1}, \boldsymbol{b}^{l_1}, \boldsymbol{k}^{l_2}, \boldsymbol{b}^{l_2}\}$ is the training parameters in the lth residual block. It is noted that the convolutional processing is conducted with zero padding to hold the dimensions of features as they go through the convolutional layers. The shortcut connection is then introduced to calculate the sum of the output of forward channels and the input features, and the output of the residual block is represented as follows.

$$\boldsymbol{x}_i^l = \sigma_{\mathrm{r}}\left[g\left(\boldsymbol{x}_i^{l-1} \middle| \boldsymbol{\theta}^l\right) + \boldsymbol{x}_i^{l-1}\right], \qquad l = 2,3,\cdots,L, \tag{21}$$

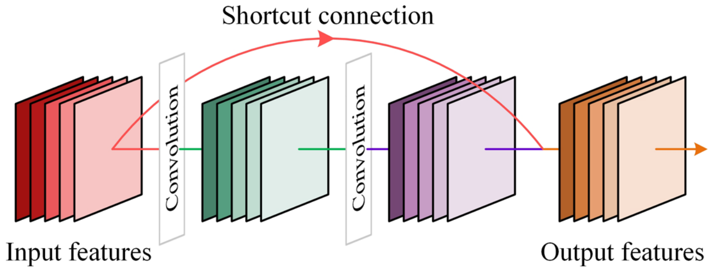


Fig. 14. Architecture of the residual block.

The deep-layer features are obtained by stacking multiple residual blocks, in which the output of the previous block is the input of the next one. The learned features are finally mapped into the target class by full-connected layers.

(2) Applications of ResNet to machine fault diagnosis

Researchers developed studies about applications of ResNet. Zhang et al. [410] constructed a diagnosis model combined with ResNet for rolling element bearings, in which the raw vibration data was directly used to train the diagnosis model. The comparison results presented the higher diagnosis accuracy than CNN-based diagnosis models. Zhao et al. [411] developed the dynamically weighted wavelet coefficients to improve the performance of ResNet-based diagnosis models, and obtained higher accuracy for fault diagnosis of planetary gearboxes under serious noise environment than other deep learning-based methods. They further proposed two multiple wavelet coefficient fusion methods [412] for ResNet-based diagnosis models, which helped learn more easily-distinguished features from the input data than the standard ResNet-based models. A data-driven diagnosis model combined with time-frequency analysis and ResNet was proposed by Ma et al. [413] for fault diagnosis of planetary gearboxes. For the key components of high-speed trains, Peng et al. [414] used ResNet to recognize the health states of wheelset bearings, and Su et al. [415] constructed a ResNet-based diagnosis model for fault diagnosis of the bogies. Both the experimental results of publications showed the superiority of the ResNet-based diagnosis model than other deep-learning approaches.

ResNet is developed to obtain higher generalization performance on the basis of the architecture of CNN. Therefore, the ResNet-based diagnosis models inherit the advantages of the CNN-based diagnosis models, and possibly obtain high diagnosis accuracy, especially for the complicated operation conditions such as varying-speed or varying-load conditions.

3.4. Epilog

This section reviews the applications of deep learning to machine fault diagnosis in the present, and divides the diagnosis process into data collection and the deep learning-based diagnosis. Such diagnosis architecture is expected to construct end-to-end diagnosis models that could directly bridge the relationship between the raw monitoring data and the health states of machines. Although some successes have been achieved, they are mostly subject to the common assumption: the labeled data are sufficient and contain completed information about the health states of machines [35]. In engineering scenarios, however, such assumption is unpractical because of two characteristics for the data collected from real-case machines [26]. 1) It is difficult for these data to contain sufficient information to reflect completed kinds of health states. The fact is that machines mostly work under the healthy state, while the faults seldom happen. Thus, it is easier to collect healthy data than faulty data. As a result, the collected data are seriously imbalanced. 2) The majority of the collected data are unlabeled. It is unrealistic to frequently stop the machines and inspect the health states due to the huge loss of wealth. According to the aforementioned two characteristics, it is necessary to train reliable diagnosis models for engineering scenarios in future.

4. Future: IFD using transfer learning theories

This section forecasts one of the most potential research prospects, i.e., applications of transfer learning to machine fault diagnosis. There are three subsections: 1) why to transfer, 2) what is transfer, and 3) how to transfer. The first subsection focuses on the motivation why transfer learning comes to be a promising topic in the future research of IFD. In the second subsection, we define transfer learning in IFD. In the third subsection, a few exploratory studies are reviewed according to the different categories of transfer learning.

4.1. Motivation of applying transfer learning to IFD in engineering scenarios

The successes of IFD mostly rely on sufficient labeled data to train diagnosis models based on machine learning. However, it takes much cost to recollect sufficient data and further label them, which is unpractical for machines in engineering scenarios. Such problem may be solved by the idea that the diagnosis knowledge could be reused across multiply related machines. For example, the diagnosis knowledge from the laboratory-used bearings may help recognize the health states of bearings in engineering scenarios. In such scenario, it is possible to simulate diverse faults and collect sufficient labeled data from laboratory-used bearings. The diagnosis models trained with them could also work for fault diagnosis of bearings in engineering scenarios if the diagnosis knowledge could be reused. Transfer learning is able to get the above purpose, in which the knowledge from one or more diagnosis tasks can be reused to other related but different ones [27]. With the help of transfer learning theories, it is dispensable to collect sufficient labeled data, which releases the common assumption in training diagnosis models based on machine learning. As a result, IFD is expected to be expanded from the academic research to engineering scenarios.

4.2. Definitions of transfer learning in IFD

In this subsection, we attempt to define the transfer problems and transfer scenarios in IFD combining with the basic definitions of transfer learning.

4.2.1. Transfer problems in IFD

For transfer learning in IFD, the diagnosis knowledge is expected to be reused from one or multiple diagnosis tasks (the source domain) to other related but different ones (the target domain), which refers to the concepts of the domain and the task. The domain is denoted as a pair of $\mathcal{D} = \{X, P(X)\}$ including the dataset $X = \{\boldsymbol{x}_i\}_{i=1}^{n}$ and its marginal probability distribution $P(X)$. The diagnosis task $\mathcal{T} = \{\mathcal{Y}, f(\cdot)\}$ consists of the label space $\mathcal{Y} = \{y_i\}_{i=1}^{n}$ and the diagnosis model $f(\cdot)$. The task contains the health information of machines by observing the health states in the label space. The diagnosis model $f(\cdot)$ can be trained with the labeled data $\{x_i, y_i\}_{i=1}^{n}$. After that, the model is further endowed with diagnosis knowledge, i.e., the relationship between the input data and their corresponding health states of machines.

Given the source domain $\mathcal{D}^{\mathrm{s}}$, the target domain $\mathcal{D}^{\mathrm{t}}$, and the diagnosis tasks $\mathcal{T}^{\mathrm{s}}$ and $\mathcal{T}^{\mathrm{t}}$, transfer problems in IFD aim to apply the diagnosis knowledge from tasks $\mathcal{T}^{s}$ to improve the performance of diagnosis model $f_{\mathrm{T}}(\cdot)$ on the related task $\mathcal{T}^{\mathrm{t}}$. It is noted that $\mathcal{D}^{s} \neq \mathcal{D}^{t}$ and $\mathcal{T}^{s} \neq \mathcal{T}^{t}$. To be specific, the source and target domains are detailed as follows [26, 27].

- The source domain is considered as the pair of $\mathcal{D}^{\mathrm{s}} = \{X^{\mathrm{s}}, P_{\mathrm{s}}(X)\}$ to provide the diagnosis knowledge from one or multiple source diagnosis tasks $\mathcal{T}^{\mathrm{s}} = \{\mathcal{Y}^{\mathrm{s}}, f_{\mathrm{s}}(\cdot)\}$. The dataset X^{s} contains n_{s} labeled samples $\{\boldsymbol{x}_i^{\mathrm{s}}, y_i^{\mathrm{s}}\}_{i=1}^{n_{\mathrm{s}}}$ following the distribution $P_{\mathrm{s}}(X)$. Given the label space $\mathcal{Y} = \{1,2,\cdots,k\}$ with k kinds of health states of machines, the label space in the source domain is subject to the condition $\mathcal{Y}^{\mathrm{s}} \subseteq \mathcal{Y}$.
- The target domain serves as the one where the diagnosis knowledge from the source

domain is reused, which is regarded as the pair of $\mathcal{D}^{\mathrm{t}} = \{X^{\mathrm{t}}, P_{\mathrm{t}}(X)\}$. Different from the samples in the source domain, the target-domain datasets X^{t} consist of n_{t} samples $\{\boldsymbol{x}_i^{\mathrm{t}}\}_{i=1}^{n_{\mathrm{t}}}$ but only a few of them are labeled. To make the matters more serious, there is none labeled samples in the target domain. These samples are drawn from the distribution $P_{\mathrm{t}}(X)$, and $P_{\mathrm{t}}(X) \neq P_{\mathrm{s}}(X)$.

- In order to ensure the success of transferring the diagnosis knowledge across domains, the label space of the source domain should overlap that of the target domain, i.e., $\Upsilon^{\mathrm{t}} \subseteq \Upsilon^{\mathrm{s}} \subseteq \Upsilon$. Such constraint can be explained by the example that it is unrealistic to use the diagnosis knowledge from motors for fault diagnosis of generators.

4.2.2. Transfer scenarios in IFD

Transfer scenarios in IFD can be divided into two categories, i.e., transfer in the identical machine (TIM) and transfer across different machines (TDM), as shown in Table 4. These two categories are both subject to the common assumption, in which the source-domain data are labeled, while there are minority of labeled data or even none of labeled data in the target domain. 1) In TIM scenarios, the source and target domain data are collected from the identical machine, but with varying operation conditions like varying speed and varying load, or various working environments [416]. These factors change the distribution of the collected data so that the diagnosis models trained by using the source domain data are unable to directly work on the target domain. 2) In TDM scenarios, the source and target domain data are collected from different but related machines like motors and generators. These data involve more complicated factors than TIM, such as different machine specifications, diverse structures, etc. [26, 35, 417]. Such factors also result in serious distribution discrepancy of the data between the source domain and the target domain. Therefore, transfer learning is expected to construct diagnosis models for transfer scenarios in IFD, which are robust to the aforementioned factors.

Table 4 Categories of transfer scenarios in machine fault diagnosis.

Transfer scenarios	Assumptions		Factors leading to transfer scenarios
	Source domain	Target domain	
Data collected from the identical machine (TIM)	Available labels	• Minority of labels • Unavailable labels	• Varying speed • Varying load • Various working environments
Data collected from different machines (TDM)	Available labels	• Minority of labels • Unavailable labels	• Different machine specifications • Diverse structures • Different measurement environments • Different working environments

4.3. Categories of transfer learning-based approaches in IFD

A few researchers have developed exploratory studies in IFD by using transfer learning theories, which are summarized in Table 5. We further divide them into four categories, i.e., feature-based approaches, GAN-based approaches, instance-based approaches, and parameter-based approaches. Among them, the studies about feature-based approaches account for the largest proportion. Furthermore, majority of studies focus on TIM scenarios, and only four publications aim at TDM scenarios.

Table 5 Summary of applications of transfer learning to machine fault diagnosis.

Approaches	References	Transfer scenarios		Methodologies
		TIM	TDM	
Feature-based	Chen et al. [418], Xie et al. [419], and Tong et al. [420, 421]	√		TCA and JDA

	Zhang et al. [416, 422], Qian et al. [423], and Peng et al. [414]	√		Deep learning and AdaBN
	Lu et al. [424], Wen et al. [425], Li et al. [426], Zhang et al. [427], Wang et al. [428], Qian et al. [429], Wang et al. [430] Yang et al. [26, 417, 443], and Xu et al. [431]	√	√	Deep transfer learning
	Wang et al. [432], Zhang et al. [433], and Zheng et al. [434]	√	√	Transfer factor analysis, and subspace alignment
GAN-based	Xie et al. [435], Li et al. [436], Han et al. [437], and Guo et al. [35]	√	√	GAN and MMD
Instance-based	Shen et al. [36]	√		TrAdaBoost
Parameter-based	Zhang et al. [438], Hasan et al. [37], Cao et al. [439], and Shao et al. [440]	√		Deep learning and fine tuning

4.3.1. Feature-based approaches

Feature-based approaches are widely investigated in transfer learning because of the capability of correcting serious across-domain discrepancy [33], such as TDM scenarios. As shown in Fig. 15, such approaches can be commonly divided into four steps [26]. First, a feature mapping is used to map the cross-domain data into a common feature space. After that, the distribution discrepancy of the features is measured by distance metrics. Furthermore, the results are propagated backward to update the parameters of the feature mapping by the minimization optimization strategy, which helps reduce the distribution discrepancy of the features. Finally, the domain-shared classifier trained with the source-domain samples is employed to work on the target domain according to the features with similar distribution.

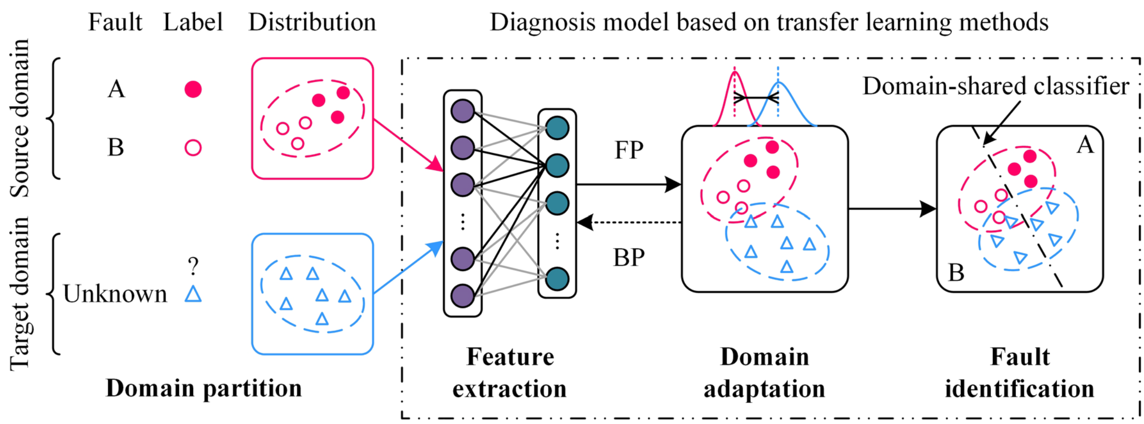


Fig. 15. Steps of feature-based transfer learning approaches [26].

(1) TCA and JDA

TCA is a typically feature-based approach [29]. This approach attempts to find a low-dimensional feature space, in which the cross-domain data are subject to small distribution discrepancy. After that, the learned features are used to train domain-shared classifiers that are mostly constructed by traditional machine learning theories. The optimization objective of TCA is shown as

$$\begin{aligned} \min_{\boldsymbol{W}} \quad & \operatorname{trace}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{KLKW}) + \mu \cdot \operatorname{trace}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W}) \\ \text{s.t.} \quad & \boldsymbol{W}^{\mathrm{T}}\boldsymbol{KHKW} = \boldsymbol{I} \end{aligned} \tag{22}$$

where $\boldsymbol{K} = [K_{i,j}] \in \Re^{(n_s+n_t)\times(n_s+n_t)}$ is the kernel matrix of the input cross-domain samples and $K_{i,j} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$, $\boldsymbol{W} = \boldsymbol{K}^{-1/2}\widetilde{\boldsymbol{W}} \in \Re^{(n_s+n_t)\times m}$ maps the cross-domain samples from the space $\Re^{n_s+n_t}$ to the m-dimensional space $\Re^m$ and $n_s + n_t > m$, μ is the tradeoff parameter to balance the contributions of the distribution adaptation and the model complexity, $\boldsymbol{H} = \boldsymbol{I}_{n_s+n_t} - 1/(n_s + n_t)\mathbf{1}\mathbf{1}^T$ is the centering matrix, and $\boldsymbol{L} = [L_{i,j}] \geq 0$ can be calculated as

$$L_{i,j} = \begin{cases} \dfrac{1}{n_\mathrm{s}^2}, & \boldsymbol{x}_i, \boldsymbol{x}_j \in X^\mathrm{s} \\ \dfrac{1}{n_\mathrm{t}^2}, & \boldsymbol{x}_i, \boldsymbol{x}_j \in X^\mathrm{t} \\ -\dfrac{1}{n_\mathrm{s} n_\mathrm{t}}, & \begin{cases} \boldsymbol{x}_i \in X^\mathrm{s}, \boldsymbol{x}_j \in X^\mathrm{t} \\ \boldsymbol{x}_i \in X^\mathrm{t}, \boldsymbol{x}_j \in X^\mathrm{s} \end{cases} \end{cases}. \tag{23}$$

The optimal feature mapping $\boldsymbol{W}^*$ obtained by solving Eq. (22) can be further used to calculate the cross-domain features $\boldsymbol{W}^*\boldsymbol{K}$ subject to similar distribution.

A few researchers have introduced TCA to reduce the distribution discrepancy of the cross-domain data in IFD. Chen et al. [418] used TCA to extract transferable features of the collected data from rolling element bearings under different operation conditions. Xie et al. [419] employed TCA and SVM-based classifier for fault diagnosis of gearboxes under different operation conditions.

TCA just adapts the marginal probability distribution of the input cross-domain data, but ignores the conditional probability distribution from the feature space to the target classes. Thus, JDA is further proposed to solve this problem, which is defined as follows [30].

$$\begin{aligned} \min_{\boldsymbol{W}} \quad & \sum_{c=1}^{C} \mathrm{trace}(\boldsymbol{W}^\mathrm{T}\boldsymbol{K}\boldsymbol{L}_c\boldsymbol{K}\boldsymbol{W}) + \lambda \cdot \|\boldsymbol{W}\|_\mathrm{F}^2 \\ \mathrm{s.t.} \quad & \boldsymbol{W}^\mathrm{T}\boldsymbol{K}\boldsymbol{H}\boldsymbol{K}\boldsymbol{W} = \boldsymbol{I} \end{aligned} \tag{24}$$

where $\boldsymbol{L}_c = \left[L_{i,j}^{(c)}\right] \geq 0$, and $\boldsymbol{L}_c = \boldsymbol{L}$ when $c = 0$. Otherwise the elements of $\boldsymbol{L}_c$ is

$$L_{i,j} = \begin{cases} \dfrac{1}{n_{\mathrm{s,(c)}}^2}, & \boldsymbol{x}_i, \boldsymbol{x}_j \in X_{(c)}^\mathrm{s} = \{\boldsymbol{x}_i | \boldsymbol{x}_i \in X^\mathrm{s} \wedge X_{y_i=c}\} \\ \dfrac{1}{n_{\mathrm{t,(c)}}^2}, & \boldsymbol{x}_i, \boldsymbol{x}_j \in X_{(c)}^\mathrm{t} = \{\boldsymbol{x}_i | \boldsymbol{x}_i \in X^\mathrm{t} \wedge X_{y_i=c}\}, \\ -\dfrac{1}{n_{\mathrm{s,(c)}} n_{\mathrm{t,(c)}}}, & \begin{cases} \boldsymbol{x}_i \in X_{(c)}^\mathrm{s}, \boldsymbol{x}_j \in X_{(c)}^\mathrm{t} \\ \boldsymbol{x}_i \in X_{(c)}^\mathrm{t}, \boldsymbol{x}_j \in X_{(c)}^\mathrm{s} \end{cases} \end{cases} \tag{25}$$

where the sub-domains $X_{(c)}^\mathrm{s}$ and $X_{(c)}^\mathrm{t}$ are the set of samples belonging to the class c in the source domain. Similar to solving TCA, the cross-domain features $\boldsymbol{W}^*\boldsymbol{K}$ can be obtained by optimizing the function shown in Eq. (24).

With regard to fault diagnosis, Tong et al. [420, 421] employed JDA to complete the diagnosis tasks of bearings respectively used in the motor and the belt conveyor idler, and achieve better diagnosis accuracy than TCA-based diagnosis models.

Diagnosis models based on TCA and JDA use the simple nonlinear mapping to extract features, which is difficult to fit the complicated distribution of data. As a result, the diagnosis models may get poor transfer results on the target domain because of the under-corrected discrepancy of the source and target domains.

(2) Deep learning and AdaBN

The advent of deep learning replaces the simple nonlinear mapping in TCA and JDA with the deep hierarchical architectures. Several publications solve the TIM scenarios just by deep learning-based diagnosis models. Zhang et al. [416] constructed a deep CNN-based diagnosis model for fault diagnosis of motor bearings under varying operation conditions and different noisy environments. Peng et al. [414] employed the ResNet-based diagnosis models for locomotive bearings, which were robust to varying operation conditions and the added noise. In Ref. [422, 423], the adaptive batch normalization (AdaBN) [441] was used to improve the performance of CNN-based diagnosis models for bearings under different operation conditions.

Deep learning-based diagnosis models are beneficial to reducing the cross-domain discrepancy by extracting deep-layer features [33]. However, they are unable to adapt serious cross-domain discrepancy in some diagnosis scenarios, such as TDM.

(3) Deep transfer learning

In order to correct the serious cross-domain discrepancy, deep transfer learning impose constraints on the parameters of deep learning-based model by minimizing the distance metric to distribution discrepancy. Maximum mean discrepancy (MMD) is commonly-used nonparametric distance to the distribution discrepancy, which is defined as follows.

$$D_{\mathcal{H}}^{2}(X,Y) \coloneqq \sup_{\Phi\in\mathcal{H}}\{E_{X\sim p}[\Phi(\boldsymbol{x})] - E_{Y\sim q}[\Phi(\boldsymbol{y})]\}, \tag{26}$$

where $\sup\{\cdot\}$ is the supremum of the input aggregate, $\mathcal{H}$ represents the reproduced kernel Hilbert space (RKHS), and $\Phi(\cdot)$ is the nonlinear mapping from the original space to RKHS. The MMD of learned cross-domain features is viewed as the regularization term of optimization objective for deep learning-based diagnosis models. For example, the publications [424, 425, 431] employed MMD to regularize the optimization objectives of stacked AE-based diagnosis models as

$$\min_{\boldsymbol{\theta}} \quad \frac{1}{2m}\sum_{i=1}^{m}\left\|\boldsymbol{x}_i^{\mathcal{D}} - \widehat{\boldsymbol{x}}_i^{\mathcal{D}}\right\|^2 + \alpha\cdot \mathrm{KL}(p\|p^l) + \beta\cdot \widehat{D}_{\mathcal{H}}^{2}[f_\theta(\boldsymbol{x}^{\mathrm{s}}), f_\theta(\boldsymbol{x}^{\mathrm{t}})], \tag{27}$$

where $\boldsymbol{x}_i^{\mathcal{D}} = \{\boldsymbol{x}_i^{\mathrm{s}}, \boldsymbol{x}_i^{\mathrm{t}}\}$ is the cross-domain samples. In Eq. (27), the second term encourages to get the sparse features, and the third term minimizes the distribution discrepancy of the represented cross-domain features. Depending on the greedy layer-wise pre-training, the stacked AE is able to extract deep-layer features subject to similar distribution, and further realize the transfer scenarios of motor bearings [424, 425], gears [424], and virtual car body-side production line [431]. The regularization terms of MMD can also be used in training CNN to construct the end-to-end diagnosis models [426, 427]. Among them, Yang et al. [26] introduced multi-layer MMD into the optimization of CNN-based diagnosis model to realize the TDM transfer scenarios of transferring diagnosis knowledge from the laboratory-used motor bearings to the locomotive bearings, which was one of the earliest work in this scenario. The proposed CNN-based diagnosis model was trained by

$$\min_{\boldsymbol{\theta}} \qquad \frac{1}{n_{\mathrm{s}}}\sum_{i=1}^{n_{\mathrm{s}}} J(\boldsymbol{y}_i^{\mathrm{s}}, \widehat{\boldsymbol{y}}_i^{\mathrm{s}}) + \alpha\cdot\frac{1}{n_{\mathrm{t}}}\sum_{j=1}^{n_{\mathrm{t}}} J(\widetilde{\boldsymbol{y}}_i^{\mathrm{t}}, \widehat{\boldsymbol{y}}_i^{\mathrm{t}}) + \beta\cdot \widehat{D}_{\mathcal{H}}^{2}(\mathcal{Z}^{\mathrm{s},\mathcal{L}}, \mathcal{Z}^{\mathrm{t},\mathcal{L}}), \tag{28}$$

where the first term was used to train a domain-shared classifier by labeled source domain samples, the second term was expected to minimize the error between the predicted labels $\hat{\boldsymbol{y}}_i^{\mathrm{t}}$ and pseudo labels $\tilde{\boldsymbol{y}}_i^{\mathrm{t}}$ for target-domain samples, which could improve the diagnosis accuracy on the target domain by reducing the among-class distance of the learned features, and the last term was to adapt the distribution of the multi-layer features $\{\mathcal{Z}^{\mathrm{s},\mathcal{L}}, \mathcal{Z}^{\mathrm{t},\mathcal{L}}\}$ both in the convolutional layers and the full-connected layers. Considering the difficulties in determine the Gaussian kernel parameters, Yang et al. [417] introduced multi-kernel MMD into the optimization of CNN-based diagnosis model, and it could be trained by solving a convex optimization problem as

$$\begin{aligned} &\min_{\boldsymbol{\theta}} \max_{k\in\mathcal{K}} \quad \frac{1}{n_{\mathrm{s}}}\sum_{i=1}^{n_{\mathrm{s}}} J(\boldsymbol{y}_i^{\mathrm{s}}, \hat{\boldsymbol{y}}_i^{\mathrm{s}}) + \alpha \cdot \frac{1}{n_{\mathrm{t}}}\sum_{j=1}^{n_{\mathrm{t}}} J(\tilde{\boldsymbol{y}}_i^{\mathrm{t}}, \hat{\boldsymbol{y}}_i^{\mathrm{t}}) + \beta \cdot \widehat{D}^2_{\mathcal{H}\sim\mathcal{K}}(\boldsymbol{x}^{\mathrm{s},\mathrm{F}_2}, \boldsymbol{x}^{\mathrm{t},\mathrm{F}_2}) \\ &\mathrm{s.t.} \quad \mathcal{K} := \left\{ k \,\middle|\, k = \sum_{u=1}^{U} \beta_u k_u, \sum_{u=1}^{U} \beta_u = 1, \beta_u \geq 0, \forall u \in \{1,2,\cdots,U\} \right\} \end{aligned} \tag{29}$$

where the weighted sum of the U-kernel MMD of the learned features was used to estimate the cross-domain discrepancy. In terms of the weaknesses of Gaussian kernel-based MMD, Yang et al [443] further proposed the polynomial kernel induced MMD to improve the transfer performance of the deep transfer learning-based diagnosis models, and the diagnosis results were higher and more robust to kernel parameters than the conventional-used Gaussian kernel induced MMD. In addition to MMD, other distance metrics are used to construct deep transfer learning-based diagnosis models. Qian et al. [429] used high-order Kullback-Leibler divergence to measure the distribution discrepancy, and further train a sparse filter-based diagnosis model for gears. Ref. [428] adapted the distribution of learned cross-domain features by reducing the center distance, which attempts for fault diagnosis of motor bearings under varying operation conditions. Wang et al. [430] proposed a deep transfer learning-based diagnosis model for power plant thermal system, in which CORAL was employed to align the covariance of the cross-domain features.

Deep transfer learning-based diagnosis models are useful to correct serious cross-domain discrepancy based on the theories of deep learning and transfer learning, which has been widely concerned for transfer scenarios in IFD. In these models, the transfer results relate to the distance metric to the distribution discrepancy. The diagnosis models may obtain poor transfer results on the target domain if the distances are unable to adequately describe the discrepancy.

(4) Other approaches

In Ref. [432], the authors found a low-dimensional latent space by the transfer factor analysis, which helped select the cross-domain features with small discrepancy. Zhang et al. [433] mapped the cross-domain samples into two d-dimensional subspaces by the principal component analysis, and the subspaces are aligned to minimize the cross-domain discrepancy. Zheng et al. [434] proposed a diagnosis model for bearing fault diagnosis, which could fuse the diagnosis knowledge from multiple operation conditions and further complete the diagnosis tasks on another condition.

4.3.2. GAN-based approaches

Generally, GAN consists of the generative model $G(\cdot)$ and the discriminative model $D(\cdot)$ [442]. The former acquires the distribution information of the target-domain samples, and further generates fake samples with the similar distribution to the target-domain samples when inputting random noise or the source-domain samples. The later focuses on training the parameters of the generative model to make the generated fake samples undistinguished from the actual samples in the target domain. The generative and discriminative models are gamed with each other in GAN, which is defined as

$$\min_{G}\max_{D} \quad E_{\boldsymbol{x}^{\mathrm{s}}\in X^{\mathrm{s}}}[\log D(G(\boldsymbol{x}^{\mathrm{s}}))] + E_{\boldsymbol{x}^{\mathrm{t}}\in X^{\mathrm{t}}}[\log(1 - D(G(\boldsymbol{x}^{\mathrm{t}})))]. \tag{30}$$

In the field of IFD, Xie et al. [435] proposed the cycle-consistent GAN for fault diagnosis of bearings under different operation conditions, in which the fake samples were generated according to the samples from other operation conditions. Li et al. [436] constructed multiple generative models for fault diagnosis of bearings under varying operation conditions, and the fake samples were generated by the sample from one fault. Han et al [437] presented a deep adversarial CNN for machines from one operation condition to another. Guo et al. [35] proposed the diagnosis models based on the adversarial learning for fault diagnosis among different bearings, which was one of the earliest works for TDM scenarios. They constructed a diagnosis model including condition recognition and domain adaptation, which was trained by the following optimization objective.

$$\begin{aligned}\min_{\boldsymbol{\theta}} \quad & \frac{1}{n_{\mathrm{s}}}\sum_{i=1}^{n_{\mathrm{s}}} J(\boldsymbol{y}_i^{\mathrm{s}}, \hat{\boldsymbol{y}}_i^{\mathrm{s}}) + \beta \cdot \widehat{D}_{\mathcal{H}}^2(\boldsymbol{x}^{\mathrm{s},\mathrm{F}_2}, \boldsymbol{x}^{\mathrm{t},\mathrm{F}_2}) \\ & - \left[\frac{1}{n_{\mathrm{s}}}\sum_{i=1}^{n_{\mathrm{s}}} logD\left(\boldsymbol{x}_i^{\mathrm{s},\mathrm{F}_2}\right) + \frac{1}{n_{\mathrm{t}}}\sum_{j=1}^{n_{\mathrm{t}}} \log\left(1 - D\left(\boldsymbol{x}_j^{\mathrm{t},\mathrm{F}_2}\right)\right)\right].\end{aligned} \tag{31}$$

By minimizing the third regularization term in Eq. (31), the diagnosis model could not distinguish the domain classes of the input cross-domain samples. Consequently, the model was able to serve well both in the source and target domains.

GAN-based approaches are able to generate fake labeled data that are similar to the actual data in the target domain. These fake data are able to help train reliable diagnosis models for the target domain. Therefore, such approaches are not affected by the condition whether labeled samples in the target domain are available, and they are expected to work for both the TIM scenarios and the TDM scenarios.

4.3.3. Instance-based approaches

In instance-based approaches, it is assumed that minority of labeled samples are labeled in the target domain, which is insufficient to train reliable diagnosis models. Therefore, the purpose of instance-based approaches focuses on using the related samples in the source domain to improve the performance of the diagnosis model $f_{\mathrm{t}}(\cdot)$ in the target domain [27]. TrAdaboost [31] is a typical instance-based approach, which derives from the Adaboost algorithm. In the method, the source and target domain samples are weighted to balance their contributions of training the diagnosis models, as shown in Fig. 16. If the given diagnosis model misclassifies a sample in the target domain, the weights for this sample will be enhanced. Furthermore, the weights of the misclassified samples in the source domain will be reduced. As a result, the decision boundary will be enforced towards the orientation of correctly classify the target-domain samples.

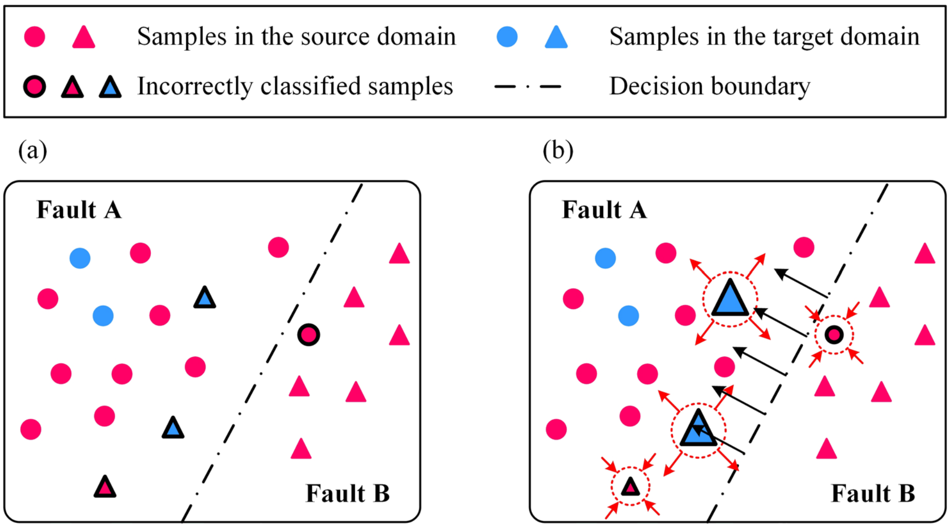


Fig. 16. Illustration of the TrAdaboost algorithm: (a) directly train the diagnosis models with the source and target domain samples, and (b) train the diagnosis models by using TrAdaboost algorithm.

Based on the architecture of TrAdaboost, Shen et al. [36] used the cross-domain features to train a set of *k*NN-based diagnosis models by TrAdaboost algorithm, which helped recognize the health states of bearings under varying operation conditions. In this method, the singular value decomposition was first employed to extract fault features from the monitoring data of motor bearings. After that, a metric to vector angle cosine was designed to select the features that were subject to small cross-domain discrepancy. Finally, the features from the source and target domains were used to train a *k*NN-based diagnosis model by TrAdaboost algorithm. The results showed that the trained diagnosis model could correctly recognize the unlabeled samples in the target domain although there were a small number of labeled samples in the target domain.

Instance-based approaches are easily-implemented for transfer scenarios. However, the transfer performance of them relates to the number of target-domain samples. Furthermore, they just are able to correct discrepancy in TIM scenarios, but incapable to implement transfer scenarios with serious discrepancy, such as TDM scenarios, because these approaches may lack the powerful capability of data fitting.

4.3.4. Parameter-based approaches

Similar to the instance-based approaches, parameter-based approaches also assume that there are minority of labeled samples in the target domain [27]. In the approaches, the parameters of diagnosis models are pre-trained with the source-domain samples. After that, the parameters of the pre-trained diagnosis models are saved and further reassigned to the diagnosis models served in the target domain. Finally, the minority of labeled samples in the target domain are used to fine-tune the target diagnosis models.

Researchers have developed the parameter-based approaches for transfer scenarios. Zhang et al. [438] and Hasan et al. [37] aimed at transfer scenarios of motor bearings subject to varying operation conditions, in which the pre-trained diagnosis model was fine-tuned by the samples from the target operation conditions. Compared with the diagnosis model just trained with a small number of target-domain samples, the fine-tuned diagnosis model presented the faster convergence rate and higher diagnosis accuracy. Furthermore, some researchers employed the pre-trained model for image identification to perform fault diagnosis tasks of machines. Cao et al. [439] proposed a deep CNN-based diagnosis model for fault diagnosis of gearboxes. The authors constructed a CNN model with 24 layers, and it was trained with the labeled images from the well-known datasets of ImageNet. Then, the well-trained parameters were used to initialize the other CNN-based diagnosis model subject to the same architecture as the pre-trained one. The diagnosis model was finally trained with the collected vibration data that were converted as the 2D image format. Shao et al. [440] presented a parameter-based approach for fault diagnosis of motor bearings and gearboxes. In this approach, the well-known VGG-16 network was expected to complete the diagnosis tasks of machines under different working conditions. The collected vibration data were first converted to the 2D time-frequency images by wavelet transform. Then, the vibration data in image format were used to fine-tune the pre-trained VGG-16, in which three convolutional blocks in the bottom were frozen.

Parameter-based approaches fine-tune the pre-trained diagnosis models for fault diagnosis tasks in the target domain, which consumes less computation resources in handling massive data. However, the transfer performance of these approaches mostly depends on the number of labeled samples in the target domain.

4.4. Epilog

Transfer learning is promising to expand IFD from academic research to engineering scenarios. The transfer problems of IFD are first defined, and the transfer scenarios are further categorized into TIM and TDM in this section. Aiming at the transfer diagnosis scenarios, a few exploratory studies are divided into feature-based approaches, GAN-based approaches, instance-based approaches, and parameter-based approaches. Among them, feature-based approaches and GAN-based approaches are widely concerned because they can accomplish TDM scenarios subject to serious cross-domain discrepancy. In contrast, the instance-based approaches and parameter-based approaches are easily-implemented, and mostly focus on TIM scenarios subject to slight cross-domain discrepancy. Furthermore, the transfer performance of these approaches may relate to the number of labeled samples in the target domain.

5. Discussions: Future challenges and a roadmap in IFD

Following the development of machine learning theories, IFD gradually releases the contribution of the human labor and automatically recognizes the health states of machines from the past up to the present, and its applications will serve to the engineering scenarios in the future period. At the end of this review, we attempt to picture the roadmap and discuss the future challenges of IFD, as shown in Fig. 17, which is expected to inspire the readers to orientate the potential trends of this field over the next five or ten years.

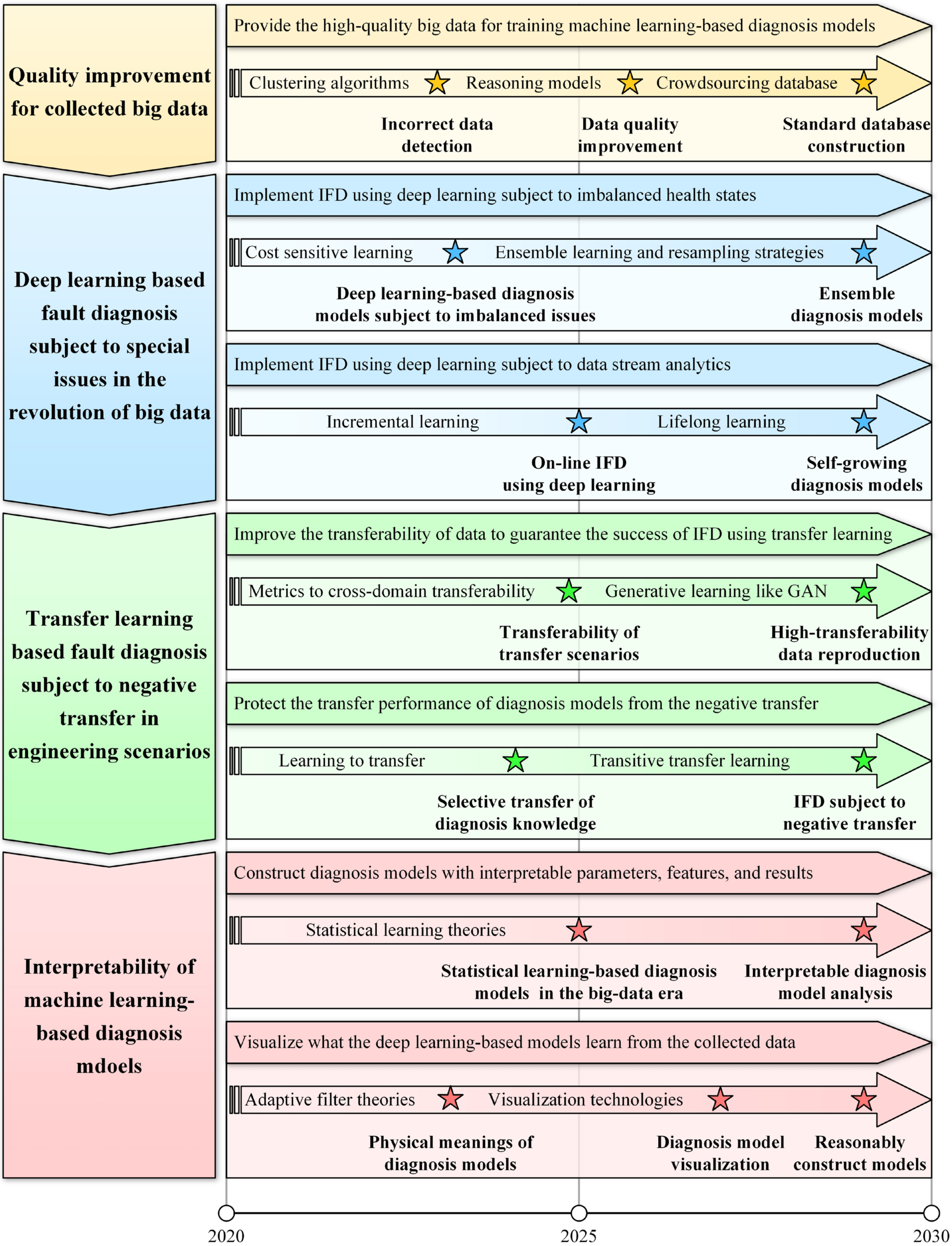


Fig. 17. Roadmap of applications of machine learning to machine fault diagnosis.

(1) How to provide the high-quality big data for training diagnosis models based on machine learning?

In the present period of IFD, the volume of the collected data is rapidly grown than ever before. However, the quality of the collected data is not always satisfactory because a portion of them might be subject to the poor quality. The poor-quality data are defined as the incorrect data with inaccurate, uncertain, incomplete, and low-timeless. Lots of factors may lead to them, such as the disturbance of working environments, the anomaly of data acquisition devices, and the interruption of data transmission. The incorrect data will result in unreliable diagnosis results when these data are directly used to train the diagnosis models by machine learning. Therefore, it is necessary to develop effective methods to clean incorrect data and improve the quality of the collected big data. The clustering algorithms and reasoning models may help separate anomaly data and further improve the quality of collected data. Furthermore, the crowdsourcing database technologies could manage the big data with low value density and improve the data quality, which help construct the standard database to provide the high-quality data for training machine learning-based diagnosis models.

(2) How to construct deep learning-based diagnosis models subject to special issues in the revolution of big data?

With the revolution of big data, two special issues come to bring negative effects on the IFD using deep learning theories, i.e., imbalanced health states and analytics for data stream. 1) Imbalanced distribution in health states of machines is a common phenomenon in engineering scenarios. For example, the data collected from the healthy state are far more sufficient than those from the faults. If the imbalanced data are used to train the deep learning-based diagnosis models, the decision boundary of the models might be enforced to shift towards the health states with minority of instances, thus the diagnosis accuracy would be reduced. To overcome the issue, the cost sensitive learning may help construct diagnosis models subject to imbalanced health states. In addition, ensemble learning, such as Adaboost and XGboost, is expected to improve the performance of deep learning-based diagnosis models when combined with the resampling strategies. 2) The monitoring data stream is regard as the increasingly-enlarged data with the continuous time series. By analyzing that, IFD is possible to provide the real-time diagnosis results. Such issue has been concerned by the researchers for years, but the unreliable data transmission and the bandwidth limitations prevent the monitoring data stream from arriving the destination in an unbroken sequence. Furthermore, the inefficient computation capability impedes the success of data stream analytics. As a result, the IFD using deep learning is just implemented based on the off-line historical data. Fortunately, the monitoring data stream is able to be collected and efficiently handled with the advent of IoT, broadband internet, and cloud computing. Therefore, the on-line IFD is encouraged to be developed to make real-time decisions on the incipient anomaly or the sudden faults of machines. The incremental learning is expected to facilitate the on-line IFD using deep learning. Furthermore, the lifelong learning may promote deep learning-based diagnosis models to constantly acquire the diagnosis knowledge from the monitoring data stream.

(3) How to protect the performance of transfer learning-based diagnosis models from the negative transfer in engineering scenarios?

The successes of transfer learning are subject to the assumption of related health information across multiple diagnosis tasks. If the assumption is invalid, the negative transfer may happen to reduce the transfer performance of the transfer learning-based diagnosis models by using the diagnosis knowledge from the source domain. Two reasons may cause the negative transfer. 1) It is possible to collect data from multiple source domains to match a common target domain, but the health information contained in these source domains is not all related to the that in the target domain in engineering scenarios. For example, the data from motors with different specifications can be collected to form multiple source domains. Some of them could provide positive diagnosis knowledge for a generator with similar physical construction to the motors, but not all of them. Consequently, it confuses the researchers to select the optimal one and guarantee the performance of the transfer learning-based diagnosis models. Therefore, it is necessary to develop metrics to cross-domain transferability, which may help select the relevant source domains. Furthermore, GAN could help generate fake high-transferability data to extend the available data for IFD using transfer learning. 2) The constructed transfer learning-based diagnosis models are incapable to extract the related health information. According to Section 4.3, several approaches can be used in IFD, but the transfer performance of them is different for a given transfer scenario due to diverse performance ceilings. To select the optimal one and guarantee the transfer success, it must take much time by experimental trials. For this issue, the architecture of learning to transfer is expected to automatically select the optimal approaches with the highest performance based on the historical knowledge. Furthermore, with the help of transitive transfer learning, the diagnosis knowledge of the source domain may be reused in the target domain if the negative transfer inevitably happens in engineering scenarios.

(4) How to improve the interpretability of the deep learning-based diagnosis models?

Although there are state-of-the-art achievements of deep learning-based diagnosis models, an open issue of black box for deep hierarchical networks still confuses the academic researchers. It is difficult to acquire how these models learn the diagnosis knowledge from the monitoring data. As a result, the diagnosis models are constructed by experimental trials once and once again rather than the strictly theoretical background. To bridge the gap, two research interests are recommended to be concerned. 1) The deep learning-based diagnosis models are trained by minimizing empirical risk, which really lacks rigorous theory. As a result, the physical meanings of them are difficult to be interpretable. The statistical learning theories, such as SVM and PGM, have rigorous theory grounds, which promote to constructed diagnosis models with easily-understood model parameters, features, and diagnosis results. Therefore, it is still worthy to investigate statistical learning in IFD with the revolution of big data. 2) The process of learning features through deep learning is similar to the filtering process. Therefore, the adaptive filter theory is beneficial to analyzing the physical meanings of deep learning-based diagnosis models. In addition, the visualization technologies, such as maximizing the activation and feature inversion, are expected to visually express what the diagnosis models learn from the input data. According to the above research, the interpretability results are helpful to construct the deep learning-based diagnosis models with the optimal architectures reasonably.

6. Conclusions

In this paper, we review the applications of machine learning to machine fault diagnosis, which is roughly divided into three periods. In the past, IFD is implemented by the steps of data collection, artificial feature extraction, and health state recognition. By using traditional machine learning theories, the diagnosis models are able to automatically recognize the health states of machines. However, the literature review presents that the artificial feature extraction still relies on the expert knowledge. With the rapid development of machine learning over the recent years, the advent of deep learning brings positive effects on the enhanced benefits. The deep learning theories construct end-to-end diagnosis models to automatically learn features from the collected data, and subsequently recognize the health states of machines. It should be concerned that the successes of deep learning-based diagnosis models are subject to enough labeled samples. Such assumption is unpractical in engineering scenarios. To bridge the gap, transfer learning theories are promising to construct diagnosis models, in which the diagnosis knowledge can be transferred across multiple diagnosis tasks. Finally, we discuss the challenges of IFD and further picture a roadmap. This review is expected to systematically present the development of IFD and provide the valuable guidelines of the future research in this field.

Acknowledgements

This research was supported by NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization (U1709208), National Natural Science Foundation of China (61673311), and National Key R&D Program of China (2018YFB1306100).

References

[1] Y. Lei, Intelligent fault diagnosis and remaining useful life prediction of rotating machinery, Butterworth-Heinemann Elsevier Ltd, Oxford, 2016.

[2] X. Dai, Z. Gao, From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis, IEEE Trans. Ind. Inform. 9 (2013) 2226-2238.

[3] A. Stetco, F. Dinmohammadi, X. Zhao, V. Robu, D. Flynn, M. Barnes, J. Keane, G. Nenadic, Machine learning methods for wind turbine condition monitoring: A review, Renewable Energy 133 (2019) 620-635.

[4] D.T. Hoang, H.J. Kang, A survey on deep learning based bearing fault diagnosis, Neurocomputing 335 (2019) 327-335.

[5] Z. Gao, C. Cecati, S.X. Ding, A survey of fault diagnosis and fault-tolerant techniques-part II: Fault diagnosis with knowledge-based and hybrid/active approaches, IEEE Trans. Ind. Electron. 62 (2015) 3768-3774.

[6] L. Duan, M. Xie, J. Wang, T. Bai, Deep learning enabled intelligent fault diagnosis: Overview and applications, J. Intell. Fuzzy Syst. 35 (2018) 5771-5784.

[7] R. Liu, B. Yang, E. Zio, X. Chen, Artificial intelligence for fault diagnosis of rotating machinery: A review, Mech. Syst. Signal Process. 108 (2018) 33-47.

[8] S. Khan, T. Yairi, A review on the application of deep learning in system health management, Mech. Syst. Signal Process. 107 (2018) 241-265.

[9] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mech. Syst. Signal Process. 115 (2019) 213-237.

[10] E. Golge, Brief history of machine learning, 2016, Available: http://www.erogol.com/brief-history-machine-learning/.

[11] J. Schmidhuber, Deep learning in neural networks: An overview, Neural Networks 61 (2015) 85-117.

[12] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273-297.

[13] T.M. Cover, P.E. Hart, Nearest neighbor pattern classification, IEEE Trans. Inf. Theory 13 (1967) 21-27.

[14] D. Koller, N. Friedman, Probabilistic graphical models: Principles and techniques, MIT Press, Cambridge, Massachusetts, 2009.

[15] S. Liao, Expert system methodologies and applications - a decade review from 1995 to 2004, Expert Syst. Appl. 28 (2005) 93-103.

[16] G.M. Knapp, H.P. Wang, Machine fault classification - a neural network approach, Int. J. Prod. Res. 30 (1992) 811-823.

[17] A. Widodo, B.S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, Mech. Syst. Signal Process. 21 (2007) 2560-2574.

[18] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436-444.

[19] Y. Bengio, Learning deep architectures for AI, Foundations and Trends in Machine Learning 2 (2009) 1-127.

[20] I. Goodfellow, Y. Bengio, Deep learning, MIT Press, Cambridge, Massachusetts, 2016.

[21] D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning representations by back-propagating errors, Nature 323 (1986) 533-536.

[22] G.E. Hinton, Training products of experts by minimizing contrastive divergence, Neural Comput. 14 (2002)

1771-1800.

[23] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, Science 313 (2006) 504-507.

[24] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Commun. ACM 60 (2017) 84-90.

[25] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, USA, 2016, pp. 770-778.

[26] B. Yang, Y. Lei, F. Jia, S. Xing, An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings, Mech. Syst. Signal Process. 122 (2019) 692-706.

[27] S.J. Pan, Q. Yang, A survey on transfer learning, IEEE Trans. Knowl. Data Eng. 22 (2010) 1345-1359.

[28] S. Thrun, L. Pratt, Learning to learn: Introduction and overview, Springer Ltd, Boston, MA, 1998.

[29] S.J. Pan, I.W. Tsang, J.T. Kwok, Q. Yang, Domain adaptation via transfer component analysis, IEEE Trans. Neural Netw. 22 (2011) 199-210.

[30] M. Long, J. Wang, G. Ding, J. Sun, P.S. Yu, Transfer feature learning with joint distribution adaptation, in: *IEEE International Conference on Computer Vision*, Sydney, Australia, 2013, pp. 2200-2207.

[31] W. Dai, Q. Yang, G. Xue, Y. Yu, Boosting for transfer learning, in: *International Conference on Machine learning*, Corvalis, USA, 2007, pp. 193-200.

[32] H. Venkateswara, S. Chakraborty, S. Panchanathan, Deep-learning systems for domain adaptation in computer vision learning transferable feature representations, IEEE Signal Process. Mag. 34 (2017) 117-129.

[33] M. Long, J. Wang, Y. Cao, J. Sun, P.S. Yu, Deep learning of transferable representation for scalable domain adaptation, IEEE Trans. Knowl. Data Eng. 28 (2016) 2027-2040.

[34] M. Long, H. Zhu, J. Wang, M.I. Jordan, Deep transfer learning with joint adaptation networks, in: *IEEE International Conference on Machine Learning*, Sydney, Australia, 2017, pp. 2208-2217.

[35] L. Guo, Y. Lei, S. Xing, T. Yan, N. Li, Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data, IEEE Trans. Ind. Electron. 66 (2019) 7316-7325.

[36] F. Shen, C. Chen, R. Yan, R.X. Gao, Bearing fault diagnosis based on svd feature extraction and transfer learning classification, in: *Prognostics and System Health Management Conference*, Beijing, China, 2015, pp. 1-6.

[37] M.J. Hasan, J.M. Kim, Bearing fault diagnosis under variable rotational speeds using stockwell transform-based vibration imaging and transfer learning, Appl. Sci.-Basel 8 (2018) 2357.

[38] J. Chen, Z. Li, J. Pan, G. Chen, Y. Zi, J. Yuan, B. Chen, Z. He, Wavelet transform based on inner product in fault diagnosis of rotating machinery: A review, Mech. Syst. Signal Process. 70-71 (2016) 1-35.

[39] Y. Wang, J. Xiang, R. Markert, M. Liang, Spectral kurtosis for fault detection, diagnosis and prognostics of rotating machines: A review with applications, Mech. Syst. Signal Process. 66-67 (2016) 679-698.

[40] Z. Qiao, Y. Lei, N. Li, Applications of stochastic resonance to machinery fault detection: A review and tutorial, Mech. Syst. Signal Process. 122 (2019) 502-536.

[41] A. Rai, S.H. Upadhyay, A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings, Tribol. Int. 96 (2016) 289-306.

[42] Y. Lei, J. Lin, M.J. Zuo, Z. He, Condition monitoring and fault diagnosis of planetary gearboxes: A review, Measurement 48 (2014) 292-305.

[43] D. Goyal, Vanraj, B.S. Pabla, S.S. Dhami, Condition monitoring parameters for fault diagnosis of fixed axis gearbox: A review, Arch. Comput. Methods Eng. 24 (2017) 543-556.

[44] A. Widodo, B.S. Yang, E.Y. Kim, A.C.C. Tan, J. Mathew, Fault diagnosis of low speed bearing based on acoustic emission signal and multi-class relevance vector machine, Nondestr. Test. Eval. 24 (2009) 313-328.

[45] D.H. Pandya, S.H. Upadhyay, S.P. Harsha, Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using apf-knn, Expert Syst. Appl. 40 (2013) 4137-4145.

[46] C. Li, R.V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, R.E. Vasquez, Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals, Mech. Syst. Signal Process. 76-77 (2016) 283-293.

[47] P.K. Wong, J. Zhong, Z. Yang, C.M. Vong, A new framework for intelligent simultaneous-fault diagnosis of rotating machinery using pairwise-coupled sparse bayesian extreme learning committee machine, Proc. Inst. Mech. Eng. Part C-J. Mech. Eng. Sci. 231 (2017) 1146-1161.

[48] J. Yoon, D. He, Planetary gearbox fault diagnostic method using acoustic emission sensors, IET Sci. Meas. Technol. 9 (2015) 936-944.

[49] T.R. Lin, A.C.C. Tan, L. Ma, J. Mathew, Condition monitoring and fault diagnosis of diesel engines using instantaneous angular speed analysis, Proc. Inst. Mech. Eng. Part C-J. Mech. Eng. Sci. 229 (2015) 304-315.

[50] J. Yang, L. Pu, Z. Wang, Y. Zhou, X. Yan, Fault detection in a diesel engine by analysing the instantaneous angular speed, Mech. Syst. Signal Process. 15 (2001) 549-564.

[51] M. Zhang, Y. Zi, L. Niu, S. Xi, Y. Li, Intelligent diagnosis of V-type marine diesel engines based on multifeatures extracted from instantaneous crankshaft speed, IEEE Trans. Instrum. Meas. 68 (2019) 722-740.

[52] A. Naha, A.K. Samanta, A. Routray, A.K. Deb, Low complexity motor current signature analysis using sub-nyquist strategy with reduced data length, IEEE Trans. Instrum. Meas. 66 (2017) 3249-3259.

[53] M.S. Safizadeh, B. Yari, Pump cavitation detection through fusion of support vector machine classifier data associated with vibration and motor current signature, Insight 59 (2017) 669-673.

[54] T. Yang, H. Pen, Z. Wang, C. Chang, Feature knowledge based fault detection of induction motors through the analysis of stator current data, IEEE Trans. Instrum. Meas. 65 (2016) 549-558.

[55] Z. Duan, T. Wu, S. Guo, T. Shao, R. Malekian, Z. Li, Development and trend of condition monitoring and fault diagnosis of multi-sensors information fusion for rolling bearings: A review, Int. J. Adv. Manuf. Technol. 96 (2018) 803-819.

[56] Y. Lei, Z. He, Y. Zi, Q. Hu, Fault diagnosis of rotating machinery based on multiple ANFIS combination with gas, Mech. Syst. Signal Process. 21 (2007) 2280-2294.

[57] Y. Lei, M.J. Zuo, Z. He, Y. Zi, A multidimensional hybrid intelligent method for gear fault diagnosis, Expert Syst. Appl. 37 (2010) 1419-1430.

[58] V. Bolon-Canedo, N. Sanchez-Marono, A. Alonso-Betanzos, A review of feature selection methods on synthetic data, Knowl. Inf. Syst. 34 (2013) 483-519.

[59] K. Zhang, Y. Li, P. Scarf, A. Ball, Feature selection for high-dimensional machinery fault diagnosis data using multiple models and radial basis function networks, Neurocomputing 74 (2011) 2941-2952.

[60] X. Zhang, Q. Zhang, M. Chen, Y. Sun, X. Qin, H. Li, A two-stage feature selection and intelligent fault diagnosis method for rotating machinery using hybrid filter and wrapper method, Neurocomputing 275 (2018) 2426-2439.

[61] F. Jiang, Z. Zhu, W. Li, Y. Ren, G. Zhou, Y. Chang, A fusion feature extraction method using EEMD and correlation coefficient analysis for bearing fault diagnosis, Appl. Sci.-Basel 8 (2018) 1621.

[62] M. Gerdes, D. Galar, D. Scholz, Decision trees and the effects of feature extraction parameters for robust sensor network design, Eksploat. Niezawodn. 19 (2017) 31-42.

[63] Y. Li, Y. Yang, G. Li, M. Xu, W. Huang, A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mrmr feature selection, Mech. Syst. Signal Process. 91 (2017) 295-312.

[64] M. Singh, A.G. Shaik, Faulty bearing detection, classification and location in a three-phase induction motor based on stockwell transform and support vector machine, Measurement 131 (2019) 524-533.

[65] Y. Lei, M.J. Zuo, Gear crack level identification based on weighted k nearest neighbor classification algorithm, Mech. Syst. Signal Process. 23 (2009) 1535-1547.

[66] K. Kira, L.A. Rendell, A practical approach to feature selection, in: *International Workshop on Machine*

Learning, Aberdeen, UK, 1992, pp. 249-256.

[67] I. Kononenko, Estimating attributes: Analysis and extensions of Relief, in: *European conference on machine learning*, Catania, Italy, 1994, pp. 171-82.

[68] M.A. Hall, L.A. Smith, Practical feature subset selection for machine learning, in: *Australasian Computer Science Conference*, Perth, Australia, 1998, pp. 181-191.

[69] H. Peng, F. Long, C. Ding, Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 1226-1238.

[70] S.C. Kothari, H. Oh, Neural networks for pattern recognition, Butterworth-Heinemann Elsevier Ltd, Oxford, 1993.

[71] B.S. Yang, T. Han, J. An, ART-KOHONEN neural network for fault diagnosis of rotating machinery, Mech. Syst. Signal Process. 18 (2004) 645-657.

[72] B.S. Yang, K.J. Kim, Application of Dempster-Shafer theory in fault diagnosis of induction motors using vibration and current signals, Mech. Syst. Signal Process. 20 (2006) 403-420.

[73] H. Liu, R. Setiono, Feature selection and classification - a probabilistic wrapper approach, in: *International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, Fukuoka, Japan, 1996, pp. 419-424.

[74] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, K. Knight, Sparsity and smoothness via the fused lasso, J. R. Stat. Soc. Ser. B-Stat. Methodol. 67 (2005) 91-108.

[75] A.N. Tikhonov, A.V. Goncharsky, V.V. Stepanov, A.G. Yagola, Numerical methods for the solution of ill-posed problems, Springer Ltd, Boston, MA, 1995.

[76] M.F. White, Expert systems for fault diagnosis of machinery, Measurement 9 (1991) 163-171.

[77] S. Sahin, M.R. Tolun, R. Hassanpour, Hybrid expert systems: A survey of current approaches and applications, Expert Syst. Appl. 39 (2012) 4609-4617.

[78] M. Krishnamurthi, D.T. Phillips, An expert system framework for machine fault-diagnosis, Comput. Ind. Eng. 22 (1992) 67-84.

[79] H.L. Gelgele, K.S. Wang, An expert system for engine fault diagnosis: Development and application, J. Intell. Manuf. 9 (1998) 539-545.

[80] C. Angeli, An online expert system for fault diagnosis in hydraulic systems, Expert Syst. 16 (1999) 115-120.

[81] S. Ebersbach, Z. Peng, Expert system development for vibration analysis in machine condition monitoring, Expert Syst. Appl. 34 (2008) 291-299.

[82] B.S. Yang, D.S. Lim, A.C.C. Tan, Vibex: An expert system for vibration fault diagnosis of rotating machinery using decision tree and decision table, Expert Syst. Appl. 28 (2005) 735-742.

[83] H.J. Lee, D.Y. Park, B.S. Ahn, Y.M. Park, J.K. Park, S.S. Venkata, A fuzzy expert system for the integrated fault diagnosis, IEEE Trans. Power Delivery 15 (2000) 833-838.

[84] S. Liu, S. Liu, An efficient expert system for machine fault diagnosis, Int. J. Adv. Manuf. Technol. 21 (2003) 691-698.

[85] J. Wu, Y. Wang, M.S.R. Bai, Development of an expert system for fault diagnosis in scooter engine platform using fuzzy-logic inference, Expert Syst. Appl. 33 (2007) 1063-1075.

[86] T. Berredjem, M. Benidir, Bearing faults diagnosis using fuzzy expert system relying on an improved range overlaps and similarity method, Expert Syst. Appl. 108 (2018) 134-142.

[87] J. Wu, P. Chiang, Y. Chang, Y. Shiao, An expert system for fault diagnosis in internal combustion engines using probability neural network, Expert Syst. Appl. 34 (2008) 2704-2713.

[88] J. Wu, C. Liu, An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network, Expert Syst. Appl. 36 (2009) 4278-4286.

[89] A. Hajnayeb, S.E. Khadem, M.H. Moradi, Design and implementation of an automatic condition-monitoring expert system for ball-bearing fault detection, Ind. Lubr. Tribol. 60 (2008) 93-100.

[90] P. Jayaswal, S.N. Verma, A.K. Wadhwani, Development of EBP-artificial neural network expert system for rolling element bearing fault diagnosis, J. Vibrat. Control 17 (2011) 1131-1148.

[91] R.A. Vingerhoeds, P. Janssens, B.D. Netten, Enhancing off-line and online condition monitoring and fault-diagnosis, Control Eng. Pract. 3 (1995) 1515-1528.

[92] A. Varma, N. Roddy, Icarus: Design and deployment of a case-based reasoning system for locomotive diagnostics, Eng. Appl. Artif. Intell. 12 (1999) 681-690.

[93] H. Wu, Y. Liu, Y. Ding, Y. Qiu, Fault diagnosis expert system for modern commercial aircraft, Aircraft Eng. Aerospace Technol. 76 (2004) 398-403.

[94] C.M. Vong, P.K. Wong, W.F. Ip, Case-based expert system using wavelet packet transform and kernel-based feature manipulation for engine ignition system diagnosis, Eng. Appl. Artif. Intell. 24 (2011) 1281-1294.

[95] B. Merainani, C. Rahmoune, D. Benazzouz, B. Ould-Bouamama, A novel gearbox fault feature extraction and classification using hilbert empirical wavelet transform, singular value decomposition, and som neural network, J. Vibrat. Control 24 (2018) 2512-2531.

[96] M.L.D. Wong, L.B. Jack, A.K. Nandi, Modified self-organising map for automated novelty detection applied to vibration signal monitoring, Mech. Syst. Signal Process. 20 (2006) 593-610.

[97] X. Chen, J. Zhou, H. Xiao, E. Wang, J. Xiao, H. Zhang, Fault diagnosis based on comprehensive geometric characteristic and probability neural network, Appl. Math. Comput. 230 (2014) 542-554.

[98] B. Zhong, J. MacIntyre, Y. He, J. Tait, High order neural networks for simultaneous diagnosis of multiple faults in rotating machines, Neural Comput. Appl. 8 (1999) 189-195.

[99] M. Barakat, M. El Badaoui, F. Guillet, Hard competitive growing neural network for the diagnosis of small bearing faults, Mech. Syst. Signal Process. 37 (2013) 276-292.

[100] M. Barakat, D. Lefebvre, M. Khalil, F. Druaux, O. Mustapha, Parameter selection algorithm with self adaptive growing neural network classifier for diagnosis issues, Int. J. Mach. Learn. Cybern. 4 (2013) 217-

233.
[101] D.M. Yang, A.F. Stronach, P. MacConnell, J. Penman, Third-order spectral techniques for the diagnosis of motor bearing condition using artificial neural networks, Mech. Syst. Signal Process. 16 (2002) 391-411.
[102] B. Samanta, K.R. Al-Balushi, Artificial neural network based fault diagnostics of rolling element bearings using time-domain features, Mech. Syst. Signal Process. 17 (2003) 317-328.
[103] Y. Yu, D. Yu, J. Cheng, A roller bearing fault diagnosis method based on emd energy entropy and ANN, J. Sound Vibrat. 294 (2006) 269-277.
[104] C. Castejon, O. Lara, J.C. Garcia-Prada, Automated diagnosis of rolling bearings using MRA and neural networks, Mech. Syst. Signal Process. 24 (2010) 289-299.
[105] B. Muruganatham, M.A. Sanjith, B. Krishnakumar, S. Murty, Roller element bearing fault diagnosis using singular spectrum analysis, Mech. Syst. Signal Process. 35 (2013) 150-166.
[106] M. Unal, M. Onat, M. Demetgul, H. Kucuk, Fault diagnosis of rolling bearings using a genetic algorithm optimized neural network, Measurement 58 (2014) 187-196.
[107] J. Zarei, M.A. Tajeddini, H.R. Karimi, Vibration analysis for bearing fault detection and classification using an intelligent filter, Mechatronics 24 (2014) 151-157.
[108] L.F. de Almeida, J.W.P. Bizarria, F.C.P. Bizarria, M.H. Mathias, Condition-based monitoring system for rolling element bearing using a generic multi-layer perceptron, J. Vibrat. Control 21 (2015) 3456-3464.
[109] H. Ahmed, A.K. Nandi, Compressive sampling and feature ranking framework for bearing fault classification with vibration signals, IEEE Access 6 (2018) 44731-44746.
[110] G. Wang, Z. Luo, X. Qin, Y. Leng, T. Wang, Fault identification and classification of rolling element bearing based on time-varying autoregressive spectrum, Mech. Syst. Signal Process. 22 (2008) 934-947.
[111] Y. Lei, Z. He, Y. Zi, Application of an intelligent classification method to mechanical fault diagnosis, Expert Syst. Appl. 36 (2009) 9941-9948.

[112] G.S. Vijay, S.P. Pai, N.S. Sriram, R. Rao, Radial basis function neural network based comparison of dimensionality reduction techniques for effective bearing diagnostics, Proc. Inst. Mech. Eng. Part J-J. Eng. Tribol. 227 (2013) 640-653.

[113] F. Jiang, Z. Zhu, W. Li, G. Zhou, S. Xia, Noise reduction in feature level and its application in rolling element bearing fault diagnosis, Adv. Mech. Eng. 10 (2018) 1-10.

[114] T. Tang, L. Bo, X. Liu, B. Sun, D. Wei, Variable predictive model class discrimination using novel predictive models and adaptive feature selection for bearing fault identification, J. Sound Vibrat. 425 (2018) 137-148.

[115] Y. Lei, Z. He, Y. Zi, Eemd method and wnn for fault diagnosis of locomotive roller bearings, Expert Syst. Appl. 38 (2011) 7334-7341.

[116] L. Wu, B. Yao, Z. Peng, Y. Guan, Fault diagnosis of roller bearings based on a wavelet neural network and manifold learning, Appl. Sci.-Basel 7 (2017) 158.

[117] I.A. Abu-Mahfouz, A comparative study of three artificial neural networks for the detection and classification of gear faults, Int. J. Gen. Syst. 34 (2005) 261-277.

[118] J. Rafiee, P.W. Tse, A. Harifi, M.H. Sadeghi, A novel technique for selecting mother wavelet function using an intelligent fault diagnosis system, Expert Syst. Appl. 36 (2009) 4862-4875.

[119] A. Hajnayeb, A. Ghasemloonia, S.E. Khadem, M.H. Moradi, Application and comparison of an ANN-based feature selection method and the genetic algorithm in gearbox fault diagnosis, Expert Syst. Appl. 38 (2011) 10205-10209.

[120] M. Cerrada, R.V. Sanchez, D. Cabrera, G. Zurita, C. Li, Multi-stage feature selection by using genetic algorithms for fault diagnosis in gearboxes based on vibration signal, Sensors 15 (2015) 23903-23926.

[121] P. Kane, A. Andhare, Application of psychoacoustics for gear fault diagnosis using artificial neural network, J. Low Freq. Noise Vib. Act. Control 35 (2016) 207-220.

[122] T. Waqar, M. Demetgul, Thermal analysis MLP neural network based fault diagnosis on worm gears,

Measurement 86 (2016) 56-66.
[123] S. Tyagi, S.K. Panigrahi, A hybrid genetic algorithm and back-propagation classifier for gearbox fault diagnosis, Appl. Artif. Intell. 31 (2017) 593-612.
[124] W. Lai, P.W. Tse, G. Zhang, T. Shi, Classification of gear faults using cumulants and the radial basis function network, Mech. Syst. Signal Process. 18 (2004) 381-389.
[125] H. Li, Y. Zhang, H. Zheng, Gear fault detection and diagnosis under speed-up condition based on order cepstrum and radial basis function neural network, J. Mech. Sci. Technol. 23 (2009) 2780-2789.
[126] H. Liu, J. Zhang, Y. Cheng, C. Lu, Fault diagnosis of gearbox using empirical mode decomposition and multi-fractal detrended cross-correlation analysis, J. Sound Vibrat. 385 (2016) 350-371.
[127] H. Chen, Y. Lu, L. Tu, Fault identification of gearbox degradation with optimized wavelet neural network, Shock Vibrat. 20 (2013) 247-262.
[128] B. Ayhan, M.Y. Chow, M.H. Song, Multiple discriminant analysis and neural-network-based monolith and partition fault-detection schemes for broken rotor bar in induction motors, IEEE Trans. Ind. Electron. 53 (2006) 1298-1308.
[129] A. Sadeghian, Z. Ye, B. Wu, Online detection of broken rotor bars in induction motors by wavelet packet decomposition and artificial neural networks, IEEE Trans. Instrum. Meas. 58 (2009) 2253-2263.
[130] H. Arabaci, O. Bilgin, Automatic detection and classification of rotor cage faults in squirrel cage induction motor, Neural Comput. Appl. 19 (2010) 713-723.
[131] E. Cabal-Yepez, M. Valtierra-Rodriguez, R.J. Romero-Troncoso, A. Garcia-Perez, R.A. Osornio-Rios, H. Miranda-Vidales, R. Alvarez-Salas, FPGA-based entropy neural processor for online detection of multiple combined faults on induction motors, Mech. Syst. Signal Process. 30 (2012) 123-130.
[132] M. Hernandez-Vargas, E. Cabal-Yepez, A. Garcia-Perez, Real-time SVD-based detection of multiple combined faults in induction motors, Comput. Electr. Eng. 40 (2014) 2193-2203.

[133] S.S. Moosavi, A. Djerdir, Y. Ait-Amirat, D.A. Khaburi, ANN based fault diagnosis of permanent magnet synchronous motor under stator winding shorted turn, Electr. Power Syst. Res. 125 (2015) 67-82.

[134] V.N. Ghate, S.V. Dudul, Cascade neural-network-based fault classifier for three-phase induction motor, IEEE Trans. Ind. Electron. 58 (2011) 1555-1563.

[135] R.H.C. Palacios, A. Goedtel, W.F. Godoy, J.A. Fabri, Fault identification in the stator winding of induction motors using PCA with artificial neural networks, J. Control Autom. Electr. Syst. 27 (2016) 406-418.

[136] T. Boukra, A. Lebaroud, G. Clerc, Statistical and neural-network approaches for the classification of induction machine faults using the ambiguity plane representation, IEEE Trans. Ind. Electron. 60 (2013) 4034-4042.

[137] A.J.C. Sharkey, G.O. Chandroth, N.E. Sharkey, A multi-net system for the fault diagnosis of a diesel engine, Neural Comput. Appl. 9 (2000) 152-160.

[138] P. Lu, M. Zhang, T. Hsu, J. Zhang, An evaluation of engine faults diagnostics using artificial neural networks, J. Eng. Gas Turbines Power-Trans. ASME 123 (2001) 340-346.

[139] J. Chen, R.B. Randall, Improved automated diagnosis of misfire in internal combustion engines based on simulation models, Mech. Syst. Signal Process. 64-65 (2015) 58-83.

[140] J. Chen, R.B. Randall, B. Peeters, Advanced diagnostic system for piston slap faults in ic engines, based on the non-stationary characteristics of the vibration signals, Mech. Syst. Signal Process. 75 (2016) 434-454.

[141] M. Khazaee, A. Banakar, B. Ghobadian, M. Mirsalim, S. Minaei, M. Jafari, P. Sharghi, Fault detection of engine timing belt based on vibration signals using data-mining techniques and a novel data fusion procedure, Struct. Health Monit. 15 (2016) 583-598.

[142] M. Khazaee, A. Banakar, B. Ghobadian, M.A. Mirsalim, S. Minaei, S.M. Jafari, Detection of inappropriate working conditions for the timing belt in internal-combustion engines using vibration signals and data mining, Proc. Inst. Mech. Eng. Part D-J. Automob. Eng. 231 (2017) 418-432.

[143] A. Zabihi-Hesari, S. Ansari-Rad, F.A. Shirazi, M. Ayati, Fault detection and diagnosis of a 12-cylinder trainset diesel engine based on vibration signature analysis and neural network, Proc. Inst. Mech. Eng. Part C-J. Mech. Eng. Sci. 233 (2019) 1910-1923.

[144] J. Wu, C. Huang, Y. Chang, Y. Shiao, Fault diagnosis for internal combustion engines using intake manifold pressure and artificial neural network, Expert Syst. Appl. 37 (2010) 949-958.

[145] J. Wu, C. Huang, An engine fault diagnosis system using intake manifold pressure signal and Wigner-Ville distribution technique, Expert Syst. Appl. 38 (2011) 536-544.

[146] Y. Shen, L. Cao, Z. Wang, S. Zhou, B. Gou, Fault diagnosis of diesel fuel ejection system based on improved wnn, in: *World Congress on Intelligent Control and Automation*, Dalian, China, 2006, pp. 5752-5755.

[147] C. Zhang, C. Guo, Y. Fan, Fault diagnosis for diesel engine based on immune wavelet neural network, in: *International Conference on Advanced Computer Control*, Shenyang, China, 2010, pp. 522-526.

[148] R.J. Kuo, Intelligent diagnosis for turbine blade faults using artificial neural networks and fuzzy-logic, Eng. Appl. Artif. Intell. 8 (1995) 25-34.

[149] P.W. Ilott, A.J. Griffiths, Fault diagnosis of pumping machinery using artificial neural networks, Proc. Inst. Mech. Eng. Part E-J. Process Mech. Eng. 211 (1997) 185-194.

[150] J. Wu, J. Kuo, An automotive generator fault diagnosis system using discrete wavelet transform and artificial neural network, Expert Syst. Appl. 36 (2009) 9776-9783.

[151] A.A. Mohammed, R.D. Neilson, W.F. Deans, P. MacConnell, Crack detection in a rotating shaft using artificial neural networks and PSD characterisation, Meccanica 49 (2014) 255-266.

[152] R.B. Walker, R. Vayanat, S. Perinpanayagam, I.K. Jennions, Unbalance localization through machine nonlinearities using an artificial neural network approach, Mech. Mach. Theory 75 (2014) 54-66.

[153] H. Malik, S. Mishra, Artificial neural network and empirical mode decomposition based imbalance fault diagnosis of wind turbine using turbsim, fast and simulink, IET Renewable Power Gener. 11 (2017) 889-902.

[154] A.C. McCormick, A.K. Nandi, Classification of the rotating machine condition using artificial neural networks, Proc. Inst. Mech. Eng. Part C-J. Mech. Eng. Sci. 211 (1997) 439-450.

[155] A.C. McCormick, A.K. Nandi, Real-time classification of rotating shaft loading conditions using artificial neural networks, IEEE Trans. Neural Netw. 8 (1997) 748-757.

[156] J. Wu, Y. Wang, P. Chiang, M.R. Bai, A study of fault diagnosis in a scooter using adaptive order tracking technique and neural network, Expert Syst. Appl. 36 (2009) 49-56.

[157] J.A.B. Villanueva, F.J. Espadafor, F. Cruz-Peragon, M.T. Garcia, A methodology for cracks identification in large crankshafts, Mech. Syst. Signal Process. 25 (2011) 3168-3185.

[158] Q. Liu, X. Yu, Q. Feng, Fault diagnosis using wavelet neural networks, Neural Process. Lett. 18 (2003) 115-123.

[159] C. Chen, C. Mo, A method for intelligent fault diagnosis of rotating machinery, Digital Signal Process. 14 (2004) 203-217.

[160] Q. Guo, H. Yu, A. Xu, A hybrid PSO-GD based intelligent method for machine diagnosis, Digital Signal Process. 16 (2006) 402-418.

[161] Z. Xiao, X. He, X. Fu, O.P. Malik, ACO-initialized wavelet neural network for vibration fault diagnosis of hydroturbine generating unit, Math. Prob. Eng. (2015) Doi: 10.1155/2015/354658.

[162] Y. Jin, C. Shan, Y. Wu, Y. Xia, Y. Zhang, L. Zeng, Fault diagnosis of hydraulic seal wear and internal leakage using wavelets and wavelet neural network, IEEE Trans. Instrum. Meas. 68 (2019) 1026-1034.

[163] B. Pang, G. Tang, C. Zhou, T. Tian, Rotor fault diagnosis based on characteristic frequency band energy entropy and support vector machine, Entropy 20 (2018) 932.

[164] Z. Wang, J. Fan, Fault early recognition and health monitoring on aeroengine rotor system, J. Aerosp. Eng. 28 (2015) 04014065.

[165] X. Jin, J. Feng, S. Du, G. Li, Y. Zhao, Rotor fault classification technique and precision analysis with kernel

principal component analysis and multi-support vector machines, J. Vibroeng. 16 (2014) 2582-2592.
[166] S. Yuan, F. Chu, Support vector machines-based fault diagnosis for turbo-pump rotor, Mech. Syst. Signal Process. 20 (2006) 939-952.
[167] X. Tang, L. Zhuang, J. Cai, C. Li, Multi-fault classification based on support vector machine trained by chaos particle swarm optimization, Knowledge-Based Syst. 23 (2010) 486-490.
[168] X. Zhang, W. Chen, B. Wang, X. Chen, Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization, Neurocomputing 167 (2015) 260-279.
[169] R. Jegadeeshwaran, V. Sugumaran, Fault diagnosis of automobile hydraulic brake system using statistical features and support vector machines, Mech. Syst. Signal Process. 52-53 (2015) 436-446.
[170] J.S. Rapur, R. Tiwari, Automation of multi-fault diagnosing of centrifugal pumps using multi-class support vector machine with vibration and motor current signals in frequency domain, J. Braz. Soc. Mech. Sci. Eng. 40 (2018) 278.
[171] J.S. Rapur, R. Tiwari, On-line time domain vibration and current signals based multi-fault diagnosis of centrifugal pumps using support vector machines, J. Nondestr. Eval. 38 (2019) 6.
[172] J.C. Platt, Fast training of support vector machines using sequential minimal optimization, MIT Press, Cambridge, Massachusetts, 1999.
[173] C.W. Hsu, C.J. Lin, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Netw. 13 (2002) 415-425.
[174] K. Bacha, S. Ben Salem, A. Chaari, An improved combination of Hilbert and park transforms for fault detection and identification in three-phase induction motors, Int. J. Electr. Power Energy Syst. 43 (2012) 1006-1016.
[175] S. Zgarni, H. Keskes, A. Braham, Nested SVDD in DAG SVM for induction motor condition monitoring,

Eng. Appl. Artif. Intell. 71 (2018) 210-215.

[176] H. Keskes, A. Braham, Recursive undecimated wavelet packet transform and DAG SVM for induction motor diagnosis, IEEE Trans. Ind. Inform. 11 (2015) 1059-1066.

[177] Y. Li, M. Xu, Y. Wei, W. Huang, A new rolling bearing fault diagnosis method based on multiscale permutation entropy and improved support vector machine based binary tree, Measurement 77 (2016) 80-94.

[178] Y. Li, M. Xu, H. Zhao, W. Huang, Hierarchical fuzzy entropy and improved support vector machine based binary tree approach for rolling bearing fault diagnosis, Mech. Mach. Theory 98 (2016) 114-132.

[179] Y. Li, Y. Yang, X. Wang, B. Liu, X. Liang, Early fault diagnosis of rolling bearings based on hierarchical symbol dynamic entropy and binary tree support vector machine, J. Sound Vibrat. 428 (2018) 72-86.

[180] D. Yang, Y. Liu, S. Li, X. Li, L. Ma, Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm, Mech. Mach. Theory 90 (2015) 219-229.

[181] Y. Li, G. Li, Y. Yang, X. Liang, M. Xu, A fault diagnosis scheme for planetary gearboxes using adaptive multi-scale morphology filter and modified hierarchical permutation entropy, Mech. Syst. Signal Process. 105 (2018) 319-337.

[182] B. Tang, T. Song, F. Li, L. Deng, Fault diagnosis for a wind turbine transmission system based on manifold learning and shannon wavelet support vector machine, Renewable Energy 62 (2014) 1-9.

[183] M. Kang, J.M. Kim, Singular value decomposition based feature extraction approaches for classifying faults of induction motors, Mech. Syst. Signal Process. 41 (2013) 348-356.

[184] S. Ben Salem, K. Bacha, A. Chaari, Support vector machine based decision for mechanical fault condition monitoring in induction motor using an advanced Hilbert-park transform, ISA Trans. 51 (2012) 566-572.

[185] C. Cai, X. Weng, C. Zhang, A novel approach for marine diesel engine fault diagnosis, Cluster Comput. 20 (2017) 1691-1702.

[186] T. Xu, Z. Yin, D. Cai, D. Zheng, Fault diagnosis for rotating machinery based on local mean decomposition

morphology filtering and least square support vector machine, J. Intell. Fuzzy Syst. 32 (2017) 2061-2070.

[187] Y. Li, K. Feng, X. Liang, M.J. Zuo, A fault diagnosis method for planetary gearboxes under non-stationary working conditions using improved Vold-Kalman filter and multi-scale sample entropy, J. Sound Vibrat. 439 (2019) 271-286.

[188] X. Jiang, S. Li, Y. Wang, A novel method for self-adaptive feature extraction using scaling crossover characteristics of signals and combining with LS-SVM for multi-fault diagnosis of gearbox, J. Vibroeng. 17 (2015) 1861-1878.

[189] M. Heidari, H. Homaei, H. Golestanian, A. Heidari, Fault diagnosis of gearboxes using wavelet support vector machine, least square support vector machine and wavelet packet transform, J. Vibroeng. 18 (2016) 860-875.

[190] Q. Jiang, Q. Zhu, B. Wang, L. Guo, Nonlinear machine fault detection by semi-supervised laplacian eigenmaps, J. Mech. Sci. Technol. 31 (2017) 3697-3703.

[191] C.M. Vong, P.K. Wong, Engine ignition signal diagnosis with wavelet packet transform and multi-class least squares support vector machines, Expert Syst. Appl. 38 (2011) 8563-8570.

[192] V. Sugumaran, V. Muralidharan, K.I. Ramachandran, Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing, Mech. Syst. Signal Process. 21 (2007) 930-942.

[193] N. Saravanan, V.N.S.K. Siddabattuni, K.I. Ramachandran, A comparative study on classification of features by svm and psvm extracted using morlet wavelet for fault diagnosis of spur bevel gear box, Expert Syst. Appl. 35 (2008) 1351-1366.

[194] L.H. Chiang, M.E. Kotanchek, A.K. Kordon, Fault diagnosis based on fisher discriminant analysis and support vector machines, Comput. Chem. Eng. 28 (2004) 1389-1401.

[195] V. Sugumaran, G.R. Sabareesh, K.I. Ramachandran, Fault diagnostics of roller bearing using kernel based

neighborhood score multi-class support vector machine, Expert Syst. Appl. 34 (2008) 3090-3098.
[196] Y. Wang, S. Kang, Y. Jiang, G. Yang, L. Song, V.I. Mikulovich, Classification of fault location and the degree of performance degradation of a rolling bearing based on an improved hyper-sphere-structured multi-class support vector machine, Mech. Syst. Signal Process. 29 (2012) 404-414.
[197] S. Dong, B. Tang, R. Chen, Bearing running state recognition based on non-extensive wavelet feature scale entropy and support vector machine, Measurement 46 (2013) 4189-4199.
[198] M. Heidari, S. Shateyi, Wavelet support vector machine and multi-layer perceptron neural network with continues wavelet transform for fault diagnosis of gearboxes, J. Vibroeng. 19 (2017) 125-137.
[199] H. Keskes, A. Braham, Z. Lachiri, Broken rotor bar diagnosis in induction machines through stationary wavelet packet transform and multiclass wavelet SVM, Electr. Power Syst. Res. 97 (2013) 151-157.
[200] F. Chen, B. Tang, R. Chen, A novel fault diagnosis model for gearbox based on wavelet support vector machine with immune genetic algorithm, Measurement 46 (2013) 220-232.
[201] X. Zhang, B. Wang, X. Chen, Intelligent fault diagnosis of roller bearings with multivariable ensemble-based incremental support vector machine, Knowledge-Based Syst. 89 (2015) 56-85.
[202] J. Zheng, H. Pan, J. Cheng, Rolling bearing fault detection and diagnosis based on composite multiscale fuzzy entropy and ensemble support vector machines, Mech. Syst. Signal Process. 85 (2017) 746-759.
[203] B.M. Ebrahimi, M.J. Roshtkhari, J. Faiz, S.V. Khatami, Advanced eccentricity fault recognition in permanent magnet synchronous motors using stator current signature analysis, IEEE Trans. Ind. Electron. 61 (2014) 2041-2052.
[204] J. Hang, J. Zhang, M. Cheng, Application of multi-class fuzzy support vector machine classifier for fault diagnosis of wind turbine, Fuzzy Sets Syst. 297 (2016) 128-140.
[205] Z. Li, S. Zhong, L. Lin, Novel gas turbine fault diagnosis method based on performance deviation model, J. Propul. Power 33 (2017) 730-739.

[206] F. Chen, B. Tang, T. Song, L. Li, Multi-fault diagnosis study on roller bearing based on multi-kernel support vector machine with chaotic particle swarm optimization, Measurement 47 (2014) 576-590.

[207] Y. Liu, J. Zhang, L. Ma, A fault diagnosis approach for diesel engines based on self-adaptive WVD, improved FCBF and PECOC-RVM, Neurocomputing 177 (2016) 600-611.

[208] L.B. Jack, A.K. Nandi, Support vector machines for detection and characterization of rolling element bearing faults, Proc. Inst. Mech. Eng. Part C-J. Mech. Eng. Sci. 215 (2001) 1065-1074.

[209] L.B. Jack, A.K. Nandi, Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms, Mech. Syst. Signal Process. 16 (2002) 373-390.

[210] A. Rojas, A.K. Nandi, Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines, Mech. Syst. Signal Process. 20 (2006) 1523-1536.

[211] A. Widodo, E.Y. Kim, J.D. Son, B.S. Yang, A.C.C. Tan, D.S. Gu, B.K. Choi, J. Mathew, Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine, Expert Syst. Appl. 36 (2009) 7252-7261.

[212] N. Li, R. Zhou, Q. Hu, X. Liu, Mechanical fault diagnosis based on redundant second generation wavelet packet transform, neighborhood rough set and support vector machine, Mech. Syst. Signal Process. 28 (2012) 608-621.

[213] A. Widodo, B.S. Yang, Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors, Expert Syst. Appl. 33 (2007) 241-250.

[214] A. Widodo, B.S. Yang, T. Han, Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors, Expert Syst. Appl. 32 (2007) 299-312.

[215] B. Samanta, K.R. Al-Balushi, S.A. Al-Araimi, Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection, Eng. Appl. Artif. Intell. 16 (2003) 657-665.

[216] M. Kang, J. Kim, J.-M. Kim, A.C.C. Tan, E.Y. Kim, B.K. Choi, Reliable fault diagnosis for low-speed

bearings using individually trained support vector machines with kernel discriminative feature analysis, IEEE Trans. Power Electron. 30 (2015) 2786-2797.

[217] X. Zhu, J. Xiong, Q. Liang, Fault diagnosis of rotation machinery based on support vector machine optimized by quantum genetic algorithm, IEEE Access 6 (2018) 33583-33588.

[218] B. Samanta, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, Mech. Syst. Signal Process. 18 (2004) 625-644.

[219] K. Zhu, X. Song, D. Xue, A roller bearing fault diagnosis method based on hierarchical entropy and support vector machine with particle swarm optimization algorithm, Measurement 47 (2014) 669-675.

[220] Z. Su, B. Tang, Z. Liu, Y. Qin, Multi-fault diagnosis for rotating machinery based on orthogonal supervised linear local tangent space alignment and least square support vector machine, Neurocomputing 157 (2015) 208-222.

[221] S. Dong, X. Xu, J. Liu, Z. Gao, Rotating machine fault diagnosis based on locality preserving projection and back propagation neural network-support vector machine model, Meas. Control 48 (2015) 211-216.

[222] Y. Li, B. Miao, W. Zhang, P. Chen, J. Liu, X. Jiang, Refined composite multiscale fuzzy entropy: Localized defect detection of rolling element bearing, J. Mech. Sci. Technol. 33 (2019) 109-120.

[223] X. Li, A. Zheng, X. Zhang, C. Li, L. Zhang, Rolling element bearing fault detection using support vector machine with improved ant colony optimization, Measurement 46 (2013) 2726-2734.

[224] S. Abbasion, A. Rafsanjani, A. Farshidianfar, N. Irani, Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine, Mech. Syst. Signal Process. 21 (2007) 2933-2945.

[225] Y. Yang, D. Yu, J. Cheng, A fault diagnosis approach for roller bearing based on imf envelope spectrum and svm, Measurement 40 (2007) 943-950.

[226] G. Xian, B. Zeng, An intelligent fault diagnosis method based on wavelet packer analysis and hybrid support

vector machines, Expert Syst. Appl. 36 (2009) 12131-12136.
[227] R. Hao, Z. Peng, Z. Feng, F. Chu, Application of support vector machine based on pattern spectrum entropy in fault diagnostics of rolling element bearings, Meas. Sci. Technol. 22 (2011) 045708.
[228] K.C. Gryllias, I.A. Antoniadis, A support vector machine approach based on physical model training for rolling element bearing fault detection in industrial environments, Eng. Appl. Artif. Intell. 25 (2012) 326-344.
[229] M.M.M. Islam, J. Kim, S.A. Khan, J.-M. Kim, Reliable bearing fault diagnosis using Bayesian inference-based multi-class support vector machines, J. Acoust. Soc. Am. 141 (2017) 89-95.
[230] A. HungLinh, J. Cheng, Y. Yang, T. Tung Khac, The support vector machine parameter optimization method based on artificial chemical reaction optimization algorithm and its application to roller bearing fault diagnosis, J. Vibrat. Control 21 (2015) 2434-2445.
[231] B.S. Yang, T. Han, W.W. Hwang, Fault diagnosis of rotating machinery based on multi-class support vector machines, J. Mech. Sci. Technol. 19 (2005) 846-859.
[232] J. Yang, Y. Zhang, Y. Zhu, Intelligent fault diagnosis of rolling element bearing based on SVMs and fractal dimension, Mech. Syst. Signal Process. 21 (2007) 2012-2024.
[233] S. Wu, P. Wu, C. Wu, J. Ding, C. Wang, Bearing fault diagnosis based on multiscale permutation entropy and support vector machine, Entropy 14 (2012) 1343-1356.
[234] S. Wu, C. Wu, T. Wu, C. Wang, Multi-scale analysis based ball bearing defect diagnostics using Mahalanobis distance and support vector machine, Entropy 15 (2013) 416-433.
[235] L. Saidi, J. Ben Ali, F. Friaiech, Application of higher order spectral features and support vector machines for bearing faults classification, ISA Trans. 54 (2015) 193-206.
[236] K. Zhu, H. Li, A rolling element bearing fault diagnosis approach based on hierarchical fuzzy entropy and support vector machine, Proc. Inst. Mech. Eng. Part C-J. Mech. Eng. Sci. 230 (2016) 2314-2322.

[237] R. Ziani, A. Felkaoui, R. Zegadi, Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized Fishers criterion, J. Intell. Manuf. 28 (2017) 405-417.
[238] M.M.M. Islam, J.M. Kim, Reliable multiple combined fault diagnosis of bearings using heterogeneous feature models and multiclass support vector machines, Reliab. Eng. Syst. Saf. 184 (2019) 55-66.
[239] X. Zhang, J. Zhou, Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines, Mech. Syst. Signal Process. 41 (2013) 127-140.
[240] Z. Liu, M.J. Zuo, H. Xu, Feature ranking for support vector machine classification and its application to machinery fault diagnosis, Proc. Inst. Mech. Eng. Part C-J. Mech. Eng. Sci. 227 (2013) 2077-2089.
[241] C. Li, R.V. Sanchez, G. Zurita, V. Cerrada, D. Cabrera, R.E. Vasquez, Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis, Neurocomputing 168 (2015) 119-127.
[242] W. Lu, W. Jiang, G. Yuan, L. Yan, A gearbox fault diagnosis scheme based on near-field acoustic holography and spatial distribution features of sound field, J. Sound Vibrat. 332 (2013) 2593-2610.
[243] F. Cheng, Y. Peng, L. Qu, W. Qiao, Current-based fault detection and identification for wind turbine drivetrain gearboxes, IEEE Trans. Ind. Appl. 53 (2017) 878-887.
[244] Z. Xing, J. Qu, Y. Chai, Q. Tang, Y. Zhou, Gear fault diagnosis under variable conditions with intrinsic time-scale decomposition-singular value decomposition and support vector machine, J. Mech. Sci. Technol. 31 (2017) 545-553.
[245] L. Liu, X. Liang, M.J. Zuo, A dependence-based feature vector and its application on planetary gearbox fault classification, J. Sound Vibrat. 431 (2018) 192-211.
[246] C. Shen, D. Wang, F. Kong, P.W. Tse, Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier, Measurement 46 (2013) 1551-1564.

[247] D.J. Bordoloi, R. Tiwari, Support vector machine based optimization of multi-fault classification of gears with evolutionary algorithms from time-frequency vibration data, Measurement 55 (2014) 1-14.

[248] X. Zhang, J. Zhao, X. Zhang, X. Ni, H. Li, F. Sun, A novel hybrid compound fault pattern identification method for gearbox based on NIC, MFDFA and WOASVM, J. Mech. Sci. Technol. 33 (2019) 1097-1113.

[249] A. Widodo, B.S. Yang, D.S. Gu, B.K. Choi, Intelligent fault diagnosis system of induction motor based on transient current signal, Mechatronics 19 (2009) 680-689.

[250] B.M. Ebrahimi, J. Faiz, Feature extraction for short-circuit fault detection in permanent-magnet synchronous motors using stator-current monitoring, IEEE Trans. Power Electron. 25 (2010) 2673-2682.

[251] M.R. Shahriar, T. Ahsan, U. Chong, Fault diagnosis of induction motors utilizing local binary pattern-based texture analysis, Eurasip J. Image Vide. (2013) Doi: 10.1186/1687-5281-2013-29.

[252] M. Kang, J.M. Kim, Reliable fault diagnosis of multiple induction motor defects using a 2-D representation of shannon wavelets, IEEE Trans. Magn. 50 (2014) 8201913.

[253] J. Kurek, S. Osowski, Support vector machine for fault diagnosis of the broken rotor bars of squirrel-cage induction motor, Neural Comput. Appl. 19 (2010) 557-564.

[254] P. Gangsar, R. Tiwari, Comparative investigation of vibration and current monitoring for prediction of mechanical and electrical faults in induction motor based on multiclass-support vector machine algorithms, Mech. Syst. Signal Process. 94 (2017) 464-481.

[255] P. Gangsar, R. Tiwari, Multifault diagnosis of induction motor at intermediate operating conditions using wavelet packet transform and support vector machine, J. Dyn. Syst. Meas. Control-Trans. ASME (2018) Doi: 10.1115/1.4039204.

[256] P. Gangsar, R. Tiwari, Diagnostics of mechanical and electrical faults in induction motors using wavelet-based features of vibration and current through support vector machine algorithms for various operating conditions, J. Braz. Soc. Mech. Sci. Eng. 41 (2019) 71.

[257] W. Sun, R. Zhao, R. Yan, S. Shao, X. Chen, Convolutional discriminative feature learning for induction motor fault diagnosis, IEEE Trans. Ind. Inform. 13 (2017) 1350-1359.

[258] J.D. Martinez-Morales, E.R. Palacios-Hernandez, D.U. Campos-Delgado, Multiple-fault diagnosis in induction motors through support vector machine classification at variable operating conditions, Electr. Eng. 100 (2018) 59-73.

[259] I.P. Tsoumas, G. Georgoulas, E.D. Mitronikas, A.N. Safacas, Asynchronous machine rotor fault diagnosis technique using complex wavelets, IEEE Trans. Energy Convers. 23 (2008) 444-459.

[260] Z. Li, X. Yan, C. Yuan, Z. Peng, Intelligent fault diagnosis method for marine diesel engines using instantaneous angular speed, J. Mech. Sci. Technol. 26 (2012) 2413-2423.

[261] S.M. Lee, T.S. Roh, D.W. Choi, Defect diagnostics of SUAV gas turbine engine using hybrid SVM-artificial neural network method, J. Mech. Sci. Technol. 23 (2009) 559-568.

[262] Y. Wang, Q. Ma, Q. Zhu, X. Liu, L. Zhao, An intelligent approach for engine fault diagnosis based on hilbert-huang transform and support vector machine, Appl. Acoust. 75 (2013) 1-9.

[263] K. Jafarian, M. Mobin, R. Jafari-Marandi, E. Rabiei, Misfire and valve clearance faults detection in the combustion engines based on a multi-sensor vibration signal monitoring, Measurement 128 (2018) 527-536.

[264] D.P. Jena, S.N. Panigrahi, Motor bike piston-bore fault identification from engine noise signature analysis, Appl. Acoust. 76 (2014) 35-47.

[265] M. Namdari, H. Jazayeri-Rad, Incipient fault diagnosis using support vector machines based on monitoring continuous decision functions, Eng. Appl. Artif. Intell. 28 (2014) 22-35.

[266] D. He, R. Li, J. Zhu, Plastic bearing fault diagnosis based on a two-step data mining approach, IEEE Trans. Ind. Electron. 60 (2013) 3429-3440.

[267] L. Jiang, J. Xuan, T. Shi, Feature extraction based on semi-supervised kernel marginal fisher analysis and its application in bearing fault diagnosis, Mech. Syst. Signal Process. 41 (2013) 113-126.

[268] L. Jiang, T. Shi, J. Xuan, Fault diagnosis of rolling bearings based on marginal Fisher analysis, J. Vibrat. Control 20 (2014) 470-480.

[269] M.S. Safizadeh, S.K. Latifi, Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell, Inf. Fusion 18 (2014) 1-8.

[270] M. Van, H.J. Kang, Two-stage feature selection for bearing fault diagnosis based on dual-tree complex wavelet transform and empirical mode decomposition, Proc. Inst. Mech. Eng. Part C-J. Mech. Eng. Sci. 230 (2016) 291-302.

[271] X. An, Y. Tang, Application of variational mode decomposition energy distribution to bearing fault diagnosis in a wind turbine, Trans. Inst. Meas. Control 39 (2017) 1000-1006.

[272] J. Ma, F. Xu, K. Huang, R. Huang, GNAR-GARCH model and its application in feature extraction for rolling bearing fault diagnosis, Mech. Syst. Signal Process. 93 (2017) 175-203.

[273] B. Yao, P. Zhen, L. Wu, Y. Guan, Rolling element bearing fault diagnosis using improved manifold learning, IEEE Access 5 (2017) 6027-6035.

[274] M.H. Gharavian, F.A. Ganj, A.R. Ohadi, H.H. Bafroui, Comparison of FDA-based and PCA-based features in fault diagnosis of automobile gearboxes, Neurocomputing 121 (2013) 150-159.

[275] Z. Li, X. Yan, Z. Tian, C. Yuan, Z. Peng, L. Li, Blind vibration component separation and nonlinear feature extraction applied to the nonstationary vibration signals for the gearbox multi-fault diagnosis, Measurement 46 (2013) 259-271.

[276] S. Park, S. Kim, J.H. Choi, Gear fault diagnosis using transmission error and ensemble empirical mode decomposition, Mech. Syst. Signal Process. 108 (2018) 262-275.

[277] Vanraj, S.S. Dhami, B.S. Pabla, Hybrid data fusion approach for fault diagnosis of fixed-axis gearbox, Struct. Health Monit. 17 (2018) 936-945.

[278] A. Glowacz, Z. Glowacz, Diagnosis of stator faults of the single-phase induction motor using acoustic

signals, Appl. Acoust. 117 (2017) 20-27.
[279] Y. Lei, Z. He, Y. Zi, A combination of WKNN to fault diagnosis of rolling element bearings, J. Vibrat. Acoust.-Trans. ASME 131 (2009) 064502.
[280] X. Zhao, M. Jia, Fault diagnosis of rolling bearing based on feature reduction with global-local margin fisher analysis, Neurocomputing 315 (2018) 447-464.
[281] F. Li, J. Wang, M.K.K. Chyu, B. Tang, Weak fault diagnosis of rotating machinery based on feature reduction with supervised orthogonal local Fisher discriminant analysis, Neurocomputing 168 (2015) 505-519.
[282] S. Dong, X. Xu, R. Chen, Application of fuzzy C-means method and classification model of optimized k-nearest neighbor for fault diagnosis of bearing, J. Braz. Soc. Mech. Sci. Eng. 38 (2016) 2255-2263.
[283] S. Dong, T. Luo, L. Zhong, L. Chen, X. Xu, Fault diagnosis of bearing based on the kernel principal component analysis and optimized k-nearest neighbour model, J. Low Freq. Noise Vib. Act. Control 36 (2017) 354-365.
[284] H. Yuan, J. Chen, G. Dong, An improved initialization method of D-KSVD algorithm for bearing fault diagnosis, J. Mech. Sci. Technol. 31 (2017) 5161-5172.
[285] J. Yu, Y. He, Planetary gearbox fault diagnosis based on data-driven valued characteristic multigranulation model with incomplete diagnostic information, J. Sound Vibrat. 429 (2018) 63-77.
[286] J. Yu, W. Huang, X. Zhao, Combined flow graphs and normal naive Bayesian classifier for fault diagnosis of gear box, Proc. Inst. Mech. Eng. Part C-J. Mech. Eng. Sci. 230 (2016) 303-313.
[287] J. Yu, M. Bai, G. Wang, X. Shi, Fault diagnosis of planetary gearbox with incomplete information using assignment reduction and flexible naive Bayesian classifier, J. Mech. Sci. Technol. 32 (2018) 37-47.
[288] Y. He, R. Wang, S. Kwong, X. Wang, Bayesian classifiers based on probability density estimation and their applications to simultaneous fault diagnosis, Inf. Sci. 259 (2014) 252-268.

[289] J. Yu, B. Ding, Y. He, Rolling bearing fault diagnosis based on mean multigranulation decision-theoretic rough set and non-naive Bayesian classifier, J. Mech. Sci. Technol. 32 (2018) 5201-5211.

[290] M.Y. Asr, M.M. Ettefagh, R. Hassannejad, S.N. Razavi, Diagnosis of combined faults in rotary machinery by non-naive Bayesian approach, Mech. Syst. Signal Process. 85 (2017) 56-70.

[291] M. Yuwono, Y. Qin, J. Zhou, Y. Guo, B.G. Celler, S.W. Su, Automatic bearing fault diagnosis using particle swarm clustering and hidden Markov model, Eng. Appl. Artif. Intell. 47 (2016) 88-100.

[292] H. Zhou, J. Chen, G. Dong, R. Wang, Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden Markov model, Mech. Syst. Signal Process. 72-73 (2016) 65-79.

[293] T. Boutros, M. Liang, Detection and diagnosis of bearing and cutting tool faults using hidden Markov models, Mech. Syst. Signal Process. 25 (2011) 2102-2124.

[294] T. Liu, J. Chen, G. Dong, Singular spectrum analysis and continuous hidden Markov model for rolling element bearing fault diagnosis, J. Vibrat. Control 21 (2015) 1506-1521.

[295] H. Jiang, J. Chen, G. Dong, T. Liu, G. Chen, Study on hankel matrix-based svd and its application in rolling element bearing fault diagnosis, Mech. Syst. Signal Process. 52-53 (2015) 338-359.

[296] O. Geramifard, J.X. Xu, S.K. Panda, Fault detection and diagnosis in synchronous motors using hidden markov model-based semi-nonparametric approach, Eng. Appl. Artif. Intell. 26 (2013) 1919-1929.

[297] Y. Jia, M. Xu, R. Wang, Symbolic important point perceptually and hidden Markov model based hydraulic pump fault diagnosis method, Sensors 18 (2018) 4460.

[298] W. Xiao, J. Chen, G. Dong, Y. Zhou, Z. Wang, A multichannel fusion approach based on coupled hidden markov models for rolling element bearing fault diagnosis, Proc. Inst. Mech. Eng. Part C-J. Mech. Eng. Sci. 226 (2012) 202-216.

[299] D. Huang, L. Ke, X. Chu, L. Zhao, B. Mi, Fault diagnosis for the motor drive system of urban transit based on improved hidden Markov model, Microelectron. Reliab. 82 (2018) 179-189.

[300] J.R. Quinlan, C4.5: Programs for machine learning, Morgan Kaufmann Publishers Inc., San Francisco, 1993.

[301] M. Amarnath, V. Sugumaran, H. Kumar, Exploiting sound signals for fault diagnosis of bearings using decision tree, Measurement 46 (2013) 1250-1256.

[302] Y. Zhang, X. Li, L. Gao, L. Wang, L. Wen, Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning, J. Manuf. Syst. 48 (2018) 34-50.

[303] T. Praveenkumar, B. Sabhrish, M. Saimurugan, K.I. Ramachandran, Pattern recognition based on-line vibration monitoring system for fault diagnosis of automobile gearbox, Measurement 114 (2018) 233-242.

[304] W. Sun, J. Chen, J. Li, Decision tree and PCA-based fault diagnosis of rotating machinery, Mech. Syst. Signal Process. 21 (2007) 1300-1317.

[305] N.R. Sakthivel, V. Sugumaran, S. Babudevasenapati, Vibration based fault diagnosis of monoblock centrifugal pump using decision tree, Expert Syst. Appl. 37 (2010) 4040-4049.

[306] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5-32.

[307] B.S. Yang, X. Di, T. Han, Random forests classifier for machine fault diagnosis, J. Mech. Sci. Technol. 22 (2008) 1716-1725.

[308] Z. Wang, Q. Zhang, J. Xiong, M. Xiao, G. Sun, J. He, Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests, IEEE Sens. J. 17 (2017) 5581-5588.

[309] G. Tang, B. Pang, T. Tian, C. Zhou, Fault diagnosis of rolling bearings based on improved fast spectral correlation and optimized random forest, Appl. Sci.-Basel 8 (2018) 1859.

[310] F. Jia, Y. Lei, J. Lin, X. Zhou, N. Lu, Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data, Mech. Syst. Signal Process. 72-73 (2016) 303-315.

[311] Y. Lei, F. Jia, J. Lin, S. Xing, S.X. Ding, An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data, IEEE Trans. Ind. Electron. 63 (2016) 3137-3147.

[312] V. Mayer-Schnberger, K. Cukier, Big data: A revolution that will transform how we live, work, and think, Hodder & Stoughton Press, Carmelite House, London, 2013.

[313] X. Xu, Y. Lei, Z. Li, An incorrect data detection method for big data cleaning of machinery condition monitoring, IEEE Trans. Ind. Electron. (2019) Doi: 10.1109/TIE.2019.2903774.

[314] G.I. Platforms, The rise of industrial big data, 2012, Available: http://www.geautomation.com/download/rise-industrial-big-data.

[315] S. Yin, X. Li, H. Gao, O. Kaynak, Data-based techniques focused on modern industry: An overview, IEEE Trans. Ind. Electron. 62 (2015) 657-667.

[316] J. Lee, Industrial big data: The revolutionary transformation and value creation in industry 4.0 era, Mechanical Industry Press, Beijing, China, 2015.

[317] H. Liu, L. Li, J. Ma, Rolling bearing fault diagnosis based on stft-deep learning and sound signals, Shock Vibrat. (2016) Doi: 10.1155/2016/6127479.

[318] X. Guo, C. Shen, L. Chen, Deep fault recognizer: An integrated model to denoise and extract features for fault diagnosis in rotating machinery, Appl. Sci.-Basel 7 (2017) 41.

[319] C. Lu, Z. Wang, W. Qin, J. Ma, Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification, Signal Process. 130 (2017) 377-388.

[320] Y. Qu, M. He, J. Deutsch, D. He, Detection of pitting in gears using a deep sparse autoencoder, Appl. Sci.-Basel 7 (2017) 515.

[321] M. Xia, T. Li, L. Liu, L. Xu, C.W. de Silva, Intelligent fault diagnosis approach with unsupervised feature learning by stacked denoising autoencoder, IET Sci. Meas. Technol. 11 (2017) 687-695.

[322] H.O.A. Ahmed, M.L.D. Wong, A.K. Nandi, Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features, Mech. Syst. Signal Process. 99 (2018) 459-477.

[323] Z. Chen, Z. Li, Fault diagnosis method of rotating machinery based on stacked denoising autoencoder, J. Intell. Fuzzy Syst. 34 (2018) 3443-3449.

[324] B.P. Duong, J.M. Kim, Non-mutually exclusive deep neural network classifier for combined modes of bearing fault diagnosis, Sensors 18 (2018) 1129.

[325] C. Li, W. Zhang, G. Peng, S. Liu, Bearing fault diagnosis using fully-connected winner-take-all autoencoder, IEEE Access 6 (2018) 6103-6115.

[326] G. Liu, H. Bao, B. Han, A stacked autoencoder-based deep neural network for achieving gearbox fault diagnosis, Math. Prob. Eng. (2018) Doi: 10.1155/2018/5105709.

[327] Z. Meng, X. Zhan, J. Li, Z. Pan, An enhancement denoising autoencoder for rolling bearing fault diagnosis, Measurement 130 (2018) 448-454.

[328] M. Sohaib, J.M. Kim, Reliable fault diagnosis of rotary machine bearings using a stacked sparse autoencoder-based deep neural network, Shock Vibrat. (2018) Doi: 10.1155/2018/2919637.

[329] J. Sun, C. Yan, J. Wen, Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning, IEEE Trans. Instrum. Meas. 67 (2018) 185-195.

[330] Y. Zheng, T. Wang, B. Xin, T. Xie, Y. Wang, A sparse autoencoder and softmax regression based diagnosis method for the attachment on the blades of marine current turbine, Sensors 19 (2019) 826.

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

[332] H. Shao, H. Jiang, H. Zhao, F. Wang, A novel deep autoencoder feature learning method for rotating machinery fault diagnosis, Mech. Syst. Signal Process. 95 (2017) 187-204.

[333] H. Liu, J. Zhou, Y. Zheng, W. Jiang, Y. Zhang, Fault diagnosis of rolling bearings with recurrent neural network based autoencoders, ISA Trans. 77 (2018) 167-178.

[334] M. Ma, C. Sun, X. Chen, Deep coupling autoencoder for fault diagnosis with multimodal sensory data, IEEE

Trans. Ind. Inform. 14 (2018) 1137-1145.
[335] H. Shao, H. Jiang, K. Zhao, D. Wei, X. Li, A novel tracking deep wavelet auto-encoder method for intelligent fault diagnosis of electric locomotive bearings, Mech. Syst. Signal Process. 110 (2018) 193-209.
[336] H. Shao, H. Jiang, Y. Lin, X. Li, A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders, Mech. Syst. Signal Process. 102 (2018) 278-297.
[337] C. Shen, Y. Qi, J. Wang, G. Cai, Z. Zhu, An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive, Eng. Appl. Artif. Intell. 76 (2018) 170-184.
[338] X. Liu, Q. Zhou, J. Zhao, H. Shen, X. Xiong, Fault diagnosis of rotating machinery under noisy environment conditions based on a 1-D convolutional autoencoder and 1-D convolutional neural network, Sensors 19 (2019) 972.
[339] H. Yu, K. Wang, Y. Li, Multiscale representations fusion with joint multiple reconstructions autoencoder for intelligent fault diagnosis, IEEE Signal Process. Lett. 25 (2018) 1880-1884.
[340] H. Shao, H. Jiang, X. Li, S. Wu, Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine, Knowledge-Based Syst. 140 (2018) 1-14.
[341] W. Mao, W. Feng, X. Liang, A novel deep output kernel learning method for bearing fault structural diagnosis, Mech. Syst. Signal Process. 117 (2019) 293-318.
[342] Z. Yang, X. Wang, J. Zhong, Representational learning for fault diagnosis of wind turbine equipment: A multi-layered extreme learning machines approach, Energies 9 (2016) 379.
[343] Z. Chen, W. Li, Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network, IEEE Trans. Instrum. Meas. 66 (2017) 1693-1702.
[344] J. Li, X. Li, D. He, Y. Qu, A novel method for early gear pitting fault diagnosis using stacked SAE and GBRBM, Sensors 19 (2019) 758.
[345] J. Wang, S. Li, B. Han, Z. An, Y. Xin, W. Qian, Q. Wu, Construction of a batch-normalized autoencoder

network and its application in mechanical intelligent fault diagnosis, Meas. Sci. Technol. 30 (2019) 015106.

[346] S.R. Saufi, Z.A. bin Ahmad, M.S. Leong, M.H. Lim, Differential evolution optimization for resilient stacked sparse autoencoder and its applications on bearing fault diagnosis, Meas. Sci. Technol. 29 (2018) 125002.

[347] X. He, D. Wang, Y. Li, C. Zhou, A novel bearing fault diagnosis method based on gaussian restricted boltzmann machine, Math. Prob. Eng. (2016) Doi: 10.1155/2016/2957083.

[348] H. Jiang, H. Shao, X. Chen, J. Huang, A feature fusion deep belief network method for intelligent fault diagnosis of rotating machinery, J. Intell. Fuzzy Syst. 34 (2018) 3513-3521.

[349] D. Han, N. Zhao, P. Shi, A new fault diagnosis method based on deep belief network and support vector machine with Teager-Kaiser energy operator for bearings, Adv. Mech. Eng. 9 (2017) 1-11.

[350] S. Tang, C. Shen, D. Wang, S. Li, W. Huang, Z. Zhu, Adaptive deep feature learning network with Nesterov momentum and its application to rotating machinery fault diagnosis, Neurocomputing 305 (2018) 1-14.

[351] J. Xie, G. Du, C. Shen, N. Chen, L. Chen, Z. Zhu, An end-to-end model based on improved adaptive deep belief network and its application to bearing fault diagnosis, IEEE Access 6 (2018) 63584-63596.

[352] H. Shao, H. Jiang, F. Wang, Y. Wang, Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet, ISA Trans. 69 (2017) 187-201.

[353] H. Shao, H. Jiang, X. Zhang, M. Niu, Rolling bearing fault diagnosis using an optimization deep belief network, Meas. Sci. Technol. 26 (2015) 115002.

[354] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, S. Wu, Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing, Mech. Syst. Signal Process. 100 (2018) 743-765.

[355] H. Shao, H. Jiang, H. Zhang, T. Liang, Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network, IEEE Trans. Ind. Electron. 65 (2018) 2727-2736.

[356] P. Tamilselvan, P.F. Wang, Failure diagnosis using deep belief learning based health state classification,

Reliab. Eng. Syst. Saf. 115 (2013) 124-135.
[357] J. Sun, R. Wyss, A. Steinecker, P. Glocker, Automated fault detection using deep belief networks for the quality inspection of electromotors, tm-Tech. Mess. 81 (2014) 255-263.
[358] V.T. Tran, F. AlThobiani, A. Ball, An approach to fault diagnosis of reciprocating compressor valves using Teager-Kaiser energy operator and deep belief networks, Expert Syst. Appl. 41 (2014) 4113-4122.
[359] J. Qiu, W. Liang, L. Zhang, X. Yu, M. Zhang, The early-warning model of equipment chain in gas pipeline based on DNN-HMM, J. Nat. Gas Sci. Eng. 27 (2015) 1710-1722.
[360] Z. Gao, C. Ma, D. Song, Y. Liu, Deep quantum inspired neural network with application to aircraft fuel system fault diagnosis, Neurocomputing 238 (2017) 13-23.
[361] J. Yin, W. Zhao, Fault diagnosis network design for vehicle on-board equipments of high-speed railway: A deep learning approach, Eng. Appl. Artif. Intell. 56 (2016) 250-259.
[362] J. He, S. Yang, C. Gan, Unsupervised fault diagnosis of a gear transmission chain using a deep belief network, Sensors 17 (2017) 1564.
[363] S. Shao, W. Sun, R. Yan, P. Wang, R.X. Gao, A deep learning approach for fault diagnosis of induction motors in manufacturing, Chin. J. Mech. Eng. 30 (2017) 1347-1356.
[364] X. Wang, J. Huang, G. Ren, D. Wang, A hydraulic fault diagnosis method based on sliding-window spectrum feature and deep belief network, J. Vibroeng. 19 (2017) 4272-4284.
[365] Y. Guo, Z. Tan, H. Chen, G. Li, J. Wang, R. Huang, J. Liu, T. Ahmad, Deep learning-based fault diagnosis of variable refrigerant flow air-conditioning system for building energy saving, Appl. Energy 225 (2018) 732-745.
[366] D. Yu, Z. Chen, K. Xiahou, M. Li, T. Ji, Q. Wu, A radically data-driven method for fault detection and diagnosis in wind turbines, Int. J. Electr. Power Energy Syst. 99 (2018) 577-584.
[367] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, T. Chen, Recent

advances in convolutional neural networks, Pattern Recognit. 77 (2018) 354-377.

[368] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, in: *International Conference on International Conference on Machine Learning*, Haifa, Israel, 2010, pp. 807-814.

[369] M.M.M. Islam, J.M. Kim, Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network, Comput. Ind. 106 (2019) 142-153.

[370] X. Ding, Q. He, Energy-fluctuated multiscale feature learning with deep convnet for intelligent spindle bearing fault diagnosis, IEEE Trans. Instrum. Meas. 66 (2017) 1926-1935.

[371] Y. Han, B. Tang, L. Deng, Multi-level wavelet packet fusion in dynamic ensemble convolutional neural network for fault diagnosis, Measurement 127 (2018) 246-255.

[372] S. Guo, T. Yang, W. Gao, C. Zhang, A novel fault diagnosis method for rotating machinery based on a convolutional neural network, Sensors 18 (2018) 1429.

[373] S. Guo, T. Yang, W. Gao, C. Zhang, Y. Zhang, An intelligent fault diagnosis method for bearings with variable rotating speed based on pythagorean spatial pyramid pooling CNN, Sensors 18 (2018) 3857.

[374] W. Sun, B. Yao, N. Zeng, B. Chen, Y. He, X. Cao, W. He, An intelligent gear fault diagnosis methodology using a complex wavelet enhanced convolutional neural network, Materials 10 (2017) 790.

[375] X. Cao, B. Chen, B. Yao, W. He, Combining translation-invariant wavelet frames and convolutional neural network for intelligent tool wear state identification, Comput. Ind. 106 (2019) 71-84.

[376] D. Zhao, T. Wang, F. Chu, Deep convolutional neural network based planet bearing fault classification, Comput. Ind. 107 (2019) 59-66.

[377] S. Li, G. Liu, X. Tang, J. Lu, J. Hu, An ensemble deep convolutional neural network model with improved D-S evidence fusion for bearing fault diagnosis, Sensors 17 (2017) 1729.

[378] R. Liu, G. Meng, B. Yang, C. Sun, X. Chen, Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine, IEEE Trans. Ind. Inform. 13 (2017) 1310-1320.

[379] C. Lu, Z. Wang, B. Zhou, Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification, Adv. Eng. Inf. 32 (2017) 139-151.

[380] F. Wang, H. Jiang, H. Shao, W. Duan, S. Wu, An adaptive deep convolutional neural network for rolling bearing fault diagnosis, Meas. Sci. Technol. 28 (2017) 095005.

[381] S. Wang, J. Xiang, Y. Zhong, Y. Zhou, Convolutional neural network-based hidden Markov models for rolling element bearing fault identification, Knowledge-Based Syst. 144 (2018) 65-76.

[382] M. Xia, T. Li, L. Xu, L. Liu, C.W. de Silva, Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks, IEEE-ASME Trans. Mechatron. 23 (2018) 101-110.

[383] L. Wen, X. Li, L. Gao, Y. Zhang, A new convolutional neural network-based data-driven fault diagnosis method, IEEE Trans. Ind. Electron. 65 (2018) 5990-5998.

[384] P. Zhou, G. Zhou, Z. Zhu, C. Tang, Z. He, W. Li, F. Jiang, Health monitoring for balancing tail ropes of a hoisting system using a convolutional neural network, Appl. Sci.-Basel 8 (2018) 1346.

[385] Z. Yuan, L. Zhang, L. Duan, A novel fusion diagnosis method for rotor system fault based on deep learning and multi-sourced heterogeneous monitoring data, Meas. Sci. Technol. 29 (2018) 115005.

[386] O. Janssens, R. Van de Walle, M. Loccufier, S. Van Hoecke, Deep learning for infrared thermal image based machine health monitoring, IEEE-ASME Trans. Mechatron. 23 (2018) 151-159.

[387] L. Eren, T. Ince, S. Kiranyaz, A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier, J. Signal Process. Syst. Signal Image Video Technol. 91 (2019) 179-189.

[388] Y. Chen, G. Peng, C. Xie, W. Zhang, C. Li, S. Liu, Acdin: Bridging the gap between artificial and real bearing damages for bearing fault diagnosis, Neurocomputing 294 (2018) 61-71.

[389] L. Eren, Bearing fault detection by one-dimensional convolutional neural networks, Math. Prob. Eng. (2017) Doi: 10.1155/2017/8617315.

[390] F. Jia, Y. Lei, N. Lu, S. Xing, Deep normalized convolutional neural network for imbalanced fault

classification of machinery and its understanding via visualization, Mech. Syst. Signal Process. 110 (2018) 349-367.

[391] D.K. Appana, A. Prosvirin, J.M. Kim, Reliable fault diagnosis of bearings with varying rotational speeds using envelope spectrum and convolution neural networks, Soft Comput. 22 (2018) 6719-6729.

[392] L. Jing, M. Zhao, P. Li, X. Xu, A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox, Measurement 111 (2017) 1-10.

[393] J. Jiao, M. Zhao, J. Lin, J. Zhao, A multivariate encoder information based convolutional neural network for intelligent fault diagnosis of planetary gearboxes, Knowledge-Based Syst. 160 (2018) 237-250.

[394] L. Jing, T. Wang, M. Zhao, P. Wang, An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox, Sensors 17 (2017) 414.

[395] Y. Han, B. Tang, L. Deng, An enhanced convolutional neural network with enlarged receptive fields for fault diagnosis of planetary gearboxes, Comput. Ind. 107 (2019) 50-58.

[396] Y. Yao, H. Wang, S. Li, Z. Liu, G. Gui, Y. Dan, J. Hu, End-to-end convolutional neural network model for gear fault diagnosis based on sound signals, Appl. Sci.-Basel 8 (2018) 1584.

[397] X. Li, J. Li, Y. Qu, D. He, Gear pitting fault diagnosis using integrated CNN and GRU network with both vibration and acoustic emission signals, Appl. Sci.-Basel 9 (2019) 768.

[398] G. Jiang, H. He, J. Yan, P. Xie, Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox, IEEE Trans. Ind. Electron. 66 (2019) 3196-3207.

[399] T. Ince, S. Kiranyaz, L. Eren, M. Askar, M. Gabbouj, Real-time motor fault detection by 1-D convolutional neural networks, IEEE Trans. Ind. Electron. 63 (2016) 7067-7075.

[400] J. Yan, H. Zhu, X. Yang, Y. Cao, L. Shao, Research on fault diagnosis of hydraulic pump using convolutional neural network, J. Vibroeng. 18 (2016) 5141-5152.

[401] D. Verstraete, A. Ferrada, E.L. Droguett, V. Meruane, M. Modarres, Deep learning enabled fault diagnosis

using time-frequency image analysis of rolling element bearings, Shock Vibrat. (2017) Doi: 10.1155/2017/5067651.

[402] D. Guo, M. Zhong, H. Ji, Y. Liu, R. Yang, A hybrid feature model and deep learning based fault diagnosis for unmanned aerial vehicle sensors, Neurocomputing 319 (2018) 155-163.

[403] Y. Xin, S. Li, C. Cheng, J. Wang, An intelligent fault diagnosis method of rotating machinery based on deep neural networks and time-frequency analysis, J. Vibroeng. 20 (2018) 2321-2335.

[404] X. Chen, L. Peng, G. Cheng, C. Luo, Research on degradation state recognition of planetary gear based on multiscale information dimension of ssd and cnn, Complexity (2019) Doi: 10.1155/2019/8716979.

[405] T. Han, C. Liu, L. Wu, S. Sarkar, D. Jiang, An adaptive spatiotemporal feature learning approach for fault diagnosis in complex systems, Mech. Syst. Signal Process. 117 (2019) 170-187.

[406] Z. Zhu, G. Peng, Y. Chen, H. Gao, A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis, Neurocomputing 323 (2019) 62-75.

[407] H. Jiang, F. Wang, H. Shao, H. Zhang, Rolling bearing fault identification using multilayer deep learning convolutional neural network, J. Vibroeng. 19 (2017) 138-149.

[408] S. Suh, H. Lee, J. Jo, P. Lukowicz, Y.O. Lee, Generative oversampling method for imbalanced data on bearing fault detection and diagnosis, Appl. Sci.-Basel 9 (2019) 746.

[409] R. Huang, Y. Liao, S. Zhang, W. Li, Deep decoupling convolutional neural network for intelligent compound fault diagnosis, IEEE Access 7 (2019) 1848-1858.

[410] W. Zhang, X. Li, Q. Ding, Deep residual learning-based fault diagnosis method for rotating machinery, ISA Trans. (2018) Doi: 10.1016/j.isatra.2018.12.025.

[411] M. Zhao, M. Kang, B. Tang, M. Pecht, Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes, IEEE Trans. Ind. Electron. 65 (2018) 4290-4300.

[412] M. Zhao, M. Kang, B. Tang, M. Pecht, Multiple wavelet coefficients fusion in deep residual networks for

fault diagnosis, IEEE Trans. Ind. Electron. 66 (2019) 4696-4706.
[413] S. Ma, F. Chu, Q. Han, Deep residual learning with demodulated time-frequency features for fault diagnosis of planetary gearbox under nonstationary running conditions, Mech. Syst. Signal Process. 127 (2019) 190-201.
[414] D. Peng, Z. Liu, H. Wang, Y. Qin, L. Jia, A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains, IEEE Access 7 (2019) 10278-10293.
[415] L. Su, L. Ma, N. Qin, D. Huang, A. Kemp, Fault diagnosis of high-speed train bogie by residual-squeeze net, IEEE Trans. Ind. Inform. (2019) Doi: 10.1109/TII.2019.2907373.
[416] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load, Mech. Syst. Signal Process. 100 (2018) 439-453.
[417] B. Yang, Y. Lei, F. Jia, S. Xing, A transfer learning method for intelligent fault diagnosis from laboratory machines to real-case machines, in: *International Conference on Sensing, Diagnostics, Prognostics, and Control*, Xian, China, 2018, pp. 35 -40.
[418] C. Chen, Z. Li, J. Yang, B. Liang, A cross domain feature extraction method based on transfer component analysis for rolling bearing fault diagnosis, in: *Chinese Control and Decision Conference*, Chongqing, China, 2017, pp. 5622-5626.
[419] J. Xie, L. Zhang, L. Duan, J. Wang, On cross-domain feature fusion in gearbox fault diagnosis under various operating conditions based on transfer component analysis, in: *IEEE International Conference on Prognostics and Health Management*, Ottawa, Canada, 2016, pp. 1-6.
[420] Z. Tong, W. Li, B. Zhang, F. Jiang, G. Zhou, Bearing fault diagnosis under variable working conditions based on domain adaptation using feature transfer learning, IEEE Access 6 (2018) 76187-76197.
[421] Z. Tong, W. Li, B. Zhang, M. Zhang, Bearing fault diagnosis based on domain adaptation using transferable

features under different working conditions, Shock Vibrat. (2018) Doi: 10.1155/2018/6714520.

[422] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals, Sensors 17 (2017) 425.

[423] W. Qian, S. Li, J. Wang, Y. Xin, H. Ma, A new deep transfer learning network for fault diagnosis of rotating machine under variable working conditions, in: *Prognostics and System Health Management Conference*, Chongqing, China, 2018, pp. 1010-1016.

[424] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, T. Zhang, Deep model based domain adaptation for fault diagnosis, IEEE Trans. Ind. Electron. 64 (2017) 2296-2305.

[425] L. Wen, L. Gao, X. Li, A new deep transfer learning based on sparse auto-encoder for fault diagnosis, IEEE Trans. Syst. Man Cybern. Syst. 49 (2019) 136-144.

[426] X. Li, W. Zhang, Q. Din, A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning, Neurocomputing 310 (2018) 77-95.

[427] B. Zhang, W. Li, X. Li, N.G. See-Kiong, Intelligent fault diagnosis under varying working conditions based on domain adaptive convolutional neural networks, IEEE Access 6 (2018) 66367-66384.

[428] C. Wang, Z. Lv, J. Zhao, W. Wang, Heterogeneous transfer learning based on stack sparse auto-encoders for fault diagnosis, in: *Chinese Automation Congress*, Xian, China, 2018, pp. 4277 -4281.

[429] W. Qian, S. Li, J. Wang, A new transfer learning method and its application on rotating machine fault diagnosis under variant working conditions, IEEE Access 6 (2018) 69907-69917.

[430] X. Wang, H. He, L. Li, A hierarchical deep domain adaptation approach for fault diagnosis of power plant thermal system, IEEE Trans. Ind. Inform. (2019) Doi: 10.1109/TII.2019.2899118.

[431] Y. Xu, Y. Sun, X. Liu, Y. Zheng, A digital-twin-assisted fault diagnosis using deep transfer learning, IEEE Access 7 (2019) 19990-19999.

[432] J. Wang, J. Xie, L. Zhang, L. Duan, A factor analysis based transfer learning method for gearbox diagnosis

under various operating conditions, in: *International Symposium on Flexible Automation*, Cleveland, USA, 2016, pp. 81-86.

[433] Z. Bo, L. Wei, T. Zhe, Z. Meng, Bearing fault diagnosis under varying working condition based on domain adaptation, 2017, Available: arXiv preprint arXiv: 1707.09890.

[434] H. Zheng, R. Wang, Y. Yang, Y. Li, M. Xu, Intelligent fault identification based on multi-source domain generalization towards actual diagnosis scenario, IEEE Trans. Ind. Electron. (2019) Doi: 10.1109/TIE.2019.2898619.

[435] Y. Xie, T. Zhang, A transfer learning strategy for rotation machinery fault diagnosis based on cycle-consistent generative adversarial networks, in: *Chinese Automation Congress*, Xian, China, 2018, pp. 1309 - 1313.

[436] X. Li, W. Zhang, Q. Ding, Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks, IEEE Trans. Ind. Electron. 66 (2019) 5525-5534.

[437] T. Han, C. Liu, W. Yang, D. Jiang, A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults, Knowledge-Based Syst. 165 (2019) 474-487.

[438] R. Zhang, H. Tao, L. Wu, Y. Guan, Transfer learning with neural networks for bearing fault diagnosis in changing working conditions, IEEE Access 5 (2017) 14347-14357.

[439] P. Cao, S. Zhang, J. Tang, Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning, IEEE Access 6 (2018) 26241-26253.

[440] S. Shao, S. McAleer, R. Yan, P. Baldi, Highly-accurate machine fault diagnosis using deep transfer learning, IEEE Trans. Ind. Inform. 15 (2019) 2446-2455.

[441] Y. Li, N. Wang, J. Shi, X. Hou, J. Liu, Adaptive batch normalization for practical domain adaptation, Pattern Recognit. 80 (2018) 109-117.

[442] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio,

Generative adversarial nets, in: *Conference on Neural Information Processing Systems*, Montreal, Canada, 2014, pp. 2672-2680.

[442] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: *Conference on Neural Information Processing Systems*, Montreal, Canada, 2014, pp. 2672-2680.

[443] B. Yang, Y. Lei, F. Jia, N. Li, Z. Du, A polynomial kernel induced distance metric to improve deep transfer learning for fault diagnosis of machines. IEEE Trans. Ind. Electron. (2019) Doi: 10.1109/TIE.2019.2953010.

Table 1 Summary of applications of ANN to machine fault diagnosis.

Table 2 Summary of applications of SVM to machine fault diagnosis.

Table 3 Summary of the applications of CNN to machine fault diagnosis

Table 4 Categories of transfer scenarios in machine fault diagnosis.

Table 5 Summary of applications of transfer learning to machine fault diagnosis.

Fig. 1. Development and milestones of IFD using machine learning.

Fig. 2. Diagnosis process of IFD using traditional machine learning theories.

Fig. 3. Architecture of expert system-based diagnosis models.

Fig. 4. Architecture of BPNN with two hidden layers.

Fig. 5. Classification by the linear SVM.

Fig. 6. Illustration of the kNN algorithm.

Fig. 7. Illustration of PGM: (a) Bayesian classifier, and (b) hidden Markov model.

Fig. 8. Illustration of the decision tree.

Fig. 9. Diagnosis process of IFD using deep learning theories.

Fig. 10. Architecture of AE.

Fig. 11. Greedy layer-wise pre-training for stacked AE.

Fig. 12. Architecture of RBM.

Fig. 13. Convolution and pooling process: (a) Convolution process by using the kernel $\boldsymbol{k}^c \in \Re^{2\times1\times1}$, and (b) pooling process by using the filter $\boldsymbol{s}^{2\times2}$.

Fig. 14. Architecture of the residual block.

Fig. 15. Steps of feature-based transfer learning approaches [26].

Fig. 16. Illustration of the TrAdaboost algorithm: (a) directly train the diagnosis models with the source and target domain samples, and (b) train the diagnosis models by using TrAdaboost

algorithm.

Fig. 17. Roadmap of applications of machine learning to machine fault diagnosis.

Table 1 Summary of applications of ANN to machine fault diagnosis.

Objects	References	Methodologies
Bearings	Yang et al. [101], Samanta et al. [102], Yu et al. [103], Castejon et al. [104], Muruganatham et al. [105], Unal et al. [106], Zarei et al. [107], Almeida et al. [108], and Ahmed et al. [109]	BPNN
	Wang et al. [110], Lei et al. [111], Vijay et al. [112], Jiang et al. [113], and Tang et al. [114]	RBFN
	Lei et al. [115], Wu et al. [116]	WNN
Gears	Abu-Mahfouz et al. [117], Rafiee et al. [118], Hajnayeb et al. [119], Cerrada et al. [120], Kane et al. [121], Waqar et al. [122], and Tyagi et al. [123]	BPNN
	Lai et al. [124], Li et al. [125], and Liu et al. [126]	RBFN
	Chen et al. [127]	WNN
Motors	Ayhan et al. [128], Sadeghian et al. [129], Arabaci et al. [130], Cabal-Yepez et al. [131], Hernandez-Vargas et al. [132], and Moosavi et al. [133]	BPNN
	Ghate et al. [134], and Palacios et al. [135]	RBFN
	Boukra et al. [136]	WNN
Engines	Sharkey et al. [137], Lu et al. [138], Chen et al. [139, 140], Khazaee et al. [141, 142], and Zabihi-Hersari et al. [143]	BPNN
	Wu et al. [144, 145]	RBFN
	Shen et al. [146], Zhang et al. [147]	WNN
Others	Kuo et al. [148], Ilott et al. [149], Wu et al. [150], Mohammed et al. [151], Walker et al. [152], Malik et al. [153], and McCormick et al. [154, 155]	BPNN
	Wu et al. [156], and Villanueva et al. [157]	RBFN

Liu et al. [158], Chen et al. [159], Guo et al. [160], Xiao et al. [161], and Jin et al. [162]	WNN

Table 2 Summary of applications of SVM to machine fault diagnosis.

Objects	References	Methodologies
Bearings	Abbasion et al. [224], Yang et al. [225], Xian et al. [226], Hao et al. [227], Gryllias et al. [228], Islam et al. [229], Sugumaran et al. [192], Zheng et al. [202], Jack et al. [208], Rojas et al. [210], HungLinh et al. [230], Kang et al. [216], and Li et al. [222]	OAA-based SVM
	Yang et al. [231], Yang et al. [232], Wu et al. [233, 234], Saidi et al. [235], Zhu et al. [236], Ziani et al. [237], Islam et al. [238], Widodo et al. [211], Zhang et al. [239], and Zhu et al. [219]	OAO-based SVM
	Sugumaran et al. [192, 195], Wang et al. [196], Dong et al. [197], Zhang et al. [201], Li et al. [177-179], Xu et al. [186], Zheng et al. [202], and Chen et al. [206]	Varieties of SVM
	Jack et al. [208, 209], Samanta et al. [215], Rojas et al. [210], Widodo et al. [211], Li et al. [223], Zhang et al. [239], Chen et al. [206], Zhu et al. [219], Dong et al. [221], HungLinh et al. [230], Kang et al. [216], Su et al. [220], Zhu et al. [217], and Li et al. [222]	SVM with optimization algorithm
Gears	Liu et al. [240], and Li et al. [241]	OAA-based SVM
	Lu et al. [242], Cheng et al. [243], Xing et al. [244], Liu et al. [245], Shen et al. [246], Jiang et al. [188], Heidari et al. [198], and Bordoloi et al. [247]	OAO-based SVM
	Saravanan et al. [193], Shen et al. [246], Tang et al. [182], Jiang et al. [188], Yang et al. [180], Heidari et al. [189, 198], Jiang et al. [190], Li et al. [63, 181, 187], and Chen et al. [200]	Varieties of SVM
	Samanta et al. [218], Li et al. [212], Chen et al. [200], Bordoloi et al. [247],	SVM with

	and Zhang et al. [248]	optimization algorithm
Motors	Widodo et al. [213, 249], Ebrahimi et al. [250], Shahriar et al. [251], Kang et al. [252], Singh et al. [64], and Ebrahimi et al. [203]	OAA-based SVM
	Kurek et al. [253], Gangsar et al. [254-256], Sun et al. [257], Martinez-Morales et al. [258], Keskes et al. [199], and Widodo et al. [214]	OAO-based SVM
	Tsoumas et al. [259], Bacha et al. [174], Salem et al. [184], Kang et al. [183], Keskes et al. [176, 199], Ebrahimi et al. [203], and Zgarni et al. [175]	Varieties of SVM
	Widodo et al. [213, 214]	Optimization
Engines	Li et al. [260], and Zhang et al. [51]	OAA-based SVM
	Lee et al. [261], Wang et al. [262], Liu et al. [207], and Jafarian et al. [263]	OAO-based SVM
	Vong et al. [191], Jena et al. [264], Liu et al. [207], Cai et al. [185], and Li et al. [205]	Varieties of SVM
	Li et al. [205]	SVM with Optimization algorithm
Others	Namdari et al. [265], and Jegadeeshwaran et al. [169]	OAA-based SVM
	Rapur et al. [170, 171], Pang et al. [163], Hang et al. [204], Tang et al. [167], and Zhang et al. [168]	OAO-based SVM
	Chiang et al. [194], Yuan et al. [166], and Jin et al. [165], Tang et al. [182], Wang et al. [164], Hang et al. [204]	Varieties of SVM
	Tang et al. [167], and Zhang et al. [168]	SVM with optimization

	algorithm

1

Table 3 Summary of the applications of CNN to machine fault diagnosis

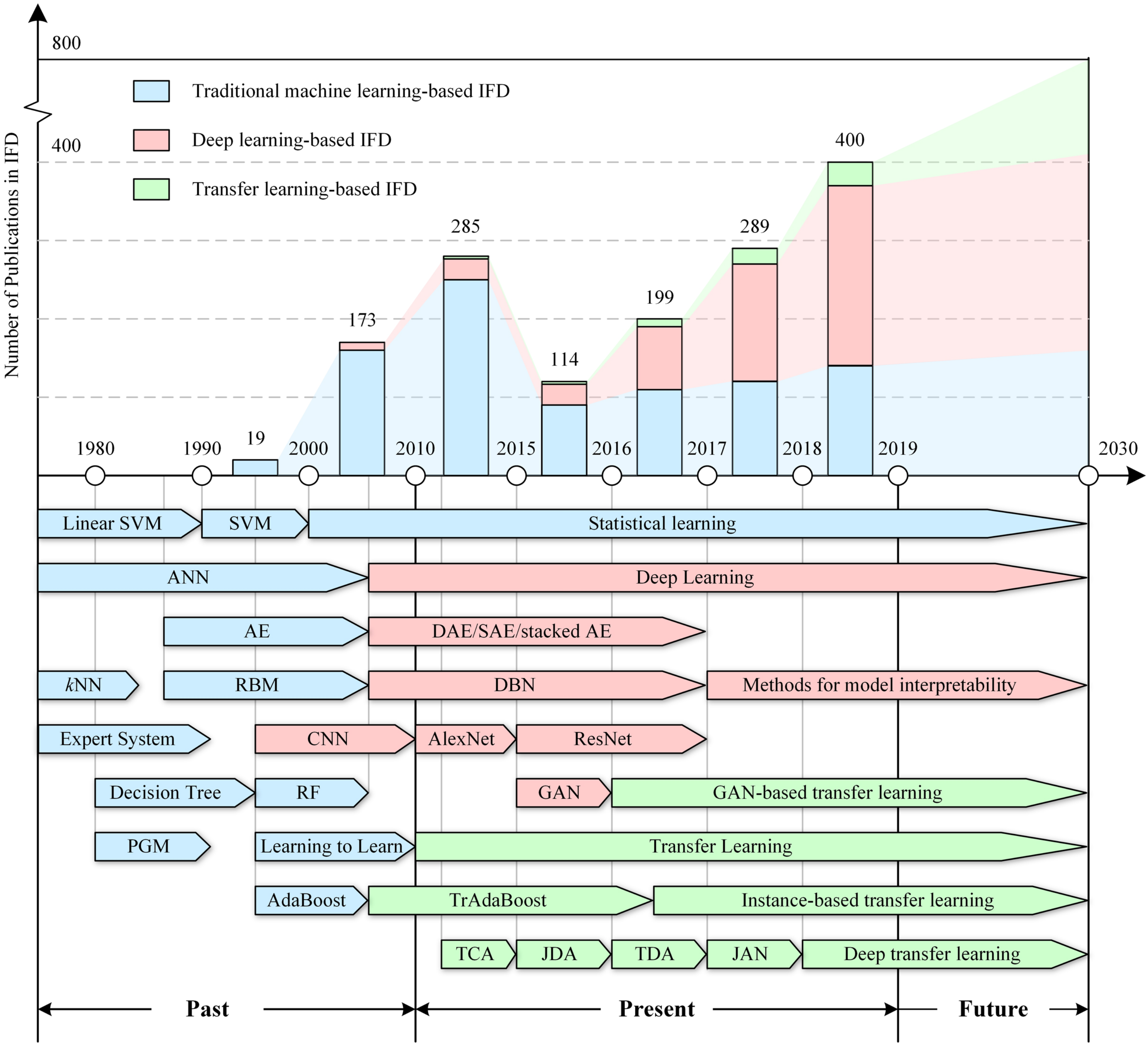
Architectures	Input data	References
2D CNN	Time-frequency spectrum	Ding et al. [370], Sun et al. [374], Verstraete et al. [401], Guo et al. [402], Han et al. [371], Xin et al. [403], Guo et al. [372, 373], Cao et al. [375], Chen et al. [404], Han et al. [405], Islam et al. [369], Zhao et al. [376], and Zhu et al. [406]
	Reshaped matrix	Jiang et al. [407], Li et al. [377], Liu et al. [378], Lu et al. [379], Wang et al. [381], Xia et al. [382], and Wang et al. [380]
	Images	Janssens et al. [386], Wen et al. [383], Yuan et al. [385], Zhou et al. [384], and Suh et al. [408]
1D CNN	Raw data or frequency spectrum	Ince et al. [399], Yan et al. [400], Eren et al. [387, 389], Jing et al. [392, 394], Appana et al. [391], Chen et al. [388], Jia et al. [390], Jiao et al. [393], Yao et al. [396], Han et al. [395], Huang et al. [409], Jiang et al. [398], and Li et al. [397]

Table 4 Categories of transfer scenarios in machine fault diagnosis.

Transfer scenarios	Assumptions		Factors leading to transfer scenarios
	Source domain	Target domain	
Data collected from the identical machine (TIM)	Available labels	• Minority of labels • Unavailable labels	• Varying speed • Varying load • Various working environments
Data collected from different machines (TDM)	Available labels	• Minority of labels • Unavailable labels	• Different machine specifications • Diverse structures • Different measurement environments • Different working environments

Table 5 Summary of applications of transfer learning to machine fault diagnosis.

Approaches	References	Transfer scenarios		Methodologies
		TIM	TDM	
Feature-based	Chen et al. [418], Xie et al. [419], and Tong et al. [420, 421]	√		TCA and JDA
	Zhang et al. [416, 422], Qian et al. [423], and Peng et al. [414]	√		Deep learning and AdaBN
	Lu et al. [424], Wen et al. [425], Li et al. [426], Zhang et al. [427], Wang et al. [428], Qian et al. [429], Wang et al. [430] Yang et al. [26, 417, 443], and Xu et al. [431]	√	√	Deep transfer learning
	Wang et al. [432], Zhang et al. [433], and Zheng et al. [434]	√	√	Transfer factor analysis, and subspace alignment
GAN-based	Xie et al. [435], Li et al. [436], Han et al. [437], and Guo et al. [35]	√	√	GAN and MMD
Instance-based	Shen et al. [36]	√		TrAdaBoost
Parameter-based	Zhang et al. [438], Hasan et al. [37], Cao et al. [439], and Shao et al. [440]	√		Deep learning and fine tuning



2 Fig. 1. Development and milestones of IFD using machine learning.

3

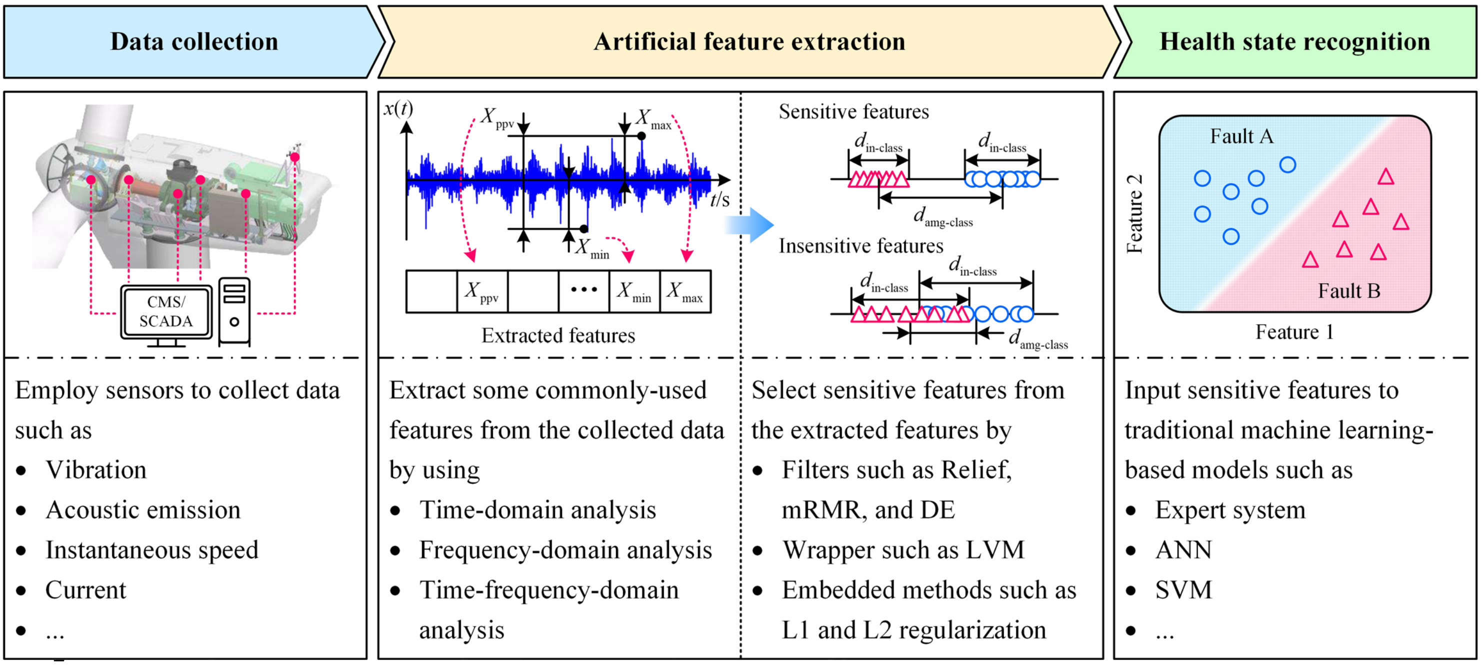


Fig. 2. Diagnosis process of IFD using traditional machine learning theories.

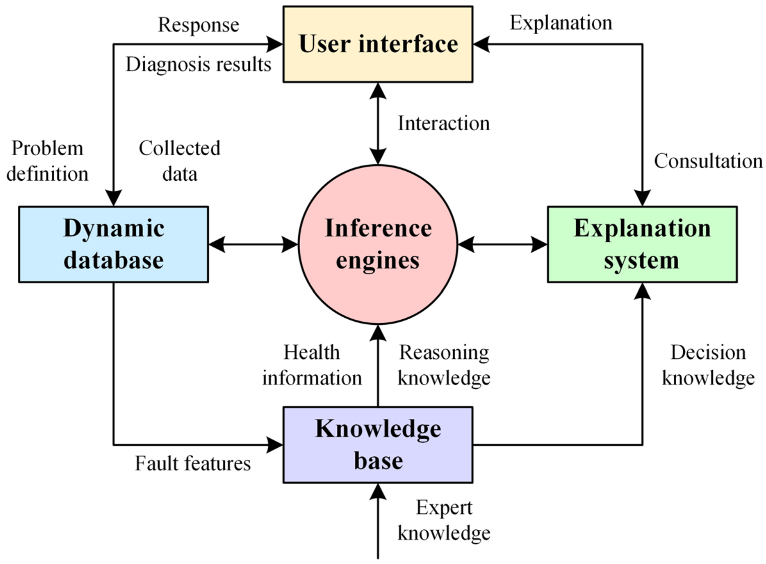


Fig. 3. Architecture of expert system-based diagnosis models.

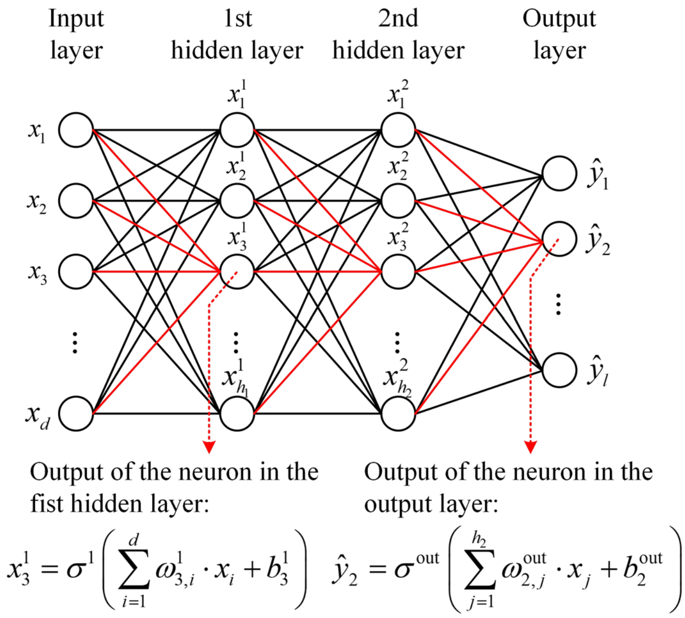


Fig. 4. Architecture of BPNN with two hidden layers.

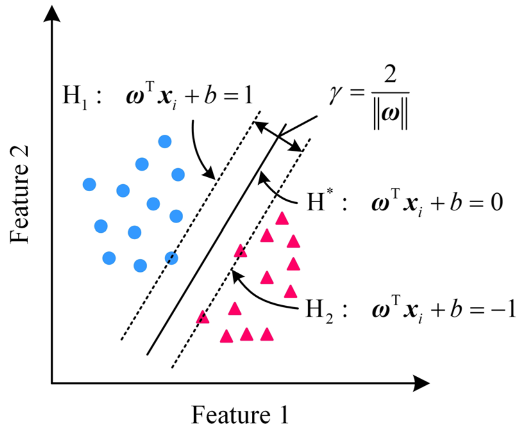


Fig. 5. Classification by the linear SVM.

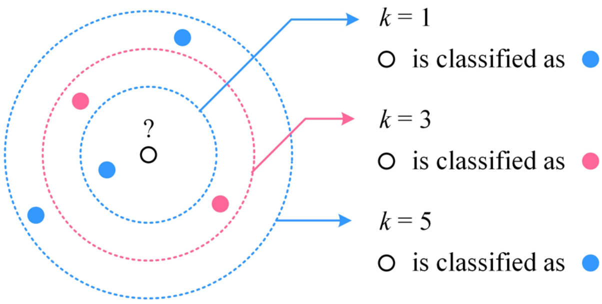


Fig. 6. Illustration of the kNN algorithm.

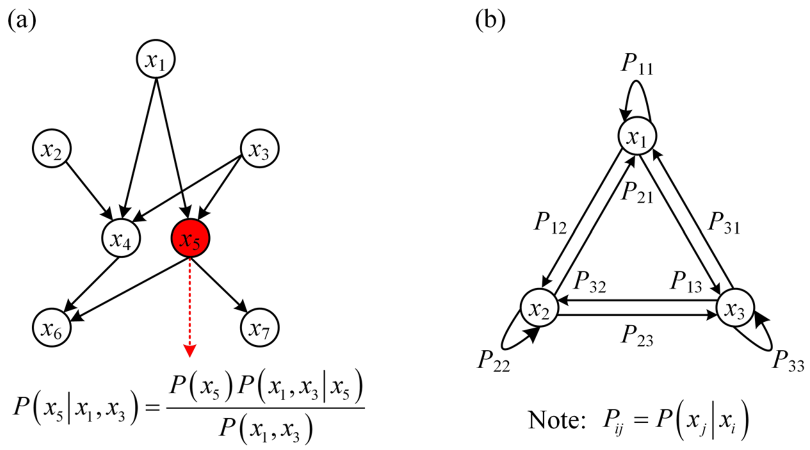


Fig. 7. Illustration of PGM: (a) Bayesian classifier, and (b) hidden Markov model.

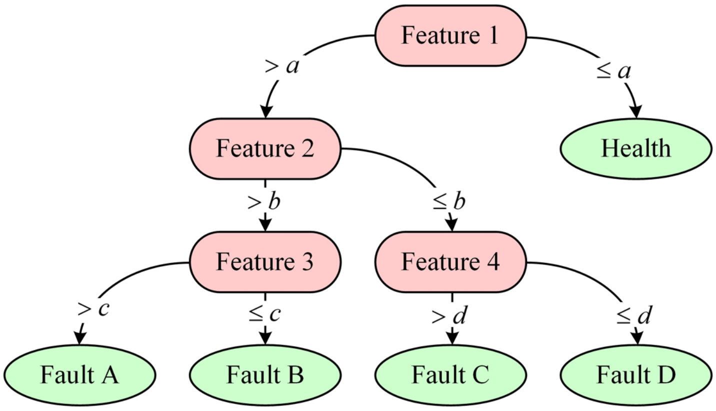


Fig. 8. Illustration of the decision tree.

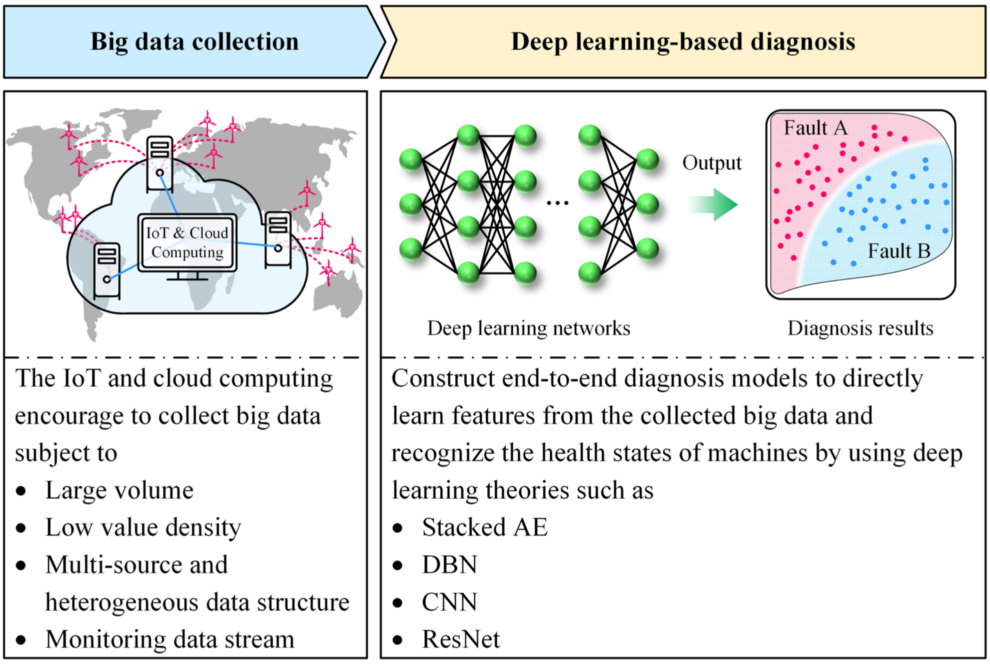


Fig. 9. Diagnosis process of IFD using deep learning theories.

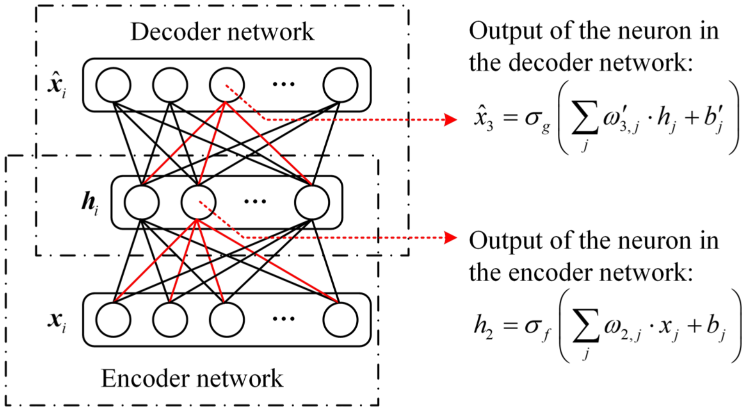


Fig. 10. Architecture of AE.

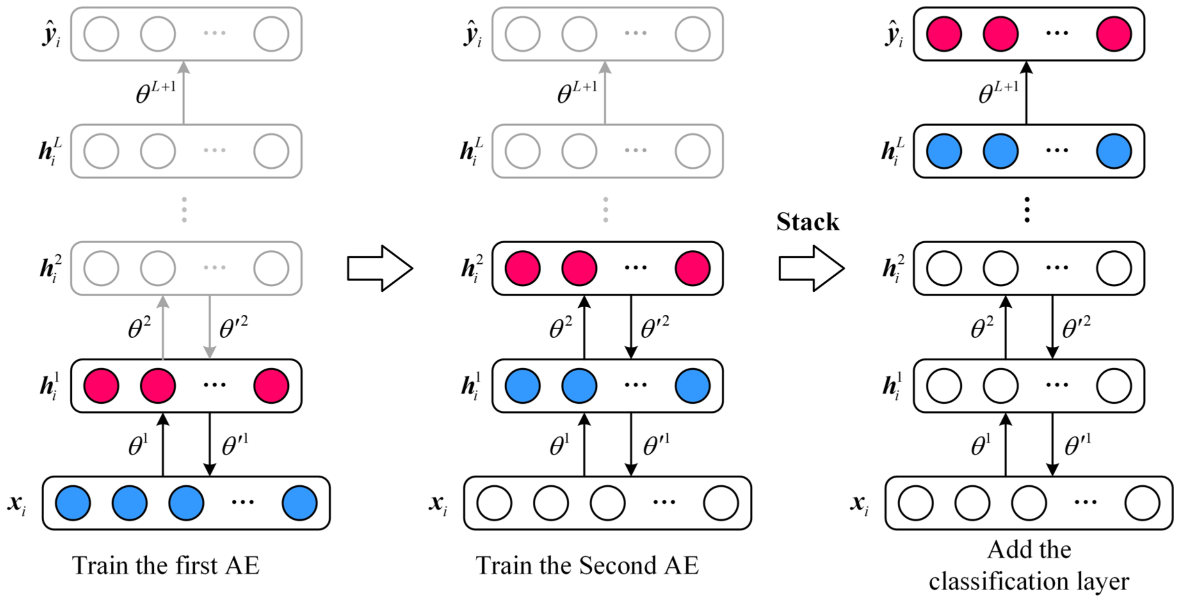


Fig. 11. Greedy layer-wise pre-training for stacked AE.

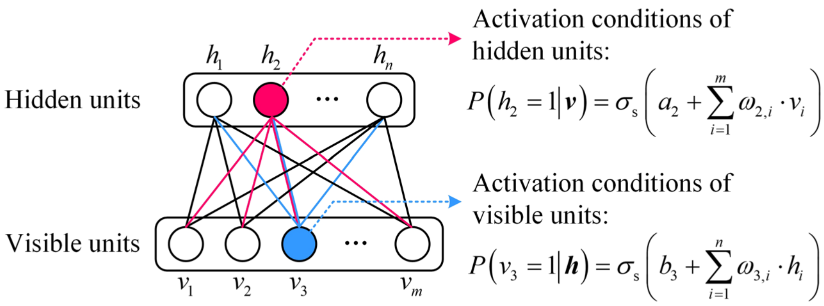


Fig. 12. Architecture of RBM.

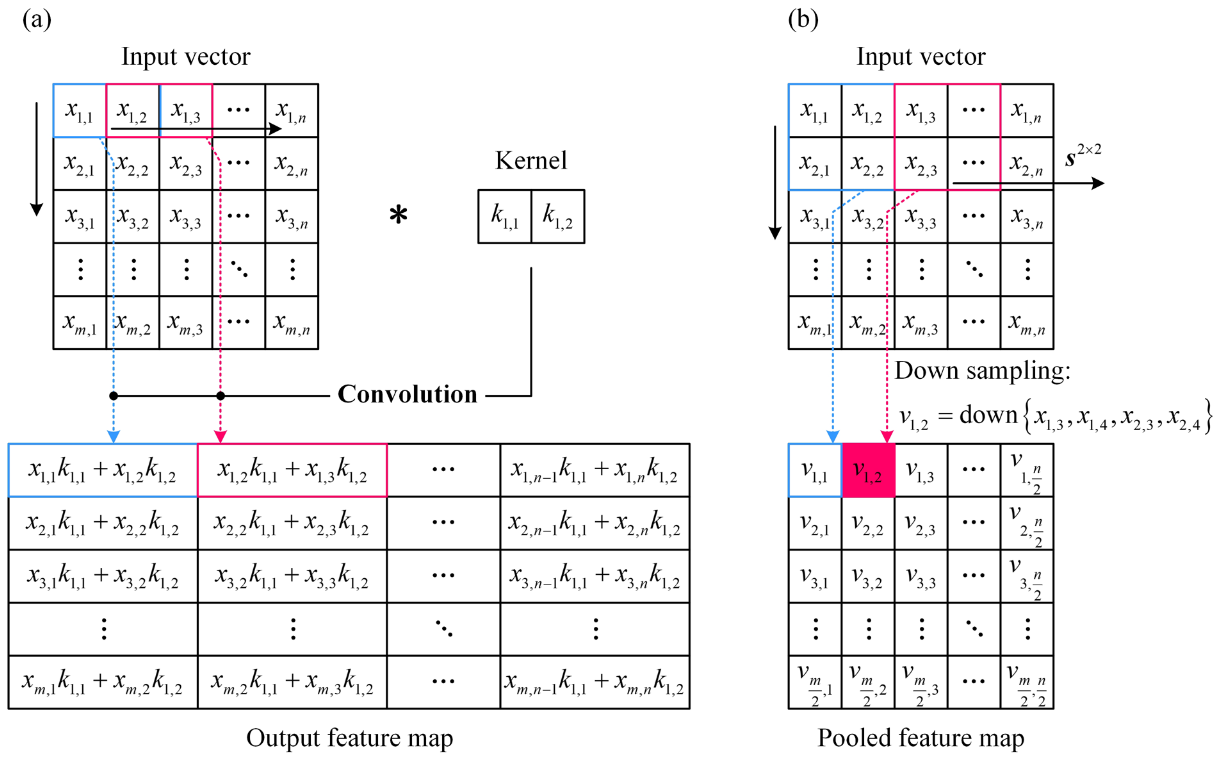


Fig. 13. Convolution and pooling process: (a) Convolution process by using the kernel $\boldsymbol{k}^c \in \Re^{2\times1\times1}$, and (b) pooling process by using the filter $\boldsymbol{s}^{2\times2}$.

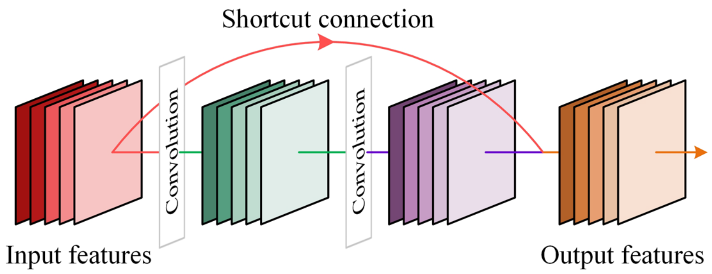


Fig. 14. Architecture of the residual block.

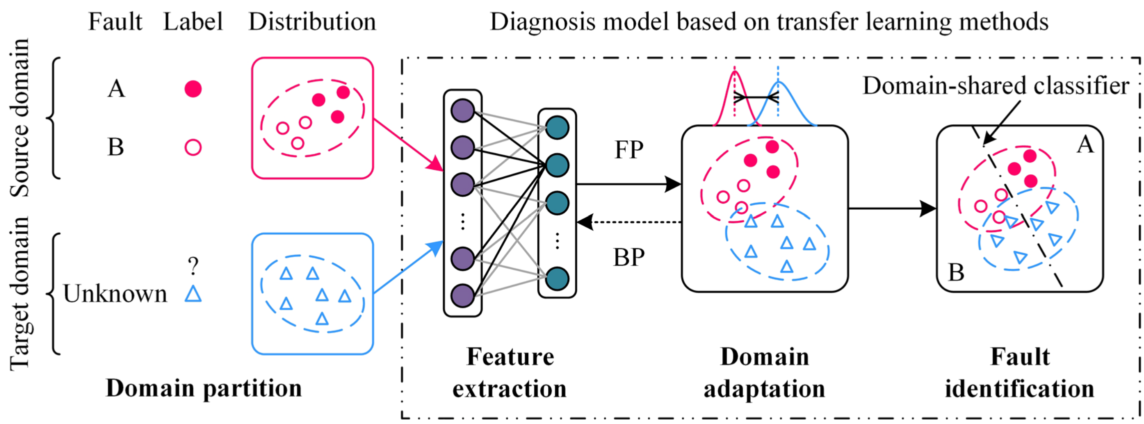


Fig. 15. Steps of feature-based transfer learning approaches [26].

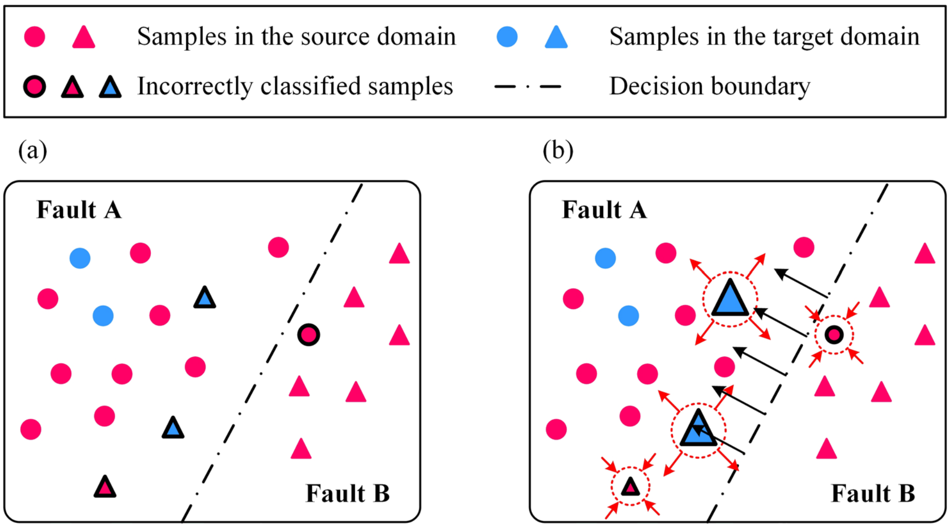


Fig. 16. Illustration of the TrAdaboost algorithm: (a) directly train the diagnosis models with the source and target domain samples, and (b) train the diagnosis models by using TrAdaboost algorithm.

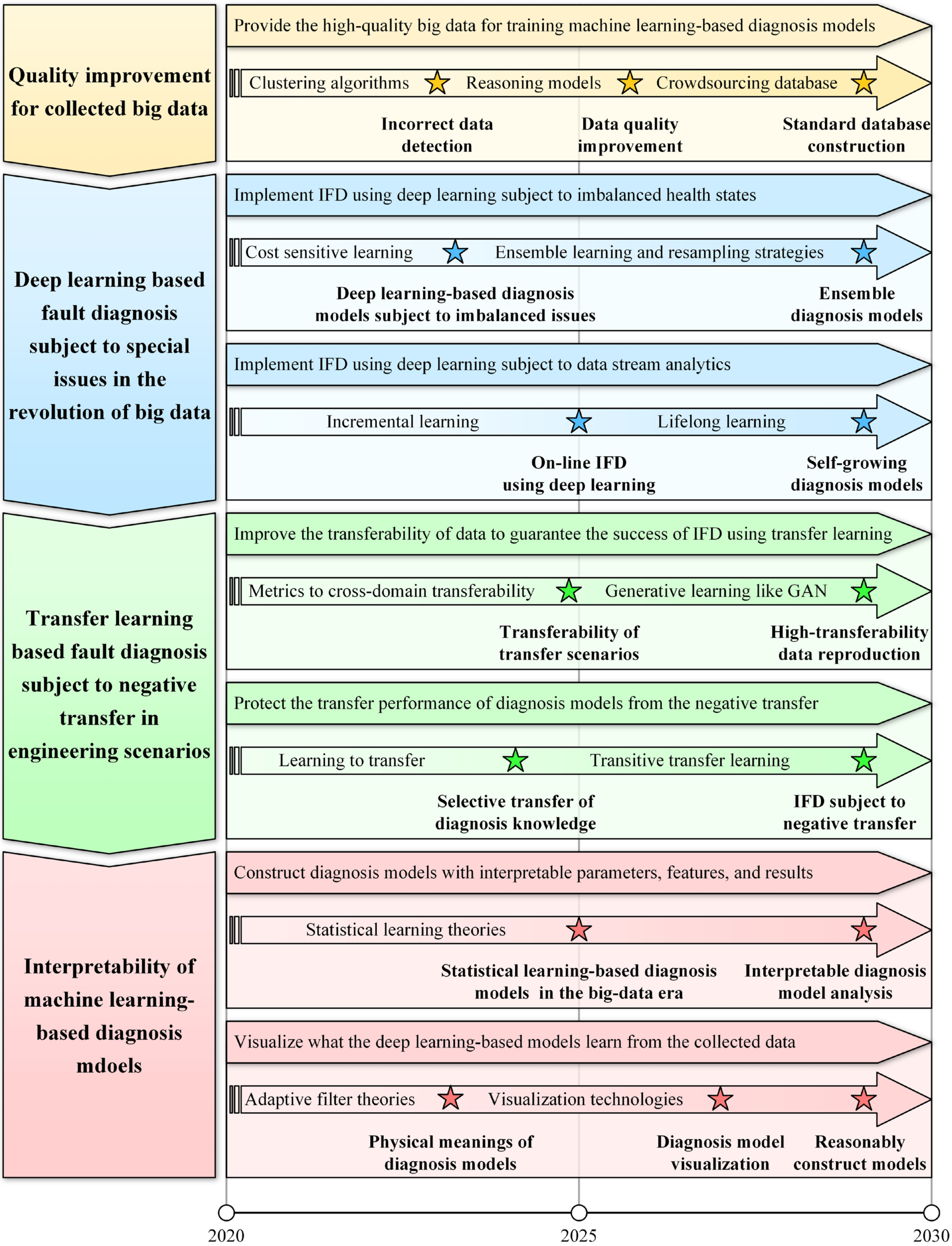


Fig. 17. Roadmap of applications of machine learning to machine fault diagnosis.