Challenges and Opportunities of Deep Learning Models for Machinery Fault Detection and Diagnosis: A Review

SYAHRIL RAMADHAN SAUFI[1,2], **ZAIR ASRAR BIN AHMAD**[1,2], **MOHD SALMAN LEONG**[1,2], **AND MENG HEE LIM**[2]

[1] School of Mechanical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Johor Bahru 81310, Malaysia
[2] Institute of Noise and Vibration, Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia

Corresponding author: Syahril Ramadhan Saufi (msyahramadhan41@gmail.com)

ABSTRACT In the age of industry 4.0, deep learning has attracted increasing interest for various research applications. In recent years, deep learning models have been extensively implemented in machinery fault detection and diagnosis (FDD) systems. The deep architecture's automated feature learning process offers great potential to solve problems with traditional fault detection and diagnosis (TFDD) systems. TFDD relies on manual feature selection, which requires prior knowledge of the data and is time intensive. However, the high performance of deep learning comes with challenges and costs. This paper presents a review of deep learning challenges related to machinery fault detection and diagnosis systems. The potential for future work on deep learning implementation in FDD systems is briefly discussed.

INDEX TERMS Deep learning, fault detection and diagnosis, current challenges, future developments.

I. INTRODUCTION

Safety and reliability are key factors in industrial operations. Rotating machinery is a vital component in many industries, and it is prone to failure due to harsh working conditions and long operational times [1], [2]. Examples of rotating machinery components including gears [3], pumps [4], bearings [5], shafts [6], blades [7], motors [8] and engines [9]. Failures in rotating machinery should be detected as early as possible to prevent critical damage [10] and sudden halt of machine operation. Failures may cause delays in operations and, consequently, tremendous economic loss [11]. For example, petrochemical industries lose around 20 billion dollars per year due to faults in their machine components [12]. According to a report by Duan *et al.* maintenance accounts for more than 60% of the total cost of aircraft engine components [13]. In the worst case, a machinery component failure may lead to loss of human life. Elasha *et al.* discussed a case

in which the failure of planet gear in an aircraft caused an accident that cost 16 people their lives [14].

Fault detection and diagnosis (FDD) is crucial to preventing unexpected breakdowns of machinery and ensuring production efficiency and operational safety. Fault detection and diagnosis systems are categorised into model-based, data/signal-based and knowledge-based approaches [15], [16]. In modern industry, signal-based or data-driven approaches have attracted more attention from researchers [17] because these approaches provide high diagnostic accuracy and do not require empirical estimation of physical parameters [18]. Data-driven approaches have two drawbacks; they depend on good data preparation, and both training and testing samples should be drawn from the same data distribution [19]. It is important to remedy these deficiencies to increase the efficiency of data-driven approaches by implementing deep learning models.

With the rapid development of computer system, artificial intelligence (AI) methods have been broadly used in many applications to assist human in interpreting data trends.

The associate editor coordinating the review of this article and approving it for publication was Kezhi Li.

Machine learning model	Fault detection (Stage 1)	Fault identification (Stage 2)	Fault size estimation (Stage 3)	Fault growth prediction/ RUL (Stage 4)
Shallow learning model	ELM [53]	RVM & SVM [54] ANN & SVM [55] ANN & C4.5 [56] SVM [57] [58]	ANN [50] SVR [59]	ANN [60] [61] [62] HMM [63]
Deep learning model	CNN [64]	CNN [65] [66] [67] AE [68] DBN [69] RBM [70]	AE [71] CNN [67]	DNN [72] AE [73] [74] RBM [75]

* RVM: Relevance vector machine; ELM: Extreme learning machine; HMM: Hidden Markov models; SVR: Support vector regression; AE: Autoencoder; RBM: restricted Boltzmann machine.
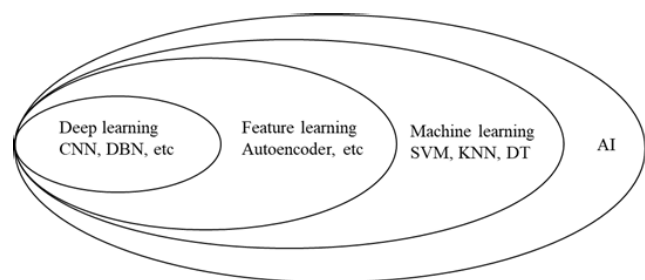

FIGURE 1. Relationship of artificial intelligence methods.

As illustrated in the diagram of AI systems in Fig. 1, deep learning (DL) is a subarea of machine learning. Deep learning has been used in many applications including natural language processing [20], image processing [21], robotics [22], FDD systems [23] and medical applications [24]. The rapid growth of DL implementation in many areas is due to several reasons, including the increase in research on machine learning models and the increase in graphics processing unit (GPU) development [25]. The complex architecture of DL requires a fast computing process to train complex architecture. Deep learning models provide an automated feature learning process that can replace the manual feature selection of traditional fault detection and diagnosis (TFDD).

In recent years, the most common deep learning models that have been extensively used in FDD systems include convolutional neural network (CNN), stacked autoencoder (SAE), restricted Boltzmann machine (RBM), deep belief network (DBN) and deep neural network (DNN) [26]. The SAE, RBM and DNN models are usually called deep neural networks, and are described in [27]–[29]. Every DL model has its own variants, which have been briefly discussed in [30]. Meanwhile, Ravi *et al.* conducted a comparative study of the advantages and disadvantages of each deep learning model [31].

This review is intended to provide a brief discussion of challenges and future development of DL applications in FDD systems for rotating machinery. It does not describe each DL architecture, as such information has been

discussed in detail in [32]–[34]. Meanwhile, the advantages and disadvantages of every DL model used in FDD systems have been discussed in [35].

The remainder of this paper is organised as follows: Section 2 presents a review of the stages of FDD. Section 3 discusses the advantage of DL over shallow learning models. Section 4 describes the challenges of DL in an FDD system. The future development of DL models in FDD systems is discussed in Section 5. Section 6 concludes the paper.

II. FAULT DETECTION AND DIAGNOSIS STAGES

Fault detection and diagnosis can be divided into four stages: fault detection, fault identification, fault severity assessment and fault growth and remaining useful life prediction. Table 1 presents the available reference on each FDD stage for shallow machine learning (SML) and DL models. In FDD systems, data acquisition is important to obtain the signal that reflects the physical condition of the machinery component. In practice, there are several types of monitoring tools including acoustic emissions (AE), vibration, pressure, oil analysis and thermal analysis [2]. Fig. 2 represents the available sensors that can be used in FDD systems, which is based on detection level [36]. Chacon *et al.* stated that AE detects faults earlier than other sensors [37] and can thus provide early detection when faults are present.

A. FAULT DETECTION (STAGE 1)

Fault detection is the examination of faults present in machinery components. A recent study found that most research on fault detection focuses on incipient faults so that the next stage of the FDD process can be performed. Incipient fault detection is important to prevent catastrophic failure [38]. Since approximately 2011 to 2015, incipient fault detection has been focused on using AE due to the high sensitivity of the AE sensor. Ferrando Chacon et al. [39], Hiremath and Reddy [40], and Kilundu *et al.* [41] are among the authors who used AE as a monitoring method for detecting incipient faults. However, in recent years, researchers including
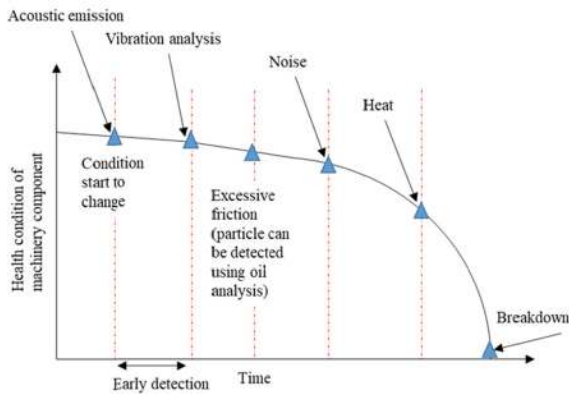
FIGURE 2. Detection level for each monitoring method [36].

FIGURE 3. Performance of deep learning model and shallow learning model [88].

Klaussen and Robbersmyr [38], Seshadrinath *et al.* [42], Zhao and Jia [43] and Lu *et al.* [44] have successfully diagnosed incipient faults using vibration analysis with more advanced signal processing techniques. Wei *et al.* discussed in detail early fault diagnosis in rotating machinery components [45].

B. FAULT IDENTIFICATION (STAGE 2)
Fault identification is the examination of the location and type of faults in machinery components. Fault location identification is usually performed on bearing components; faults can be located in the outer race, inner race or rolling element [46]. Meanwhile, fault type identification is performed in gear component, and types include chipped gear tooth [47], missing gear tooth [47], worn gear [48] and tooth root crack [49]. Most DL models have been extensively used up to this stage 2.

C. FAULT SEVERITY ASSESSMENT (STAGE 3)
Fault severity assessment is the process of determining the size of a fault located in a mechanical component. Kumar *et al.* have proposed a potential method of estimating fault size in bearing components using the integration of signal processing and a machine learning (ML) model [50]. The proposed system requires a good signal processing method to reveal the vibration peak produced when contact occurs between the rolling element and the surface of a fault.

D. FAULT GROWTH AND REMAINING USEFUL LIFE PREDICTION (STAGE 4)
The remaining useful life (RUL) process is the prediction of the life cycle of a machinery component. Meanwhile, the prediction of fault growth is based on the size of the fault after a certain cycle. The output of RUL analysis is the predicted period/time of the component breakdown. A Stage 4 analysis might be useful in an FDD system if the output RUL prediction is expressed in terms of the relationship between time and size of fault growth. To the best of our knowledge, no current research has done this analysis. In common, the features that are more sensitive to fault characteristics will be used. Si et al. [51] conducted a review of statistical approaches for determining remaining useful life. Meanwhile, Nguyen *et al.*
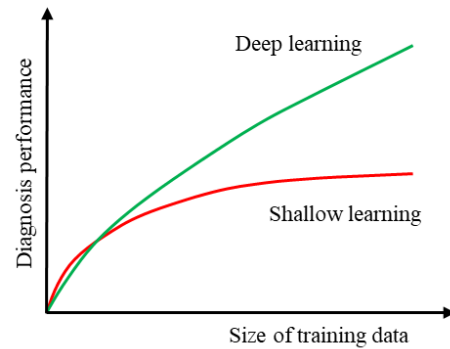
provided a detailed discussion of failure prognostics using a deep learning model [52].

III. A COMPARATIVE STUDY OF DEEP LEARNING AND SHALLOW LEARNING MODELS IN FDD SYSTEMS
In recent decades, FDD systems have typically used the support vector machine (SVM), k-nearest neighbour (KNN), decision tree (DT) and artificial neural network (ANN) models. These types of models are sometimes called shallow machine learning (SML) models [11]. The FDD systems based on SML models are called traditional fault detection and diagnosis (TFDD) systems, which are based on statistical data-driven [76]. For a model to perform well in TFDD systems, the massive number of features should be carefully selected to reduce the computational complexity of the SML model and prevent degradation of classification accuracy [77]. In the SML model, the feature selection process is introduced to reduce irrelevant and redundant features [78]. Hui *et al.* proposed an improved wrapper method to select the best features for an SVM model, which improved its computational efficiency [79].

Moreover, SML models have another deficiency in term of training size. At a certain stage, model performance will become stagnant, as shown in Fig. 3. In contrast, DL models are highly effective when using large training samples [80]. Cao *et al.* conducted a comprehensive study on training data size where the authors proved that the diagnosis performance of a DL model increased as the size of the training data set increased [81].

Traditional fault detection and diagnosis systems require five important steps, as shown in Fig. 4: data acquisition, data processing, feature extraction, feature selection/dimensional reduction and feature classification [82]. Finding suitable methods for each step is crucial and requires a trial-and-error process. Each application of a traditional FDD system uses a unique set of methods. For example, Xu et al. performed the TFDD process on bearing components by using modified distance discriminant technique, fuzzy ARTMAP, correlation measure, Bayesian belief method and a selective ensemble of multiple classifiers [83]. Meanwhile, Cerrada et al. conducted gearbox fault diagnosis using the following combination
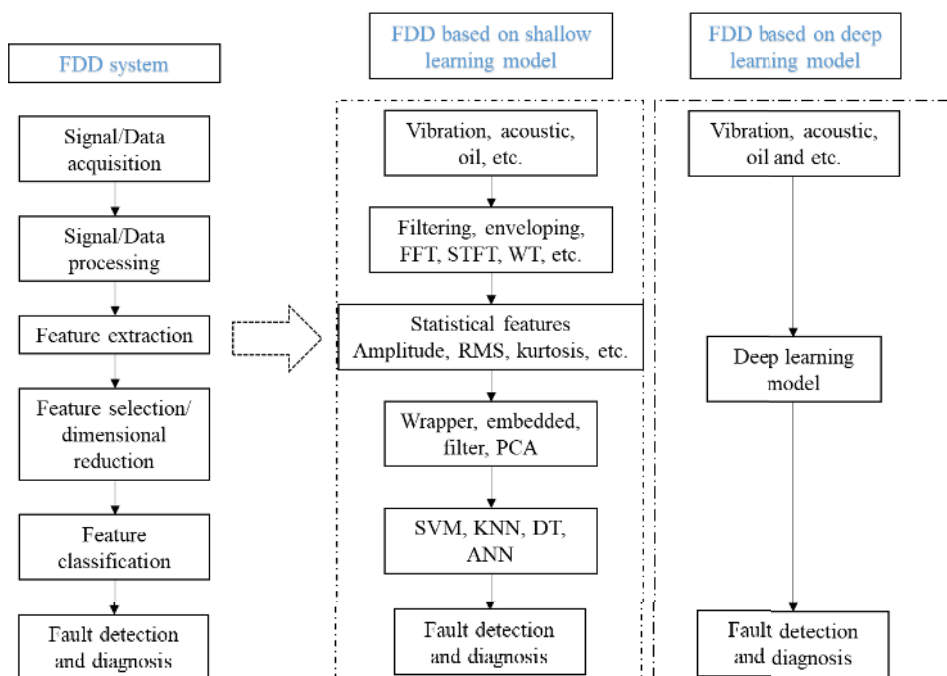
FIGURE 4. The difference between deep learning and shallow learning models in fault detection and diagnosis systems.

methods: correlation-based feature selection, entropy-based feature selection using random forest (RF), the Mamdami-type fuzzy model, hierarchical clustering and membership degree estimation using a fuzzy approach [84].

Deep learning models do not require the combination of methods for each TFDD step; the process is done using multiple hidden layers of deep learning architecture [33]. However, under certain conditions, DL models require a signal/data processing step if the signal/data is too noisy. For example, Liu et al. used variational mode decomposition (VMD) to process the vibration signal obtained from the planetary gearbox and fed the processed signal into a CNN model [85]. Shao et al. used a dual-tree complex wavelet packet to refine the fault characteristics of the vibration signal and fed the data into an adaptive deep belief network (DBN) [86]. In addition, Guan et al. used empirical mode decomposition (EMD) to extract the fault signal that carries more information regarding the physical condition of the machine [87]. The signal reconstructed by EMD is transformed into the frequency domain, and the DBN model is used for the fault diagnosis process. However, the previous research did not compare the performance of the DL model between raw time domain and processed signal. Hence, the performance of the deep leaning model with and without signal processing cannot be discussed.

A. CHALLENGES OF DEEP LEARNING MODELS IN FDD SYSTEMS

The challenges of a DL model are related to its architecture and training process. Even though there is extensive published literature on DL implementations in FDD systems,

they require prior knowledge regarding their architecture. Currently, several programming modules such as MATLAB, R and Python have been used for the development of DL. Due to different styles of code and training processes, diagnosis performance might be different for each programming module. For the last few decades, the deep architecture of the DL model has been difficult to train. The greedy layer-wise pre-training steps proposed by Hinton et al. have reduced the difficulty of the training process [89]. However, several challenges must still be overcome for the DL model to perform effectively in FDD analysis.

B. THE TRAINING PROCESS OF A DEEP LEARNING MODEL

The training process is a crucial part of a DL model. The model is first trained by being given a set of example data, and the parameters (bias and weight) are fine-tuned using a backpropagation algorithm. However, three factors should be taken into consideration before and during the training process: characteristics of input data, size of the dataset and the architecture DL model which are discussed in detail as follows.

1) CHARACTERISTICS OF INPUT DATA

Division of the continuous signal into several segments is a common process in FDD systems. The segmented signal is distributed into training and testing samples for the analysis. There are two types of segmentation processes that can be performed on a continuous signal, as shown in Fig. 5: segmentation without overlapping and segmentation with overlapping. Overlapping is a data enhancement or data
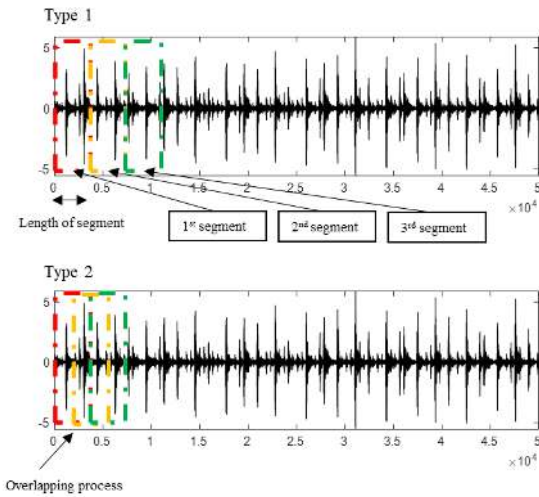
FIGURE 5. Signal segmentation process. Type 1: Non-overlap segmentation; Type 2: Overlap segmentation (Red box: 1st segment, Yellow box: 2nd segment and Green box: 3rd segment).
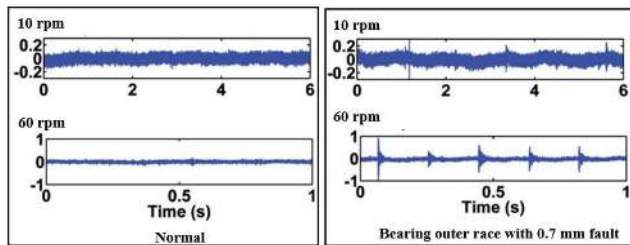

FIGURE 6. The difference in bearing signal characteristics based on shaft speed. (Left: Normal dataset; Right: Fault dataset) [100].

TABLE 2. Segmentation equation.

| Reference	Formula
[95]	$$n = \frac{l - N}{m + 1}$$ n: number of segments; l: signal length; N: number of data points; m: length of shift
[96]	$$n = (N - 1) \times \left(L_{seg} - L_{overlap} \right) + L_{seg}$$ n: number of segments, Loverlap: length of overlapping, Lseg: length of each segment, N: the number of segments
[97]	$x_k(i) = x[i + (h - 1)M]$, for i=1, 2,…,k; h=1,2,…,H h: hth segment sampled, M: number of data point at overlapping area between two segments, k; segment length from the filter size, H: total number of sampled segments.
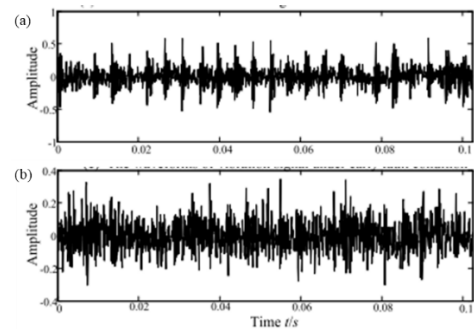
FIGURE 7. The difference of bearing signal characteristics based on fault size; a) Vibration signal with severe fault, b) Vibration signal with incipient fault [45].

augmentation process to increase the number of data samples for better generalisation of DL models [90]. Determining the correct segmentation length is a crucial part of the analysis in order to preserve the important features in each segmented signal. For example, Duong and Kim segmented the continuous signal into lengths of 500 and 1000 data points for bearing fault analysis. Further details regarding signal segmentation can be found in the research of Jing et al. who conducted a comprehensive study on segment length of a continuous vibration signal [91]. Meanwhile, Chen et al. did an analysis of segment length from 64 to 1024 with the increment of 64 segment length using the DBN model. The result showed that diagnosis performance increased with the increase in segment length [92]. Several authors discussed the relation of data point number and segment length by proposing the mathematical equations shown in Table 2. It should be noted that knowledge of the data is required to select the length of segmentation. For example, Liu et al. used 50% overlap on the continuous vibration signal [93]. Ma et al. overlapped the signal by 0.1 seconds out of the total length of 0.5 seconds [94].

Other challenges of the segmentation process arise when dealing with low-speed and incipient fault signals. In those conditions, the interaction of the fault and the rolling element is at a low energy level, which causes the fault peak amplitude

to occur rarely in the signal, which in turn makes the information difficult to extract [98]. An example of a low-speed signal is shown in Fig 6. It can be seen that as the speed increases, the fault peak amplitude becomes more frequent and obvious. Another example of a signal from low-speed operation can be found in the work of Kim et al. [99]. Figs. 7(a) and 7(b) represent the sample signals of large and incipient faults, respectively. In Fig. 7(a), the fault peak amplitude is present in the signal at regular intervals, which provides similar information on each segment signal. However, the peaks from low-speed operation (Fig. 6a) and the incipient fault (Fig. 7b) are not frequent, and each segmented signal might contain different information, which would decrease the quality of the diagnosis analysis.

Each segmented signal can be fed into a DL model in several forms, as listed in Table 3. Input characteristics for DL models are essential for accurate diagnosis. Saufi et al. conducted a study in which each type of input data produced a different diagnosis performance [71]. The authors used a stacked sparse autoencoder (SSAE) model to analyse multiple types of input, proving the versatility of a DL model

TABLE 3. Input types.

List of input types
•   Raw time domain [97]
•   Raw frequency domain [109]
•   Statistical features from time domain [110]
•   Statistical features from frequency domain [91] [110]
•   Statistical features from time-frequency domain [110]
•   Combination of statistical features from all domains [103]
•   Image of time-frequency domain [111]

over a SML model. They concluded that the selection of input types is a crucial part of a better FDD system. In recent years, researchers have tended to use raw time domain and time-frequency image data because they do not require expending much effort on manual data preparation. For example, Wu *et al.* used a 1D time domain with a CNN model and proved that their proposed diagnosis system outperformed a shallow learning model with statistical features [101]. Extraction of statistical features from each signal domain is time intensive and requires expertise to perform the analysis. In addition, it is difficult to determine the features that should be used, because the features vary according to applications [102]. The total number of statistical features from all domains can add up more than 100; Cerrada et al. extracted 359 statistical features from all domains to diagnose gear components [103], and Chen *et al.* extracted 256 statistical features for FDD analysis on gearbox components using the CNN model [104]. The details of the statistical features formula can be found in [105]–[107]. Xiang *et al.* used a teager computed order (TCO) spectrum with a stacking autoencoder to perform fault diagnosis of bearing components under fluctuating speed and variable load [108]. The authors extracted 42 statistical features from three domains and fed them into an SML model. The results showed that the proposed DL model outperformed the SML model. During the experiment, Xiang *et al.* used two bearings with similar geometry and fault characteristics—one for training data and another for testing data—in order to simulate an actual industrial fault diagnosis.

2) SIZE OF DATASET

A large dataset is essential for a DL model, especially for a mechanical signal. However, until now the size of training data has depended on user expertise, and there are no guidelines to ensure that the data is adequate for the training process. Most recent studies used large datasets to train their models. For example, Shao *et al.* used approximately 6,000 to 9,000 training samples to train a CNN model [112]. Guo *et al.* sampled 32,256 data points to achieve 99% diagnosis accuracy [113]. The authors used Pythagorean spatial pyramid pooling CNN to increase the robustness of the model when using data from variable rotating speed conditions. Meanwhile, Jian et al. achieved a satisfactory diagnosis result using 5,000 training samples for bearing components [90]. Zhong *et al.* fed a generative adversarial network 3,000 training samples for each air handling unit fault condition [114]. The current challenge facing DL models is to handle more

TABLE 4. List of input dimensions used on deep learning models.

Reference	Input dimensions	DL model
[117]	256	Enhancement denoising autoencoder
[119]	512	Convolution neural network
[118]	512, 1,024, 2,048, 4,096	Deep autoencoder
[120]	784	Denoising auto-encoder and contractive auto-encoder
[111]	784, 9,216	Deep convolutional neural network
[67]	1,024	Hierarchical adaptive deep convolution neural network
[68]	1,024	Stacked sparse autoencoder-based deep neural network
[70]	1,024	Gaussian restricted Boltzmann machine
[121]	1,024	Deep neural network
[122]	1,024	Improved convolutional deep belief network
[123]	1,024	Fully connected winner-take-all autoencoder
[26]	1,024	Deep convolutional neural network
[124]	1,024	Deep coupled dense convolutional network
[90]	2,048	Adaptive one-dimensional convolution neural network with wide kernel
[125]	3,600	Explainable convolutional neural network
[126]	4,096	Deep belief network
[127]	4,096, 6,144	Deep convolutional neural network
[94]	5,120	Coupling autoencoder
[91]	1,024, 2,048, 4096, 6,144, 8,192, 10,240, 12,288	Convolutional neural network

testing samples than training models. Currently, most DL models have been trained with more than 50% more training samples than testing samples. The capability of DL models to perform FDD with a small number of samples has not been determined. The difficulty of obtaining enough data samples in industrial environments to train a deep learning model has been discussed in [115]. In addition, large datasets are rarely available in industrial applications [116]. Meng et al. addressed the difficulty of obtaining large training samples by using Type 2 data processing, as shown in Fig. 5 [117].

In addition to the number of samples, the size of the input dimension should be considered during the analysis. Common sizes of input features used for DL models are listed in Table 4. According to analysis by Shao *et al.* the training time of a deep learning model significantly increases with an increase in input dimensions [118]. It is worth mentioning that the segmentation signal length sometimes equals to the input dimension as several DL models are directly fed with the segmented signal in raw time domain.

TABLE 5. List of hyperparameters in deep learning models.

Deep learning model	Specific hyperparameter	General hyperparameter
Convolutional neural network	Kernel size [64] Pooling factor Zero padding Number of strides	Epochs Number of hidden nodes Number of hidden layers Activation function
Sparse autoencoder	Sparsity proportion [24] Sparsity regulariser	Backpropagation algorithm Weight decay Learning rate Momentum Dropout fraction

3) THE ARCHITECTURE OF DEEP LEARNING

In general, ML architecture contains parameters and hyperparameters. Bias and weight are common parameters in the ML model that are usually randomly initialized and fine-tuned using the backpropagation algorithm. However, hyperparameters must be pre-set before an ML model can perform a training process. As they are based on deep architecture, DL models have more hyperparameter settings than SML models. Manual selection of hyperparameters is a difficult task, and guessing the values of the optimal hyperparameters is time intensive [128]. Darwish et al. concluded that it is difficult to determine accurate and efficient DL architecture in a reasonable time [129]. A list of specific and general DL hyperparameters is presented in Table 5.

Among the DL hyperparameters, the activation function should be manually set by the user because it is difficult to optimise algorithmically. Activation function selection requires user experience and knowledge. The variety of activation functions available for deep learning models during their analysis of bearing fault diagnosis have been discussed in [109]. Deep learning and SML models generally use sigmoid functions for activation. However, according to Dahl et al. rectified linear units (ReLU) achieved the same training error as sigmoid functions with a quicker training process [130]. Zhao and Kang diagnosed planetary gearbox conditions by using a ReLU activation function in a CNN model [131]. The authors conducted the data collection process for 56 seconds for each fault condition. Long duration of data collection is useful to increase the training and testing datasets. Ma et al. performed a fault diagnosis on bearing components by proposing a concatenated rectified linear unit (CReLU) in a CNN model [132]. The authors concluded that the proposed model is small, light and fast and can achieve satisfactory diagnosis performance. Meanwhile, Shao et al. conducted a comparative study on several types of activation functions in a deep autoencoder model [133]. The authors conducted the analysis using the Case Western Reserve University (CWRU) dataset with 12 bearing conditions where the highest diagnosis accuracy was 97.18%. With 400 input dimensions, the authors were capable of diagnosing the dataset.

Another hyperparameter that should be manually set is the backpropagation algorithm. In recent years, ADAgrad,

RMSprop, stochastic gradient descent (SGD), Nesterov, ADAdelta and Adam are popular backpropagation algorithms for deep learning models [134]. Jian et al. conducted a comparative study of ADAdelta, Adam and RMSprop for their proposed adaptive one-dimensional CNN model [90]. For motor bearing fault diagnosis, the authors achieved higher diagnosis performance with RMSprop. Liu et al. implemented RMSprop on a recurrent neural network-based autoencoder for fault diagnosis of rolling bearings [93]. Gong et al. conducted a comparative study between stochastic gradient descent (SGD), ADAgrad, ADAdelta, RMSProp and Adam using a rotating machinery dataset [135]. The result showed that Adam produced higher accuracy compared to other backpropagation algorithms. However, the algorithms differed only slightly in terms of classification accuracy, training time and test time. Lai et al. performed a comprehensive analysis of the backpropagation algorithm and found a slight difference between RMSProp and Adam in diagnosis performance [136]. Pan et al. diagnosed bearing conditions accurately with an implementation of the Adam algorithm in the CNN model [137]. Tang et al. used Nesterov momentum for a DBN model that outperformed the standard DBN model using a gearbox dataset [138].

New variants of Adam and AMSgrad are proposed by Luo et al. to overcome the poor generalisation of ADAgrad, RMSprop and Adam [139]. The new variants are called ADAbound and AMSbound. If these new backpropagation algorithms provide significant improvements, especially for complex deep architecture, they could present a challenge to the current algorithms. Implementation of these algorithms in DL models for FDD systems is useful to increase their performance.

The hyperparameter that rarely gets attention is the number of hidden layers, even though the hidden layer size is part of the definition of a DL model. According to analysis by Ren et al. a deep network must consist of an input layer and output layer separated by two or more hidden layers [140]. The specific hyperparameters of each deep learning model need to be set in each hidden layer. Thus, selecting the number of hidden layers is a crucial step that affects the selection of other hyperparameters. Wang et al. stated that the proper selection of hidden layers is important to avoid computational complexity [141]. Heo and Lee performed an
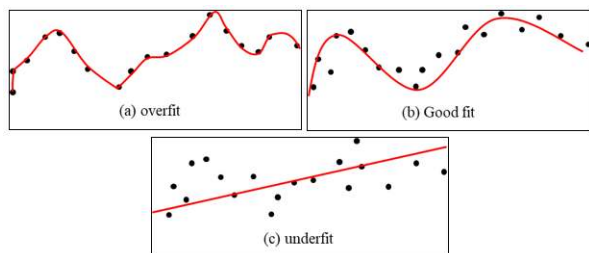
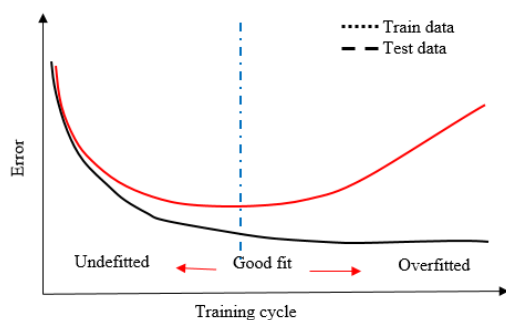FIGURE 8. Training conditions of deep learning models.

FIGURE 9. Relationship of train and test error during training process.

analysis in which the diagnosis performance increased with an increase in the number of hidden layers [142]. However, an analysis by Saufi et al. showed that diagnosis performance varies according to the number of hidden layers [128], which is similar to the result achieved by Sohaib and Kim [68] and Jing et al. [91].

During the training process, the DL model can be in a condition of good fit, underfit or overfit, as illustrated in Fig 8. Fig. 9 represents the relationship between prediction error and training cycle of the DL model. Dong et al. briefly explained the nature of overfit and underfit as they occur in ML models [143]. The ideal condition is a good fit, where the model reaches high training and testing accuracy.

The model will experience an underfit condition when the training and testing accuracy are low because the model is unable to learn the pattern input features due to a smaller number of training data points. According to Saurabh, underfitting occurs when the number of hidden neurons is low compared to the complexity of the input data [144].

In the overfit condition, the model achieves high training accuracy and low testing accuracy. Too many training cycles (epochs) may cause the model to memorise the features instead of generalising the feature patterns. In terms of FDD analysis, the noisy data may contribute to an overfitting problem wherein the model captures the noise along with the important features. Srivastava et al. stated that implementation of dropout and L1/L2 regularisation is capable of reducing the overfitting problem [145]. Shao et al. validated their proposed model performance by presenting training error and classification rates to prove that

the model did not experience an overfitting problem [146]. In addition, the authors indicated that overfitting problems occur when the number of hidden neurons increases. Similarly, Sheela and Deepa stated that random selection of the number of hidden neurons may cause a ML model to underfit or overfit [147].

As mentioned in the previous section, several authors applied signal processing methods to noisy input data to remove unwanted noise. The training conditions are closely related to DL hyperparameters. Widodo et al. emphasised the importance of hyperparameter selection to reduce overfitting problems [54]. He and He mentioned deep learning based signal processing has been developed for bearing fault diagnosis [97].

C. IMPLEMENTATION OF REAL MACHINERY SYSTEMS

Most of the DL applications available in public articles use experimental datasets for FDD analysis. Few studies have been found that use a real machinery system. Experimental and real machinery datasets differ in quality. An experimental dataset is collected under controlled conditions with a less complex system and less environment disturbance. A real machinery system, on the other hand, is a complex structure, and the collected data contain the information from unrelated components of interest. Xu et al. have questioned how high-quality raw data can be obtained in a cost-effective way [148]. In FDD systems, data quality depends on the types of sensors, data acquisition setup, the surrounding environment and the duration of the data collection. Zhang et al. emphasised that sensor problem result in poor data quality and they conducted a thorough review of how to handle low quality data using DL models [149]. High data quality is essential to ensure satisfactory performance of a DL model.

Wang et al. stated that most studies they reviewed on wind turbine gearboxes aimed to refine the fault-related signal more clearly by filtering the background noise and the unrelated vibration components [150]. Xu et al. suggested that under practical conditions, the data cleaning procedure is important to ensure good performance of a deep learning model [151]. The accuracy of a deep learning model is reduced if the input signal/data contains noise [18]. Teng et al. and Saidi et al. performed a fault diagnosis on real wind turbine gearboxes [61], [152] and it is noted that satisfactory diagnosis performance is difficult to achieve in real applications. Sadoughi et al. diagnosed bearing conditions in hay balers, and their proposed model achieved diagnosis performance of less than 95% [66]. Shao et al. diagnosed locomotive bearing conditions using a deep autoencoder and achieved diagnosis performance of up to 87.8% [118]. However, the performance reached 94.05% for the experimental dataset [118]. Xie et al. diagnosed bearing conditions on an experimental test rig and achieved diagnosis performance of between 98% and 100% for every bearing condition [153]. Yang et al. achieved approximately 99.57% to 100% accuracy when diagnosing bearing conditions from the CWRU dataset.

Liu *et al.* diagnosed a natural gas reciprocating compressor component using advanced signal processing consisting of linear mode decomposition and a stack denoising autoencoder [154]. The author achieved diagnosis performance of 92.72% under the condition of low signal-to-noise ratio. Peng *et al.* diagnosed wheelset bearings in a high speed train using a deeper 1D convolutional neural network; the model achieved between 89.7% and 99.9% diagnosis performance with different noise conditions in the dataset [155]. The authors used data augmentation to increase the sample size to 329,752 samples. Fault diagnosis on real machinery systems will be a great challenge in future years as current systems become larger and more complex with a high potential to break down on any given day [156].

Moreover, every ML model, including deep learning, requires labelled data, which are usually difficult to obtain in real industries [157]. In order to solve this problem, Li et al. proposed data augmentation methods that can create an artificial additional sample during the deep learning training process [157]. The authors demonstrated five types of data augmentation: gaussian noise, masking noise, amplitude, time stretching and signal translation. Han *et al.* used the adversarial learning framework with the CNN model in order to deal with small labelled samples [127]. They analysed training datasets of different sizes and found that the diagnosis performance increased as the number of training data samples increased.

Furthermore, imbalanced datasets commonly occur in industry environments, where the datasets for normal and fault conditions do not have similar sample sizes [158]. Chen et al. proposed graph-based rebalance semi-supervised learning (GRSSL) in order to deal with an imbalanced, unlabelled dataset [159]. The authors used the CWRU dataset to validate their proposed model. Zhang *et al.* briefly discussed the conditions of imbalanced datasets, on which they used a synthetic oversampling technique called weighted minority oversampling. The authors used a deep learning model to effectively learn the input features for accurate fault diagnosis on an intelligent maintenance system (IMS) bearing dataset [160]. Chen *et al.* analysed imbalanced SCADA data for fault detection in wind turbines using a CNN model [161]. The proposed model was able to separate the data from different classes effectively using principal component analysis. Meanwhile, Jia *et al.* diagnosed bearing conditions with imbalanced fault classification using a normalized CNN model with a neuron activation maximization (NAM) algorithm [162]. According to the authors' result, diagnosis performance decreases when the percentage of dataset imbalance increases.

IV. DEVELOPMENT AND MODIFICATION OF DEEP LEARNING MODELS IN FDD SYSTEMS

The versatility of deep learning models allows for better FDD systems compared to SML models [32]. In this section, several non-related articles on machinery fault diagnosis have been used for recommendation and reference
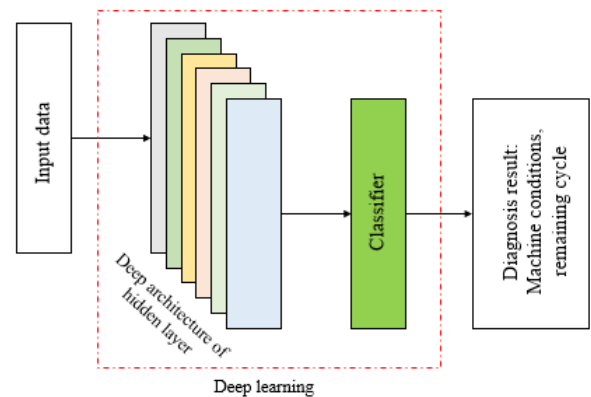

FIGURE 10. Architecture of a DL model for FDD systems.

in a future study. In this section, potential avenues for development and modifications of DL models in FDD systems are discussed.

A. COMBINATION OF DEEP ARCHITECTURE WITH SML MODEL

Deep learning contains a number of hidden layers that are used for feature extraction. The last layer of a DL model is the classification layer, which is used to classify the extracted features. The architecture of a DL model is illustrated in Fig. 10. Softmax regression is typically used as a classification layer in deep learning models [163], [164], but KNN, DT, SVM, ELM and ANN can also be used. The integration of a sparse autoencoder and an SVM model was proposed by Ju et al., who used various UCI datasets to test their proposed model [165]. SVM is a well-known model that can handle data with low sample sizes and high-dimensional features [166]. This integration may provide a great opportunity for dealing with small data sample sizes in the machinery component. Meanwhile, Xu *et al.* combined CNN with random forest (RF) ensemble learning [151]. They used the CNN model to extract the low-level features and input the features into the RF for classification. Li *et al.* diagnosed gearbox conditions using two RBM layers for the feature extraction process and RF as an output classifier [167]. The authors achieved a slight difference in diagnosis performance by changing the output classifier with SVM and KNN model. Li *et al.* combined DBN and random forest to classify spacecraft electrical signals; this method performed well in term of classification accuracy and computational efficiency [168]. The data used by the authors contained 1,000 features, and the data is distributed with 56% of training data and 44% of testing data. This approach is capable of reducing the computational load and enhancing the classifier's performance. Moreover, Monteiro *et al.* used SVM in order to improve the decision level of a CNN model, which produced a great improvement in training time and diagnosis accuracy [169]. Cheng *et al.* successfully diagnosed wind turbine gearboxes based on the current signal using autoencoder and SVM models, achieving an overall performance of 89.3% [170].

During the analysis, the authors used the Hilbert transform for amplitude demodulation. Haidong *et al.* reached approximately 95% diagnosis accuracy for 12 classes of bearing conditions from the CWRU dataset by combining a deep autoencoder with extreme learning machine (ELM) [5]. Similarly, Mao *et al.* proposed a combination of autoencoder and ELM in order to achieve satisfactory diagnosis performance with acceptable processing time [171]. The authors discussed 71 statistical features from time domain, bispectrum, EMD, and the wavelet packet decomposition, where each input data is sensitive to certain types of bearing conditions. Li *et al.* proposed a deep ELM model in which the sparsity and neighbourhood are integrated into the deep model and used to analyse a bearing fault dataset. The proposed model achieved 97% accuracy for experimental data [172]. The experiment was conducted on a Spectra Quest rotor experimental platform and the data was collected at a 12 kHz sampling rate. The proposed diagnosis system outperformed other DL models including a stacked denoising autoencoder and a CNN. Meanwhile, Wang *et al.* used the CNN model to learn the data features automatically from the CWRU dataset raw signal [173]. The features were fed into a hidden Markov model in order to classify the signals based on bearing conditions. The authors achieved approximately 98% diagnosis accuracy for 12 classes of bearing conditions, which is slightly higher compared to Haidong *et al.*'s model.

Instead of being used as a classifier in a DL model, the SML model can be used for data processing along with deep architecture. For example, Hu *et al.* used a multi-grained scanning forest ensemble at each DBM hidden layer in order to perform accurate diagnosis on industrial machinery systems [174]. Hu *et al.* conducted a comprehensive study of the relationship between the number of classes and the diagnosis accuracy, noting that performance decreases as number of classes increases. In addition, Liu *et al.* used RNN to analyse multiple time sequence data and a denoising autoencoder to analyse the CWRU dataset [93]. The authors added noise to the dataset to create signal-to-noise ratios (SNRs) ranging from 1% to 10%. The performance of the proposed model increased from 96.98% to 99.75% as the SNR increased, and the analysis indicated that the noise contained in the signal might reduce the deep learning model performance. Li et al. proposed using deep stacking least squares SVM (LS-SVM) with a one-against-all strategy to extract features from the CWRU bearing signal [175]. They found that the classification accuracy increased along with the increase in stacking layer modules.

B. HYPERPARAMETER OPTIMISATION

Hyperparameter selection is an important process, as the performance of a deep learning model is highly affected by the hyperparameters [176]. Optimisation of hyperparameters for DL can be automated using several methods, as listed in Table 6. Optimisation algorithms are computationally intensive, with the computational load depending on the hyperparameter size. Even with automated hyperparameter

TABLE 6. Potential algorithms for deep learning hyperparameter selection.

Optimisation algorithm
• Random method [183]
• Grid search method [26]
• Bayesian optimisation [184]
• Metaheuristic optimisation [129], [185]
• Optimisation method proposed by Mirjalili [186]–[188]

optimisation, human intervention is still needed to decide on the search space for the hyperparameter boundaries [130]. Until now, the optimisation process was limited to numerical hyperparameters.

Wahab *et al.* conducted a detailed analysis of metaheuristics optimisation algorithms including particle swarm optimisation, differential evolution (DE), genetic algorithm (GA), artificial ant colony, artificial bee colony, glowworm search optimisation and cuckoo search optimisation [177]. They used a benchmark function to analyse these algorithms. Meanwhile, nature-inspired algorithms have been briefly reviewed by Fizter Jr. *et al.* and they listed all available optimisation algorithms along with references [178]. Beheshti and Mariyam conducted a comprehensive review of metaheuristic algorithms [179]. The detailed analysis by these authors provides a great deal of information on the implementation of optimisation algorithms in DL models for FDD systems.

New optimisation methods have been proposed by Mirjalili including grey wolf optimiser, salp swarm optimisation, grasshopper optimisation, ant lion optimisation and whale optimisation [180], [181]. It is worth mentioning that the methods proposed by Mirjalili are more user-friendly because they require fewer parameter settings for the algorithms, which can be easily used for DL hyperparameter selection.

Several studies have demonstrated the potential of automated hyperparameter selection in DL models for FDD systems. For example, Saufi et al. compared the performance of random search, grid search, Bayesian optimisation, GA and DE and proved that all methods have the ability to optimise DL hyperparameters [128]. However, DE produced slightly higher performance compared to other optimisation algorithms. Haidong *et al.* used an artificial fish swarm algorithm and particle swarm optimisation to optimise deep autoencoder and deep belief network hyperparameters, respectively [118], [182].

C. IMPLEMENTATION OF REMAINING USEFUL LIFE AND PROGNOSIS ANALYSIS

Most prognoses and predictions of RUL rely heavily on statistical features. Ren *et al.* extracted three time domain features and one frequency domain feature to estimate the remaining useful life of bearing components [140]. Twenty features related to pressure, temperature, fan and core speed, fuel flow and coolant bleed have been used to predict the possible remaining operational life of aircraft gas turbines [74].

Zhang *et al.* used a deep learning-based long short-term memory (LSTM) approach to estimate the RUL of aircraft engines. Their model reached the lowest prediction root mean square error (RMSE) at 16.12 [189]. Haidong *et al.* used a continuous deep belief network (DBN) with locally linear embedding (LLE) to estimate the remaining useful life of bearing components [190]. The authors used the IMS dataset to test their proposed RUL system, and the model achieved the lowest prediction error of 0.7801%. The authors proposed a comprehensive index feature where the performance of DBN using proposed feature outperformed other six statistical features. Ren *et al.* fed multiple features from time, frequency and time-frequency domains into a deep autoencoder and deep neural network, and the model produced a mean square error (MSE) of 0.2 [191]. Similarly, Wang *et al.* used features from time, frequency and time-frequency domains to predict the useful life of bearing components using a two-dimensional CNN model [192]. The proposed CNN model outperformed other models in terms of prediction time and error rate. Meanwhile, Li et al. predicted the degradation process of a turbofan engine using deep CNN with raw signal datasets [193]. According to the analysis, the prediction error rate decreases with an increase in convolutional layers. Zhu *et al.* diagnosed bearing component using a wavelet transformation representation and multiscale CNN [194]. Li *et al.* carried out a prediction analysis using an image of the short-time Fourier transform and a CNN model and proved that the RUL process can be used with time-frequency images as input data [195]. Deutsch and He predicted the life of bearings and gear components using a DBN-feedforward neural network model [196]. Ellefsen et al. proposed a combination of RBM and LSTM for detecting turbofan engine degradation [197]. The authors used a GA to optimise random hyperparameters of the proposed model. Al-Dulaimi *et al.* proposed a hybrid model of CNN and LSTM in order to estimate the remaining useful life of turbofan engines [198]. In addition, Zhang et al. used a deep recurrent neural network to analyse a similar dataset of turbofan engines [199]. The authors used temperature, pressure and speed sensors during the analysis, which demonstrated that under real-world conditions, multiple types of sensors should be taken into consideration.

Root mean square error (RMSE), mean absolute percentage error (MAPE) and mean absolute error (MAE) have always been used to examine the performance of deep learning models for RUL analysis and are generally used together during the analysis [52]. In addition, it was proved that RULs models can use time-frequency images and raw time domain signals. However, use of these types of input data for RUL is still at an early phase and requires more study in future years.

D. INTEGRATION OF MULTIPLE SENSOR TYPES

A great deal of information is required to examine the condition of large machinery such as a wind turbine component. Vibration analysis is the usual method used in FDD systems and has been proven to be effective as reported by
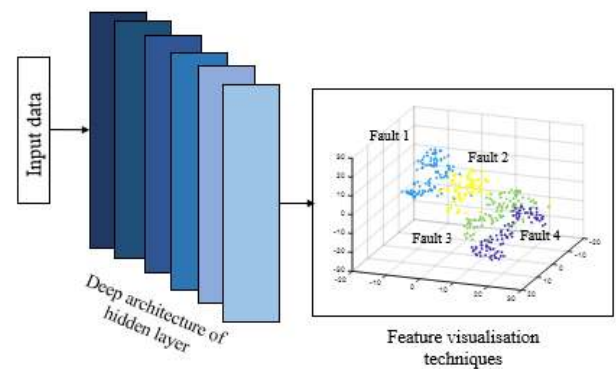

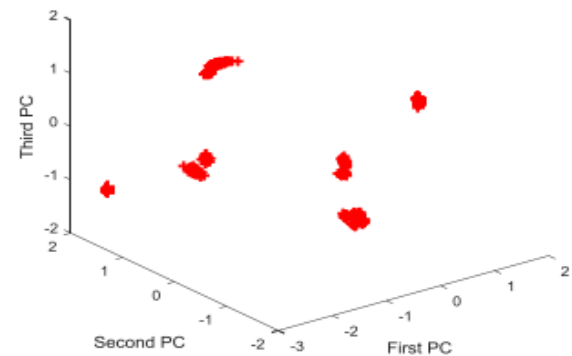**FIGURE 11.** Fault visualisation method.

FIGURE 12. First three principal components of the features from CWRU bearing data [207].

Hui *et al.* [200]. Most of the reference articles in this review paper used vibration analysis to test the proposed models. However, under certain circumstances, a single type of sensor may not provide enough information regarding the condition of the machine. According to Sarkar *et al.* the time series signal obtained from a single sensor may not carry sufficient information regarding the evolving fault [201]. They also successfully diagnosed aircraft gas turbine engines based on combining information from multiple sensors such as pressure, temperature and speed. Wang et al. stated that there are 41 parameters from various sensors that can be used for monitoring wind turbine conditions [202]. Li *et al.* performed a data fusion process by combining vibration and acoustic emission (AE) data for gearbox fault diagnosis [167]. They found that the combined data (AE and vibration) produced higher diagnosis performance compared to the analysis without data fusion. In addition, Jing *et al.* used multiple types of sensor (accelerometer, microphone, current sensor, optical encoder) with a deep convolutional neural network to diagnose planetary gearbox condition [203]. Jing *et al.* conducted a study in which they combined all of the sensor data as an input data for their proposed DL model. Multi-type sensor data fusion is among the future opportunities for development, since DL models have the ability to analyse high-dimensional features. Numerous sensors of different sizes may contribute to a multidimensional data space [204], which would give DL a great advantage over SML.
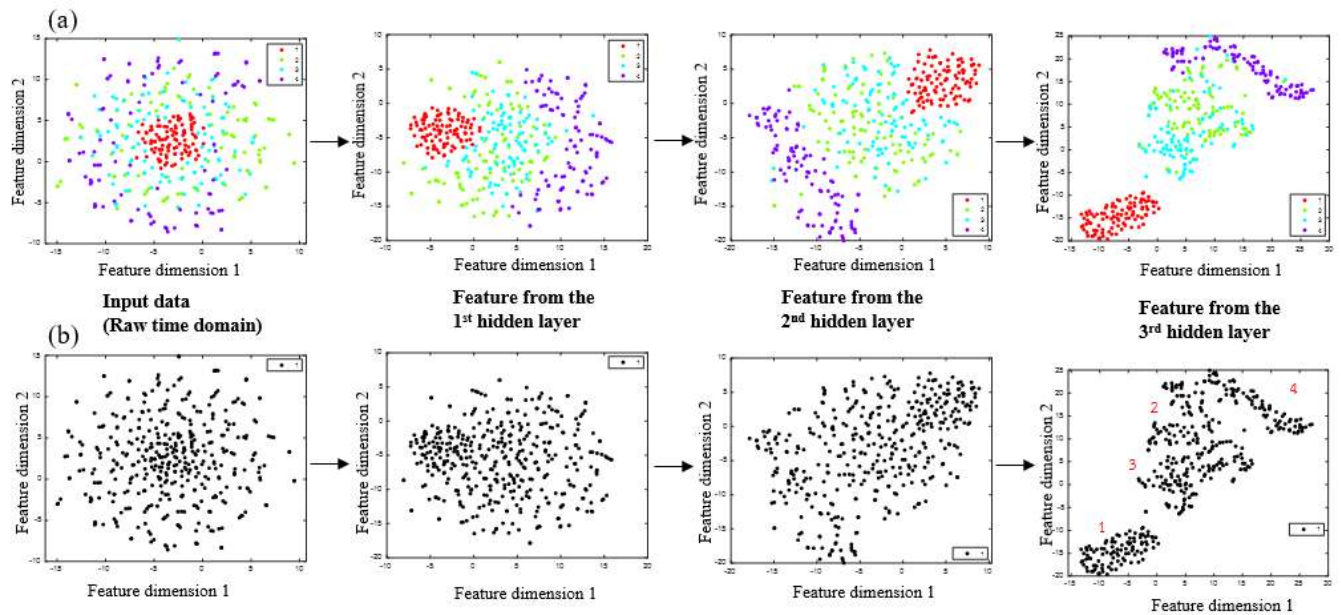
*Red: Normal conditions (Class 1); Green: Outer race fault (Class 2); Blue: Ball fault (Class 3); Purple: Inner race fault (Class 4).

FIGURE 13. Feature learning and t-SNE analysis: a) Analysis with known target data; b) Analysis with unknown target data.

E. FAULT VISUALISATION ANALYSIS

A DL model contains two processes: feature learning, in which the model extracts the features automatically from the input data; and feature classification using a fully connected layer [81]. Integration of feature learning with visual feature representation may provide a fault visualisation method, as shown in Fig. 11. There are several types of feature visualisation including t-distributed stochastic neighbour embedding (t-SNE), principal component analysis (PCA), locally-linear embedding (LLE), linear discriminant analysis, locality preserving projection and isometric projection [205]. The most common method used to analyse machinery data feature is a t-SNE model. The T-SNE model was developed by Van Der Maaten and Hinton [206]. However, there is no proof of the effectiveness of t-SNE over other methods, which provides a great opportunity for future study.

Yang *et al.* used features from the frequency domain and reduced the dimensions of the extracted features before inputting them into a DNN model for bearing fault diagnosis [207]. Then, the features processed by DNN were visualised using the PCA model. The authors proved that feature clustering can be automated. An example of their results is shown in Fig. 12, where a clear boundary can be noticed for each class condition.

Moreover, in this paper, a simple analysis based on time-frequency image input data and a stacked sparse autoencoder (SSAE), with a t-SNE model used to analyse the dataset from CWRU. SSAE and t-SNE are both based on default architecture in MATLAB module. The raw time domain signal, with segment length of 800 without an overlapping process, is transformed into a time-frequency image and it is directly fed to the SSAE model.

Fig. 13(a) consists of the t-SNE visualisation from each SSAE hidden layer with the given target values. The visualisation contains four classes, and the feature representation improves at each SSAE hidden layer. Meanwhile, Fig 13(b) represents the analysis without the target value, and the feature is visualised in a single colour. One limitation of current diagnosis systems is that they require target values for better analysis. However, in Fig 13(b), it can be seen that even without a target value, the bearing conditions can be classified into several groups. Classes 2 and 3 might be misinterpreted during the analysis since both overlap. The Fig. 13(b) visualisation demonstrates that fault visualisation can provide initial information during the FDD analysis. However, this analysis is insufficient, since the SSAE and t-SNE models are run using the default setup. If all aforementioned challenges have been met, fault visualisation methods can be improved, and clear separation between each fault class can be produced.

F. THE AVAILABLE ONLINE MACHINERY DATASETS

The vast majority of publications regarding FDD systems use the CWRU dataset. However, there are several datasets that can be obtained from online databases. Table 7 lists these datasets and the advantages and disadvantages of each. The CWRU dataset can be used as a starting point to determine the performance of the proposed model. Unlike other datasets, the CWRU database consists of four operating conditions with three fault severities. A DL model can be tested with multi-fault conditions: up to 12 bearing conditions for each operating condition. Mao *et al.* achieved satisfactory diagnosis performance and a clear separation between each fault condition using feature visualisation [208]. Zhang *et al.* provided a comprehensive review of the performance of every

TABLE 7. List of available online machinery datasets.

Dataset	Description	Ref.	Advantages	Disadvantages
Bearing	Case Western Reserve University	[213]	Contains up to 12 bearing fault conditions with 4 speed conditions.	Several datasets do not contain long enough data samples.
	Prognostic Data Repository: Centre for Intelligent Maintenance Systems (IMS)	[214]	Available for fault detection, fault identification and RUL analysis.	-
	Machinery Fault Database (MaFaulDa)	[215]	Consists of bearing faults, misalignment and imbalance faults with four load conditions.	The selection of proper sensor is important, because the data was collected using two triaxial accelerometer and a microphone.
	Paderborn University	[216]	Five different type of fault have been introduced at bearing outer race and inner race. The data was collected using vibration and stator current [209].	-
	IEEE PHM 2012 data challenge	[217]	The experimental procedure provides a clear understanding of how to conduct an accelerated degradation test. The data is collected using two sensors: vibration and temperature [218].	Because the data is used for competition, the data conditions are quite difficult to analyse.
Gearbox	High-speed turbine	[219]	The data was collected from real gearbox systems.	Limited class sample with only two gear conditions.
	IEEE PHM 2009 data challenge	[220]	Contains various type of spur and helical gear health condition.	The data has been used for competition, where the data class should be manually determined based on signal characteristics.
Blade	UNSW turbine test rig	[221]	Consists of normal blade, normal blade with air jet and fault blade conditions.	Limited data samples with only one fault simulation.
Turbofan engine	Prognostics CoE at NASA Ames (PHM 2008)	[222]	The experiment was run with noise introduced to simulate real-world conditions [223].	-

deep learning model tested on the CWRU dataset [209]. On the other hand, the data provided by machinery fault database (MaFaulDa) is useful for the next analysis since there is more than one types of sensor. The data is collected from microphone and vibration sensors. Meanwhile, IMS bearing datasets can be used for fault identification and RUL analysis, since the data has been run for endurance test. Unlike in other bearing datasets, the fault in IMS bearing components is naturally grown during the run to test failure. Saufi *et al.* used the CWRU, IMS and MaFaulDa datasets to examine the performance of their proposed model [128]. The authors achieved more than 98% diagnosis accuracy for all datasets. Guo *et al.* performed a transfer learning process using CWRU as the source domain and IMS dataset as the target domain [210]. The RUL analysis can also be done using the IEEE PHM 2012 dataset, the experiment for which was run using the PRONOSTIA platform. The datasets from Paderborn University consist of three types of bearing conditions—normal, outer race faults and inner race faults—with artificial and realistic damage conditions. Using Paderborn datasets, DL models can be tested with several types of bearing damage from stator current signal that cannot

be obtained from other bearing datasets. Zhu *et al.* achieved 97.15% and 82.05% average accuracy using the CWRU and Paderborn university datasets, respectively [211]. The authors determined that the time-frequency analysis provides higher diagnosis accuracy compared to frequency spectrum and time domain segment data. However, Chen *et al.* achieved 94.5% for Paderborn dataset using a deep inception net with an atrous convolution neural network [115]. It should be noted that the MaFaulDa and Paderborn datasets receive little attention in the development of DL models.

There are two datasets for gear fault conditions available online: high-speed turbine and IEEE PHM data challenge. The high-speed turbine dataset includes normal and fault conditions of gear sets. The experiment that generated this dataset was conducted using the actual gearbox of a wind turbine system. Furthermore, the IEEE PHM 2009 datasets provide more fault conditions in the gearbox system. The experimental setup and dataset details have been discussed in [101] and [212]. The turbine blade dataset from UNSW contains three types of conditions: normal blade, normal blade with an air jet and blade fault. The turbine experimental rig contains a 19-blade arrangement. Most of the gear and blade datasets

are not extensively used for DL models which provide a good opportunity for further development. Finally, all of the listed data in Table 7 was collected at high rotation speeds; DL models cannot be examined with low-speed conditions.

V. CONCLUSION

Fault detection and diagnosis (FDD) for rotating machinery is important in order to increase operational time and prevent unexpected breakdowns. With continued rapid advances in computer technology, DL models will continue to be powerful and attractive for use in FDD systems. However, the performance of DL brings with it several challenges. Therefore, in this paper, the challenges and development of DL implementations in FDD systems for rotating machinery have been reviewed.

REFERENCES

[1] M. Zhao, J. Lin, X. Xu, and X. Li, "Multi-fault detection of rolling element bearings under harsh working condition using IMF-based adaptive envelope order analysis," *Sensors*, vol. 14, no. 1, pp. 20320–20346, 2014.

[2] X. Chen, S. Wang, B. Qiao, and Q. Chen, "Basic research on machinery fault diagnostics: Past, present, and future trends," *Front. Mech. Eng.*, vol. 13, no. 2, pp. 264–291, Jun. 2018.

[3] Y. Qu, D. He, J. Yoon, B. Van Hecke, E. Bechhoefer, and J. Zhu, "Gearbox tooth cut fault diagnostics using acoustic emission and vibration sensors—A comparative study," *Sensors*, vol. 14, no. 1, pp. 1372–1393, 2014.

[4] N. R. Sakthivel, V. Sugumaran, and S. Babudevasenapati, "Vibration based fault diagnosis of monoblock centrifugal pump using decision tree," *Expert Syst. Appl.*, vol. 37, no. 6, pp. 4040–4049, Jun. 2010.

[5] S. Haidong, J. Hongkai, L. Xingqiu, and W. Shuaipeng, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine," *Knowl.-Based Syst.*, vol. 140, pp. 1–14, Jan. 2018.

[6] A. R. Mohanty and S. Fatima, "Shaft misalignment detection by thermal imaging of support bearings," *IFAC-PapersOnLine*, vol. 28, no. 21, pp. 554–559, 2015.

[7] J. Kubiak, G. Urquiza, J. A. Rodriguez, G. González, I. Rosales, G. Castillo, and J. Nebradt, "Failure analysis of the 150 MW gas turbine blades," *Eng. Failure Anal.*, vol. 16, no. 6, pp. 1794–1804, Sep. 2009.

[8] A. Choudhary *et al.*, "Condition monitoring and fault diagnosis of induction motors: A review," *Arch. Comput. Methods Eng.*, Sep. 2018. doi: 10.1007/s11831-018-9286-z.

[9] B. Fan, S. Feng, Y. Che, J. Mao, and Y. Xie, "An oil monitoring method of wear evaluation for engine hot tests," *Int. J. Adv. Manuf. Technol.*, vol. 94, nos. 9–12, pp. 3199–3207, Feb. 2018.

[10] A. Hase, H. Mishina, and M. Wada, "Fundamental study on early detection of seizure in journal bearing by using acoustic emission technique," *Wear*, vols. 346–347, pp. 132–139, Jan. 2016.

[11] L. Wang, Z. Zhang, H. Long, J. Xu, and R. Liu, "Wind turbine gearbox failure identification with deep neural networks," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1360–1368, Jun. 2017.

[12] V. Venkatasubramanian, R. Rengaswamy, K. Yin, and S. N. Kavuri, "A review of process fault detection and diagnosis: Part I: Quantitative model-based methods," *Comput. Chem. Eng.*, vol. 27, no. 3, pp. 293–311, Mar. 2003.

[13] Z. Duan, T. Wu, S. Guo, T. Shao, R. Malekian, and 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.*, vol. 96, nos. 1–4, pp. 803–819, 2018.

[14] F. Elasha, M. Greaves, D. Mba, and D. Fang, "A comparative study of the effectiveness of vibration and acoustic emission in diagnosing a defective bearing in a planetry gearbox," *Appl. Acoust.*, vol. 115, pp. 181–195, Jan. 2017.

[15] L. Eren, T. Ince, and S. Kiranyaz, "A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier," *J. Signal Process. Syst.*, vol. 91, no. 2, pp. 179–189, Feb. 2019.

[16] A. L. Ellefsen, V. Æsøy, S. Ushakov, and H. Zhang, "A comprehensive survey of prognostics and health management based on deep learning for autonomous ships," *IEEE Trans. Rel.*, vol. 68, no. 2, pp. 720–740, Jun. 2019.

[17] H. Wang, S. Li, L. Song, and L. Cui, "A novel convolutional neural network based fault recognition method via image fusion of multi-vibration-signals," *Comput. Ind.*, vol. 105, pp. 182–190, Feb. 2019.

[18] J.-H. Zhong, J. Zhang, J. Liang, and H. Wang, "Multi-fault rapid diagnosis for wind turbine gearbox using sparse Bayesian extreme learning machine," *IEEE Access*, vol. 7, pp. 773–781, 2018.

[19] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 1, pp. 136–144, Jan. 2017.

[20] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *J. Mach. Learn. Res.*, vol. 12 pp. 2493–2537, Aug. 2011.

[21] N. Wahab, A. Khan, and Y. S. Lee, "Two-phase deep convolutional neural network for reducing class skewness in histopathological images based breast cancer detection," *Comput. Biol. Med.*, vol. 85, pp. 86–97, Jun. 2017.

[22] H. A. Pierson and M. S. Gashler, "Deep learning in robotics: A review of recent research," *Adv. Robot.*, vol. 31, no. 16, pp. 821–835, Aug. 2017.

[23] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.

[24] A. M. Karim, M. S. Güzel, M. R. Tolun, H. Kaya, and F. V. Çelebi, "A new generalized deep learning framework combining sparse autoencoder and Taguchi method for novel data classification and processing," *Math. Problems Eng.*, vol. 2018, May 2018, Art. no. 3145947. doi: 10.1155/2018/3145947.

[25] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Trans. Signal Inf. Process.*, vol. 3, pp. 1–29, Apr. 2014.

[26] G. Qiu, Y. Gu, and Q. Cai, "A deep convolutional neural networks model for intelligent fault diagnosis of a gearbox under different operational conditions," *Measurement*, vol. 145, pp. 94–107, Oct. 2019.

[27] B. P. Duong and J.-M. Kim, "Non-mutually exclusive deep neural network classifier for combined modes of bearing fault diagnosis," *Sensors*, vol. 18, no. 4, p. 1129, 2018.

[28] Z. Chen, S. Deng, X. Chen, C. Li, R.-V. Sanchez, and H. Qin, "Deep neural networks-based rolling bearing fault diagnosis," *Microelectron. Rel.*, vol. 75, pp. 327–333, Aug. 2017.

[29] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *Neurocomputing*, vol. 234, pp. 11–26, Apr. 2017.

[30] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016.

[31] D. Ravi, C. Wong, F. Deligianni, M. Berthelot, J. Andreu-Perez, B. Lo, and G.-Y. Yang, "Deep learning for health informatics," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 4–21, Jan. 2017.

[32] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.

[33] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring: A survey," Dec. 2016, *arXiv:1612.07640v1*. [Online]. Available: https://arxiv.org/abs/1612.07640v1

[34] M. Längkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognit. Lett.*, vol. 42, pp. 11–24, Jun. 2014.

[35] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mech. Syst. Signal Process.*, vol. 107, pp. 241–265, Jul. 2018.

[36] *Vibration Condition Monitoring Services*. Accessed: Jun. 27, 2019. [Online]. Available: https://reckersmech.com/vibration-condition-monitoring-north-india/

[37] F. Elasha, M. Greaves, and D. Mba, "Planetary bearing defect detection in a commercial helicopter main gearbox with vibration and acoustic emission," *Struct. Health Monit.*, vol. 17, no. 5, pp. 1192–1212, 2018.

[38] A. Klausen and K. G. Robbersmyr, "Cross-correlation of whitened vibration signals for low-speed bearing diagnostics," *Mech. Syst. Signal Process.*, vol. 118, pp. 226–244, Mar. 2019.

[39] J. L. F. Chacon, V. Kappatos, W. Balachandran, and T.-H. Gan, "A novel approach for incipient defect detection in rolling bearings using acoustic emission technique," *Appl. Acoust.*, vol. 89, pp. 88–100, Mar. 2015.

[40] N. Hiremath and D. M. Reddy, "Bearing fault detection using acoustic emission signals analyzed by empirical mode decomposition," *Int. J. Res. Eng. Technol.*, vol. 3, no. 3, pp. 426–431, 2014.

[41] B. Kilundu, X. Chiementin, J. Duez, and D. Mba, "Cyclostationarity of Acoustic Emissions (AE) for monitoring bearing defects," *Mech. Syst. Signal Process.*, vol. 25, no. 6, pp. 2061–2072, Aug. 2011.

[42] J. Seshadrinath, B. Singh, and B. K. Panigrahi, "Incipient interturn fault diagnosis in induction machines using an analytic wavelet-based optimized Bayesian inference," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 25, no. 5, pp. 990–1001, May 2014.

[43] M. Zhao and X. Jia, "A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery," *Mech. Syst. Signal Process.*, vol. 94, pp. 129–147, Sep. 2017.

[44] W. Lu, Y. Li, Y. Cheng, D. Meng, B. Liang, and P. Zhou, "Early fault detection approach with deep architectures," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 7, pp. 1679–1689, Jul. 2018.

[45] Y. Wei, Y. Li, M. Xu, and W. Huang, "A review of early fault diagnosis approaches and their applications in rotating machinery," *Entropy*, vol. 21, no. 4, pp. 1–26, 2019.

[46] M. Yuwono, Y. Qin, J. Zhou, Y. Guo, B. G. Celler, and S. W. Su, "Automatic bearing fault diagnosis using particle swarm clustering and Hidden Markov Model," *Eng. Appl. Artif. Intell.*, vol. 47, pp. 88–100, Jan. 2016.

[47] D. Yang, Y. Liu, S. Li, X. Li, and L. Ma, "Gear fault diagnosis based on support vector machine optimized by artificial bee colony algorithm," *Mechanism Mach. Theory*, vol. 90, pp. 219–229, Aug. 2015.

[48] J. Cheng, Y. Yang, and Y. Yang, "A rotating machinery fault diagnosis method based on local mean decomposition," *Digit. Signal Process.*, vol. 22, no. 2, pp. 356–366, Mar. 2012.

[49] Y. Qu, E. Bechhoefer, D. He, and J. Zhu, "A new acoustic emission sensor based gear fault detection approach," *Int. J. Prognostics Health Manage.*, vol. 4, pp. 1–14, Jul. 2013.

[50] A. Kumar and R. Kumar, "Least square fitting for adaptive wavelet generation and automatic prediction of defect size in the bearing using Levenberg–Marquardt backpropagation," *J. Nondestruct. Eval.*, vol. 36, no. 1, p. 7, 2017.

[51] X. S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation—A review on the statistical data driven approaches," *Eur. J. Oper. Res.*, vol. 213, no. 1, pp. 1–14, 2011.

[52] K. T. P. Nguyen and K. Medjaher, "A new dynamic predictive maintenance framework using deep learning for failure prognostics," *Rel. Eng. Syst. Saf.*, vol. 188, pp. 251–262, Aug. 2019.

[53] M. F. Isham, M. S. Leong, M. H. Lim, and Z. A. Ahmad, "Variational mode decomposition for rotating machinery condition monitoring using vibration signals," *Trans. Nanjing Univ. Aeronaut. Astronaut.*, vol. 35, no. 1, pp. 38–50, 2018.

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

[55] N. Saravanan, V. N. S. K. Siddabattuni, and K. I. Ramachandran, "Fault diagnosis of spur bevel gear box using artificial neural network (ANN), and proximal support vector machine (PSVM)," *Appl. Soft Comput.*, vol. 10, no. 1, pp. 344–360, Jan. 2010.

[56] C. Rajeswari, B. Sathiyabhama, S. Devendiran, and K. Manivannan, "A gear fault identification using wavelet transform, rough set based GA, ANN and C4.5 algorithm," *Procedia Eng.*, vol. 97, pp. 1831–1841, Dec. 2014.

[57] M. Kang, J. Kim, J.-M. Kim, A. C. C. Tan, E. Y. Kim, and 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.*, vol. 30, no. 5, pp. 2786–2797, May 2015.

[58] Z. Liu, H. Cao, X. Chen, Z. He, and Z. Shen, "Multi-fault classification based on wavelet SVM with PSO algorithm to analyze vibration signals from rolling element bearings," *Neurocomputing*, vol. 99, pp. 399–410, Jan. 2013.

[59] C. Shen, D. Wang, Y. B. Liu, K. Fanrang, and P. W. Tse, "Recognition of rolling bearing fault patterns and sizes based on two-layer support vector regression machines," *Smart Struct. Syst.*, vol. 13, no. 3, pp. 453–471, 2014.

[60] H. Qiu, J. Lee, J. Lin, and G. Yu, "Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics," *J. Sound Vibrat.*, vol. 289, nos. 4–5, pp. 1066–1090, 2006.

[61] W. Teng, X. Zhang, Y. Liu, A. Kusiak, and Z. Ma, "Prognosis of the remaining useful life of bearings in a wind turbine gearbox," *Energies*, vol. 10, no. 1, p. 32, 2016.

[62] Z. Zhang, Y. Wang, and K. Wang, "Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network," *J. Intell. Manuf.*, vol. 24, no. 6, pp. 1213–1227, 2013.

[63] A. Soualhi, G. Clerc, and H. Razik, "Fault prognosis based on Hidden Markov models," in *Proc. IEEE Workshop Elect. Mach. Design Control Diagnosis*, Mar. 2015, pp. 271–278.

[64] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, Nov. 2016.

[65] Z. Wei, G. Peng, and C. Li, "Bearings fault diagnosis based on convolutional neural networks with 2-D representation of vibration signals as input," in *Proc. MATEC Web Conf.*, 2017, Art. no. 13001.

[66] M. Sadoughi and C. Hu, "Physics-based convolutional neural network for fault diagnosis of rolling element bearings," *IEEE Sensors J.*, vol. 19, no. 11, pp. 4181–4192, Jun. 2019.

[67] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis," *Measurement*, vol. 93, pp. 490–502, Nov. 2016.

[68] M. Sohaib and J. M. Kim, "Reliable fault diagnosis of rotary machine bearings using a stacked sparse autoencoder-based deep neural network," *Shock Vib.*, vol. 2018, Apr. 2018, Art. no. 2919637. doi: 10.1155/2018/2919637.

[69] J. Tao, Y. Liu, and D. Yang, "Bearing fault diagnosis based on deep belief network and multisensor information fusion," *Shock Vib.*, vol. 2016, Aug. 2016, Art. no. 9306205. doi: 10.1155/2016/9306205.

[70] X.-H. He, D. Wang, Y.-F. Li, and C.-H. Zhou, "A novel bearing fault diagnosis method based on Gaussian restricted Boltzmann machine," *Math. Problems Eng.*, vol. 2016, Dec. 2016, Art. no. 2957083. doi: 10.1155/2016/2957083.

[71] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Low-speed bearing fault diagnosis based on ArSSAE model using acoustic emission and vibration signals," *IEEE Access*, vol. 7, pp. 46885–46897, 2019.

[72] M. Xia, T. Li, L. Liu, L. Xu, S. Gao, and C. W. de Silva, "Remaining useful life prediction of rotating machinery using hierarchical deep neural network," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2017, pp. 2778–2783.

[73] R. M. Hasani, G. Wang, and R. Grosu, "An automated autoencoder correlation-based health-monitoring and prognostic method for machine bearings," 2017, *arXiv:1703.06272*. [Online]. Available: https://arxiv.org/abs/1703.06272

[74] J. Ma, H. Su, W.-L. Zhao, and B. Liu, "Predicting the remaining useful life of an aircraft engine using a stacked sparse autoencoder with multilayer self-learning," *Complexity*, vol. 2018, Jul. 2018, Art. no. 3813029. doi: 10.1155/2018/3813029.

[75] J. Deutsch and D. He, "Using deep learning based approaches for bearing remaining useful life predication," in *Proc. Annu. Conf. Prognostics Health Manage. Soc.*, 2015, pp. 1–7.

[76] P. K. Wong, J. Zhong, Z. Yang, and C. M. Vong, "Sparse Bayesian extreme learning committee machine for engine simultaneous fault diagnosis," *Neurocomputing*, vol. 174, pp. 331–343, Jan. 2016.

[77] X. Yu, F. Dong, E. J. Ding, S. P. Wu, and C. Y. Fan, "Rolling bearing fault diagnosis using modified LFDA and EMD with sensitive feature selection," *IEEE Access*, vol. 6, pp. 3715–3730, 2018.

[78] M. Van and 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. C, J. Mech. Eng. Sci.*, vol. 230, no. 2, pp. 291–302, 2016.

[79] K. H. Hui, C. S. Ooi, M. H. Lim, M. S. Leong, and S. M. Al-Obaidi, "An improved wrapper-based feature selection method for machinery fault diagnosis," *PLoS ONE*, vol. 12, no. 12, 2017, Art. no. e0189143.

[80] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: A review," *Mech. Syst. Signal Process.*, vol. 108, pp. 33–47, Aug. 2018.

[81] P. Cao, S. Zhang, and J. Tang, "Preprocessing-free gear fault diagnosis using small datasets with deep convolutional neural network-based transfer learning," *IEEE Access*, vol. 6, pp. 26241–26253, 2018.

[82] M. Syahril R. M. Saufi, Z. A. Ahmad, M. H. Lim, and M. S. Leong, "A review on signal processing techniques for bearing diagnostics," *Int. J. Mech. Eng. Technol.*, vol. 8, no. 6, pp. 327–337, 2017.

[83] Z. Xu, Y. Li, Z. Wang, and J. Xuan, "A selective fuzzy ARTMAP ensemble and its application to the fault diagnosis of rolling element bearing," *Neurocomputing*, vol. 182, pp. 25–35, Mar. 2016.

[84] M. Cerrada, C. Li, R.-V. Sánchez, F. Pacheco, D. Cabrera, and J. V. de Oliveira, "A fuzzy transition based approach for fault severity prediction in helical gearboxes," *Fuzzy Sets Syst.*, vol. 337, pp. 52–73, Apr. 2018.

[85] C. Liu, G. Cheng, X. Chen, and Y. Pang, "Planetary gears feature extraction and fault diagnosis method based on VMD and CNN," *Sensors*, vol. 18, no. 5, p. 1523, 2018.

[86] H. Shao, H. Jiang, F. Wang, and Y. Wang, "Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet," *ISA Trans.*, vol. 69, pp. 187–201, Jul. 2017.

[87] Z. Guan, Z. Liao, K. Li, and P. Chen, "A precise diagnosis method of structural faults of rotating machinery based on combination of empirical mode decomposition, sample entropy, and deep belief network," *Sensors*, vol. 19, no. 3, p. 591, 2019.

[88] A. Zappone, M. Di Renzo, and M. Debbah, "Wireless networks design in the era of deep learning: Model-based, AI-based, or both?" *IEEE Trans. Commun.*, to be published.

[89] G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.

[90] X. Jian, W. Li, X. Guo, and R. Wang, "Fault diagnosis of motor bearings based on a one-dimensional fusion neural network," *Sensors*, vol. 19, no. 1, p. 122, 2019.

[91] L. Jing, M. Zhao, P. Li, and X. Xu, "A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox," *Measurement*, vol. 111, pp. 1–10, Dec. 2017.

[92] Z. Chen, X. Zeng, W. Li, and G. Liao, "Machine fault classification using deep belief network," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf.*, May 2016, pp. 1–6.

[93] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Trans.*, vol. 77, pp. 167–178, Jun. 2018.

[94] M. Ma, C. Sun, and X. Chen, "Deep coupling autoencoder for fault diagnosis with multimodal sensory data," *IEEE Trans. Ind. Informat.*, vol. 14, no. 3, pp. 1137–1145, Mar. 2018.

[95] H. Chen, N. Hu, Z. Cheng, L. Zhang, and Y. Zhang, "A deep convolutional neural network based fusion method of two-direction vibration signal data for health state identification of planetary gearboxes," *Measurement*, vol. 146, pp. 268–278, Nov. 2019. doi: 10.1016/j.measurement.2019.04.093.

[96] S. Huang, J. Tang, J. Dai, and Y. Wang, "Signal status recognition based on 1DCNN and its feature extraction mechanism analysis," *Sensors*, vol. 19, no. 9, pp. 1–19, 2019.

[97] M. He and D. He, "A new hybrid deep signal processing approach for bearing fault diagnosis using vibration signals," *Neurocomputing*, to be published.

[98] S. T. Kandukuri, A. Klausen, H. R. Karimi, and K. G. Robbersmyr, "A review of diagnostics and prognostics of low-speed machinery towards wind turbine farm-level health management," *Renew. Sustain. Energy Rev.*, vol. 53, pp. 697–708, Jan. 2016.

[99] E. Y. Kim, A. C. C. Tan, J. Mathew, and B. S. Yang, "Condition monitoring of low speed bearings: A comparative study of the ultrasound technique versus vibration measurements," *Aust. J. Mech. Eng.*, vol. 5, no. 2, pp. 177–189, 2008.

[100] W. Moustafa, O. Cousinard, F. Bolaers, K. Sghir, and J. P. Dron, "Low speed bearings fault detection and size estimation using instantaneous angular speed," *J. Vib. Control*, vol. 22, no. 15, pp. 3413–3425, 2016.

[101] C. Wu, P. Jiang, C. Ding, F. Feng, and T. Chen, "Intelligent fault diagnosis of rotating machinery based on one-dimensional convolutional network," *Comput. Ind.*, vol. 108, pp. 53–61, Jun. 2019.

[102] M. H. Zhao, M. Kang, B. Tang, and M. Pecht, "Multiple wavelet coefficients fusion in deep residual networks for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4696–4706, Jun. 2019.

[103] M. Cerrada, G. Zurita, D. Cabrera, R.-V. Sánchez, M. Artés, and C. Li, "Fault diagnosis in spur gears based on genetic algorithm and random forest," *Mech. Syst. Signal Process.*, vols. 70–71, pp. 87–103, Mar. 2016.

[104] Z. Chen, C. Li, and R. Sanchez, "Gearbox fault identification and classification with convolutional neural networks," *Shock Vib.*, vol. 2015, Apr. 2015, Art. no. 390134. doi: 10.1155/2015/390134.

[105] X. Yan and M. Jia, "A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing," *Neurocomputing*, vol. 313, pp. 47–64, Nov. 2018.

[106] W. Caesarendra and T. Tjahjowidodo, "A review of feature extraction methods in vibration-based condition monitoring and its application for degradation trend estimation of low-speed slew bearing," *Machines*, vol. 5, no. 4, p. 21, 2017.

[107] D. Goyal, B. S. Pabla, and S. S. Dhami, "Condition monitoring parameters for fault diagnosis of fixed axis gearbox: A review," *Arch. Comput. Methods Eng.*, vol. 24, no. 3, pp. 543–556, Jul. 2017.

[108] Z. Xiang, X. Zhang, W. Zhang, and X. Xia, "Fault diagnosis of rolling bearing under fluctuating speed and variable load based on TCO Spectrum and Stacking Auto-encoder," *Measurement*, vol. 138, pp. 162–174, May 2019.

[109] S. Ma and F. Chu, "Ensemble deep learning-based fault diagnosis of rotor bearing systems," *Comput. Ind.*, vol. 105, pp. 143–152, Feb. 2019.

[110] C. Li, R. V. Sánchez, G. Zurita, M. Cerrada, and D. Cabrera, "Fault diagnosis for rotating machinery using vibration measurement deep statistical feature learning," *Sensors*, vol. 16, no. 6, p. 895, 2016.

[111] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock Vib.*, vol. 2017, Oct. 2017, Art. no. 5067651. doi: 10.1155/2017/5067651.

[112] S. Shao, S. McAleer, R. Yan, and P. Baldi, "Highly accurate machine fault diagnosis using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2446–2455, Apr. 2019.

[113] S. Guo, T. Yang, W. Gao, C. Zhang, and Y. Zhang, "An intelligent fault diagnosis method for bearings with variable rotating speed based on pythagorean spatial pyramid pooling CNN," *Sensors*, vol. 18, no. 11, p. 3857, 2018.

[114] C. Zhong, K. Yan, Y. Dai, N. Jin, and B. Lou, "Energy efficiency solutions for buildings: Automated fault diagnosis of air handling units using generative adversarial networks," *Energies*, vol. 12, no. 3, p. 527, 2019.

[115] C. Yuanhang, P. Gaoliang, X. Chaohao, Z. Wei, L. Chuanhao, and L. Shaohui, "ACDIN: Bridging the gap between artificial and real bearing damages for bearing fault diagnosis," *Neurocomputing*, vol. 294, pp. 61–71, Jun. 2018.

[116] S. Ramanathan, "5 challenges for deep learning," EE Times, San Francisco, CA, USA, Tech. Rep., 2018. [Online]. Available: https://www.eetimes.com/author.asp?section_id=36&doc_id=1333535

[117] Z. Meng, X. Zhan, J. Li, and Z. Pan, "An enhancement denoising autoencoder for rolling bearing fault diagnosis," *Measurement*, vol. 130, pp. 448–454, Dec. 2018.

[118] H. Shao, H. Jiang, H. Zhao, and F. Wang, "A novel deep autoencoder feature learning method for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 95, pp. 187–204, Oct. 2017.

[119] J. J. Hasan, M. M. M. Islam, and J.-M. Kim, "Acoustic spectral imaging and transfer learning for reliable bearing fault diagnosis under variable speed conditions," *Measurement*, vol. 138, pp. 620–631, May 2019.

[120] H. Shao, H. Jiang, F. Wang, and H. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Knowl.-Based Syst.*, vol. 119, pp. 200–220, Mar. 2017.

[121] H. O. A. Ahmed, M. L. D. Wong, and A. K. Nandi, "Effects of deep neural network parameters on classification of bearing faults," in *Proc. 42nd Annu. Conf. IEEE Ind. Electron. Soc. (IECON)*, Oct. 2016, pp. 0–5.

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

[123] C. Li, W. Zhang, G. Peng, and S. Liu, "Bearing fault diagnosis using fully-connected winner-take-all autoencoder," *IEEE Access*, vol. 6, pp. 6103–6115, 2017.

[124] J. Jiao, M. Zhao, J. Lin, and C. Ding, "Deep coupled dense convolutional network with complementary data for intelligent fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9858–9867, Dec. 2019.

[125] J. Grezmak, P. Wang, C. Sun, and R. X. Gao, "Explainable convolutional neural network for gearbox fault diagnosis," *Procedia CIRP*, vol. 80, pp. 476–481, May 2019.

[126] M. Gan, C. Wang, and C. Zhu, "Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 92–104, May 2016.

[127] T. Han, C. Liu, W. Yang, and D. Jiang, "A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults," *Knowl.-Based Syst.*, vol. 165, pp. 474–487, Feb. 2019.

[128] S. R. Saufi, Z. A. bin Ahmad, M. S. Leong, and M. H. Lim, "Differential evolution optimization for resilient stacked sparse autoencoder and its applications on bearing fault diagnosis," *Meas. Sci. Technol.*, vol. 29, no. 12, 2018, Art. no. 125002.

[129] A. Darwish, A. E. Hassanien, and S. Das, "A survey of swarm and evolutionary computing approaches for deep learning," in *Artificial Intelligence Review*. Amsterdam, The Netherlands: Springer, 2019, pp. 1–46. doi: 10.1007/s10462-019-09719-2.

[130] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 8609–8613.

[131] M. Zhao, M. Kang, B. Tang, and M. Pecht, "Deep residual networks with dynamically weighted wavelet coefficients for fault diagnosis of planetary gearboxes," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4290–4300, May 2018.

[132] S. Ma, W. Liu, W. Cai, Z. Shang, and G. Liu, "Lightweight deep residual CNN for fault diagnosis of rotating machinery based on depthwise separable convolutions," *IEEE Access*, vol. 7, pp. 57023–57036, 2019.

[133] H. Shao, H. Jiang, Y. Lin, and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Mech. Syst. Signal Process.*, vol. 102, pp. 278–297, Mar. 2018.

[134] M. Andrychowicz, M. Denil, S. Gomez, M. W. Hoffman, D. Pfau, T. Schaul, B. Shillingford, and N. de Freitas, "Learning to learn by gradient descent by gradient descent," pp. 1–17, 2016, *arXiv:1606.04474*. [Online]. Available: https://arxiv.org/abs/1606.04474

[135] W. Gong, H. Chen, Z. Zhang, M. Zhang, R. Wang, C. Guan, and Q. Wang, "A novel deep learning method for intelligent fault diagnosis of rotating machinery based on improved CNN-SVM and multichannel data fusion," *Sensors*, vol. 19, no. 7, p. 1693, 2019.

[136] C.-F. Lai, W.-C. Chien, L. T. Yang, and W. Qiang, "LSTM and edge computing for big data feature recognition of industrial electrical equipment," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2469–2477, Apr. 2019.

[137] H. Pan, X. He, S. Tang, and F. Meng, "An improved bearing fault diagnosis method using one-dimensional CNN and LSTM," *J. Mech. Eng.*, vol. 64, nos. 7–8, pp. 443–452, May 2018.

[138] S. Tang, C. Shen, D. Wang, S. Li, W. Huang, and Z. Zhu, "Adaptive deep feature learning network with Nesterov momentum and its application to rotating machinery fault diagnosis," *Neurocomputing*, vol. 305, pp. 1–14, Aug. 2018.

[139] L. Luo, Y. Xiong, Y. Liu, and X. Sun, "Adaptive gradient methods with dynamic bound of learning rate," pp. 1–19, 2019, *arXiv:1902.09843*. [Online]. Available: https://arxiv.org/abs/1902.09843

[140] L. Ren, J. Cui, Y. Sun, and X. Cheng, "Multi-bearing remaining useful life collaborative prediction: A deep learning approach," *J. Manuf. Syst.*, vol. 43, pp. 248–256, Apr. 2017.

[141] Y.-R. Wang, Q. Jin, G.-D. Sun, and C.-F. Sun, "Planetary gearbox fault feature learning using conditional variational neural networks under noise environment," *Knowl.-Based Syst.*, vol. 163, pp. 438–449, Jan. 2019.

[142] S. Heo and J. H. Lee, "Fault detection and classification using artificial neural networks," *IFAC-PapersOnLine*, vol. 51, no. 18, pp. 470–475, 2018.

[143] L. Dong, S. Liu, and H. Zhang, "A method of anomaly detection and fault diagnosis with online adaptive learning under small training samples," *Pattern Recognit.*, vol. 64, pp. 374–385, Apr. 2017.

[144] S. Karsoliya, "Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture," *Int. J. Eng. Trends Technol.*, vol. 3, no. 6, pp. 714–717, 2012.

[145] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.

[146] S.-Y. Shao, W.-J. Sun, R.-Q. Yan, P. Wang, and R. X. Gao, "A deep learning approach for fault diagnosis of induction motors in manufacturing," *Chin. J. Mech. Eng.*, vol. 30, no. 6, pp. 1347–1356, 2017.

[147] K. G. Sheela and S. N. Deepa, "Review on methods to fix number of hidden neurons in neural networks," *Math. Problems Eng.*, vol. 2013, May 2013, Art. no. 425740. doi: 10.1155/2013/425740.

[148] Y. Xu, Y. Sun, J. Wan, X. Liu, and Z. Song, "Industrial big data for fault diagnosis: Taxonomy, review, and applications," *IEEE Access*, vol. 5, pp. 17368–17380, 2017.

[149] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Inf. Fusion*, vol. 42, pp. 146–157, Nov. 2017.

[150] T. Wang, Q. Han, F. Chu, and Z. Feng, "Vibration based condition monitoring and fault diagnosis of wind turbine planetary gearbox: A review," *Mech. Syst. Signal Process.*, vol. 126, pp. 662–685, Jul. 2019.

[151] G. Xu, M. Liu, Z. Jiang, D. Söffker, and W. Shen, "Bearing fault diagnosis method based on deep convolutional neural network and random forest ensemble learning," *Sensors*, vol. 19, no. 5, p. 1088, 2019.

[152] L. Saidi, J. B. Ali, E. Bechhoefer, and M. Benbouzid, "Wind turbine high-speed shaft bearings health prognosis through a spectral Kurtosis-derived indices and SVR," *Appl. Acoust.*, vol. 120, pp. 1–8, May 2017.

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

[154] Y. Liu, L. Duan, Z. Yuan, N. Wang, and J. Zhao, "An intelligent fault diagnosis method for reciprocating compressors based on LMD and SDAE," *Sensors*, vol. 19, no. 5, p. 1041, 2019.

[155] D. Peng, Z. Liu, H. Wang, Y. Qin, and L. Jia, "A novel deeper one-dimensional CNN with residual learning for fault diagnosis of wheelset bearings in high-speed trains," *IEEE Access*, vol. 7, pp. 10278–10293, 2019.

[156] L. Sun, H. Bai, G. Zhao, and X. Wang, "Review of diagnosis technique for equipment faults and its development trend," in *Proc. MATEC Web Conf.*, vol. 22, 2015, p. 03007.

[157] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation," in *Journal of Intelligent Manufacturing*. New York, NY, USA: Springer, 2018, pp. 1–20. doi: 10.1007/s10845-018-1456-1.

[158] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, "Learning from class-imbalanced data: Review of methods and applications," *Expert Syst. Appl.*, vol. 73, pp. 220–239, May 2017.

[159] X. Chen, Z. Wang, Z. Zhang, L. Jia, and Y. Qin, "A semi-supervised approach to bearing fault diagnosis under variable conditions towards imbalanced unlabeled data," *Sensors*, vol. 18, no. 7, p. 2097, 2018.

[160] Y. Zhang, X. Li, L. Gao, L. Wang, and L. Wen, "Imbalanced data fault diagnosis of rotating machinery using synthetic oversampling and feature learning," *J. Manuf. Syst.*, vol. 48, pp. 34–50, Jul. 2018.

[161] L. Chen, G. Xu, Q. Zhang, and X. Zhang, "Learning deep representation of imbalanced SCADA data for fault detection of wind turbines," *Measurement*, vol. 139, pp. 370–379, Jun. 2019.

[162] F. Jia, Y. Lei, N. Lu, and S. Xing, "Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization," *Mech. Syst. Signal Process.*, vol. 110, pp. 349–367, Sep. 2018.

[163] C. Li, D. Zhao, S. Mu, W. Zhang, N. Shi, and L. Li, "Fault diagnosis for distillation process based on CNN–DAE," *Chin. J. Chem. Eng.*, vol. 27, no. 3, pp. 598–604, Mar. 2019.

[164] E. Hosseini-Asl, J. M. Zurada, and O. Nasraoui, "Deep learning of part-based representation of data using sparse autoencoders with nonnegativity constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 12, pp. 2486–2498, Dec. 2016.

[165] Y. Ju, J. Guo, and S. Liu, "A deep learning method combined sparse autoencoder with SVM," in *Proc. Int. Conf. Cyber-Enabled Distrib. Comput. Knowl. Discovery*, Sep. 2015, pp. 257–260.

[166] Y. Yu, T. McKelvey, and S.-Y. Kung, "A classification scheme for 'high-dimensional-small-sample-size' data using soda and ridge-SVM with microwave measurement applications," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2013, pp. 3542–3546.

[167] C. Li, R.-V. Sanchez, G. Zurita, M. Cerrada, D. Cabrera, and R. E. Vásquez, "Gearbox fault diagnosis based on deep random forest fusion of acoustic and vibratory signals," *Mech. Syst. Signal Process.*, vols. 76–77, pp. 283–293, Aug. 2016.

[168] K. Li, N. Yu, P. Li, S. Song, Y. Wu, Y. Li, and M. Liu, "Multi-label spacecraft electrical signal classification method based on DBN and random forest," *PLoS ONE*, vol. 12, no. 5, 2017, Art. no. e0176614.

[169] R. P. Monteiro, M. Cerrada, D. R. Cabrera, R. V. Sánchez, and C. J. A. Bastos-Filho, "Using a support vector machine based decision stage to improve the fault diagnosis on gearboxes," *Comput. Intell. Neurosci.*, vol. 2019, Feb. 2019, Art. no. 1383752. doi: 10.1155/2019/1383752.

[170] F. Cheng, J. Wang, L. Qu, and W. Qiao, "Rotor current-based fault diagnosis for DFIG wind turbine drivetrain gearboxes using frequency analysis and a deep classifier," in *Proc. IEEE Ind. Appl. Soc. Annu. Meeting (IAS)*, vol. 54, no. 2, Oct. 2017, pp. 1–9.

[171] W. Mao, W. Feng, X. Liang, and X. Zhang, "A novel deep output kernel learning method for bearing fault structural diagnosis," *Mech. Syst. Signal Process.*, vol. 117, no. 2, pp. 293–318, 2019.

[172] K. Li, M. Xiong, F. Li, L. Su, and J. Wu, "A novel fault diagnosis algorithm for rotating machinery based on a sparsity and neighborhood preserving deep extreme learning machine," *Neurocomputing*, vol. 350, pp. 261–270, Jul. 2019.

[173] S. Wang, J. Xiang, Y. Zhong, and Y. Zhou, "Convolutional neural network-based hidden Markov models for rolling element bearing fault identification," *Knowl.-Based Syst.*, vol. 144, pp. 65–76, Mar. 2018.

[174] G. Hu, H. Li, Y. Xia, and L. Luo, "A deep Boltzmann machine and multi-grained scanning forest ensemble collaborative method and its application to industrial fault diagnosis," *Comput. Ind.*, vol. 100, pp. 287–296, Sep. 2018.

[175] X. Li, Y. Yang, H. Pan, J. Cheng, and J. Cheng, "A novel deep stacking least squares support vector machine for rolling bearing fault diagnosis," *Comput. Ind.*, vol. 110, pp. 36–47, Sep. 2019.

[176] T. Domhan, J. T. Springenberg, and F. Hutter, "Speeding up automatic hyperparameter optimization of deep neural networks by extrapolation of learning curves," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, vol. 15, 2015, pp. 3460–3468.

[177] M. N. Ab Wahab, S. Nefti-Meziani, and A. Atyabi, "A comprehensive review of swarm optimization algorithms," *PLoS ONE*, vol. 10, no. 5, 2015, Art. no. e0122827.

[178] I. Fister, Jr., X.-S. Yang, I. Fister, J. Brest, and D. Fister, "A brief review of nature-inspired algorithms for optimization," *Elektrotehni ki Vestnik*, vol. 80, no. 3, pp. 116–122, 2013.

[179] Z. Beheshti and S. M. Shamsuddin, "A review of population-based meta-heuristic algorithm," *Int. J. Adv. Soft Comput. Appl.*, vol. 5, no. 1, pp. 1–35, 2015.

[180] S. Mirjalili, "The ant lion optimizer," *Adv. Eng. Softw.*, vol. 83, pp. 80–98, May 2015.

[181] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014.

[182] H. D. Shao, H. K. Jiang, X. Zhang, and M. G. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, no. 11, p. 115002, Nov. 2015.

[183] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012.

[184] J. T. Springenberg, A. Klein, S. Falkner, and F. Hutter, "Bayesian optimization with robust Bayesian neural networks," in *Proc. NIPS*, vol. 4, 2016, pp. 4134–4142.

[185] L. M. R. Rere, M. I. Fanany, and A. M. Arymurthy, "Metaheuristic algorithms for convolution neural network," *Comput. Intell. Neurosci.*, vol. 2016, 2016, Art. no. 1537325. doi: 10.1155/2016/1537325.

[186] H. Guo *et al.*, "Deep neural network and whale optimization algorithm to assess flyrock induced by blasting," in *Engineering With Computers*. London, U.K.: Springer, 2019, pp. 1–14. doi: 10.1007/s00366-019-00816-y.

[187] I. Aljarah, H. Faris, and S. Mirjalili, "Optimizing connection weights in neural networks using the whale optimization algorithm," *Soft Comput.*, vol. 22, no. 1, pp. 1–15, 2016.

[188] E. M. Hassib, A. I. El-Desouky, L. M. Labib, and E.-S. M. El-kenawy, "WOA+BRNN: An imbalanced big data classification framework using Whale optimization and deep neural networkk," in *Soft Computing*. Berlin, Germany: Springer, 2019. doi: 10.1007/s00500-019-03901-y.

[189] J. Zhang, P. Wang, R. Yan, and R. X. Gao, "Deep learning for improved system remaining life prediction," *Procedia CIRP*, vol. 72, pp. 1033–1038, Jan. 2018.

[190] H. Shao, H. Jiang, X. Li, and T. Liang, "Rolling bearing fault detection using continuous deep belief network with locally linear embedding," *Comput. Ind.*, vol. 96, pp. 27–39, Apr. 2018.

[191] L. Ren, Y. Sun, J. Cui, and L. Zhang, "Bearing remaining useful life prediction based on deep autoencoder and deep neural networks," *J. Manuf. Syst.*, vol. 48, pp. 71–77, Jul. 2018.

[192] Q. Wang, B. Zhao, H. Ma, J. Chang, and G. Mao, "A method for rapidly evaluating reliability and predicting remaining useful life using two-dimensional convolutional neural network with signal conversion," *J. Mech. Sci. Technol.*, vol. 33, no. 6, pp. 2561–2571, Jun. 2019.

[193] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Rel. Eng. Syst. Saf.*, vol. 172, pp. 1–11, Apr. 2018.

[194] J. Zhu, N. Chen, and W. Peng, "Estimation of bearing remaining useful life based on multiscale convolutional neural network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3208–3216, Apr. 2019.

[195] X. Li, W. Zhang, and Q. Ding, "Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction," *Rel. Eng. Syst. Saf.*, vol. 182, pp. 208–218, Feb. 2019.

[196] J. Deutsch and D. He, "Using deep learning-based approach to predict remaining useful life of rotating components," *IEEE Trans. Syst. Man, Cybern., Syst.*, vol. 48, no. 1, pp. 11–20, Jan. 2017.

[197] A. L. Ellefsen, E. Bjørlykhaug, V. Æsøy, S. Ushakov, and H. Zhang, "Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture," *Rel. Eng. Syst. Saf.*, vol. 183, pp. 240–251, Mar. 2019.

[198] A. Al-Dulaimi, S. Zabihi, A. Asif, and A. Mohammadi, "A multimodal and hybrid deep neural network model for remaining useful life estimation," *Comput. Ind.*, vol. 108, pp. 186–196, Jun. 2019.

[199] A. Zhang, H. Wang, S. Li, Y. Cui, Z. Liu, G. Yang, and J. Hu, "Transfer learning with deep recurrent neural networks for remaining useful life estimation," *Appl. Sci.*, vol. 8, no. 12, p. 2416, Nov. 2018.

[200] K. H. Hui, M. H. Lim, M. S. Leong, and S. M. Al-Obaidi, "Dempster-Shafer evidence theory for multi-bearing faults diagnosis," *Eng. Appl. Artif. Intell.*, vol. 57, pp. 160–170, Jan. 2017.

[201] S. Sarkar, S. Sarkar, K. Mukherjee, A. Ray, and A. Srivastav, "Multi-sensor information fusion for fault detection in aircraft gas turbine engines," *Proc. Inst. Mech. Eng. G, J. Aerosp. Eng.*, vol. 227, no. 12, pp. 1988–2001, 2013.

[202] W. Yang, R. Court, and J. Jiang, "Wind turbine condition monitoring by the approach of SCADA data analysis," *Renew. Energy*, vol. 53, pp. 365–376, May 2013.

[203] L. Jing, T. Wang, M. Zhao, and P. Wang, "An adaptive multi-sensor data fusion method based on deep convolutional neural networks for fault diagnosis of planetary gearbox," *Sensors*, vol. 17, no. 2, p. 414, 2017.

[204] M. vokelj, S. Zupan, and I. Prebil, "Multivariate and multiscale monitoring of large-size low-speed bearings using ensemble empirical mode decomposition method combined with principal component analysis," *Mech. Syst. Signal Process.*, vol. 24, no. 4, pp. 1049–1067, May 2010.

[205] S. Zhang and W. Li, "Bearing condition recognition and degradation assessment under varying running conditions using NPE and SOM," *Math. Problems Eng.*, vol. 2014, May 2014, Art. no. 781583. doi: 10.1155/2014/781583.

[206] L. van der Maaten and G. Hinton, "Visualizing high-dimensional data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[207] Y. Yang, P. Fu, and Y. He, "Bearing fault automatic classification based on deep learning," *IEEE Access*, vol. 6, pp. 71540–71554, 2018.

[208] W. Mao, J. He, Y. Li, and Y. Yan, "Bearing fault diagnosis with auto-encoder extreme learning machine: A comparative study," *J. Mech. Eng. Sci.*, vol. 231, no. 8, pp. 1560–1578, 2017.

[209] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Machine learning and deep learning algorithms for bearing fault diagnostics—A comprehensive review," 2019, *arXiv:1901.08247*. [Online]. Available: https://arxiv.org/abs/1901.08247

[210] L. Guo, Y. Lei, S. Xing, T. Yan, and N. Li, "Deep convolutional transfer learning network: A new method for intelligent fault diagnosis of machines with unlabeled data," *IEEE Trans. Ind. Electron.*, vol. 66, no. 9, pp. 7316–7325, Sep. 2018.

[211] Z. Zhu, G. Peng, Y. Chen, and H. Gao, "A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis," *Neurocomputing*, vol. 323, pp. 62–75, Jan. 2019.

[212] H. Al-Atat, D. Siegel, and J. Lee, "A systematic methodology for gearbox health assessment and fault classification," *Int. J. Prognostics Health Manage.*, vol. 2, no. 1, pp. 1–16, 2011.

[213] K. Loparo. *Case Western Reserve University Bearing Data Center Website*. Accessed: May 25, 2018. [Online]. Available: https://csegroups.case.edu/bearingdatacenter/pages/project-history

[214] "Prognostic data repository: Bearing data set NSF I/UCRC center for intelligent maintenance systems," Aeronaut. Space Admin., Tech. Rep., 2010. [Online]. Available: http://ti.arc.nasa.gov/tech/dash/pcoe/prognostic-data-repository/

[215] F. M. Ribeiro. (2017). *MaFaulDa-Machinery Fault Database*. Accessed: Nov. 30, 2017. [Online]. Available: http://www02.smt.ufrj.br/~offshore/mfs/index.html#TOC1

[216] C. Lessmeier, O. Enge-Rosenblatt, C. Bayer, and D. Zimmer, "Data Acquisition and Signal Analysis from Measured Motor Currents for Defect Detection in Electromechanical Drive Systems," in *Proc. Eur. Conf. Prognostics Health Manage. Soc.*, 2014, pp. 1–10.

[217] *IEEE PHM 2012 Data Challenge Bearing Dataset │ Acoustics and Vibration Database*. Accessed: Aug.. 1, 2019. [Online]. Available: http://data-acoustics.com/measurements/bearing-faults/bearing-6/

[218] R. K. Singleton, E. G. Strangas, and S. Aviyente, "Extended Kalman filtering for remaining-useful-life estimation of bearings," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1781–1790, Mar. 2015.

[219] E. Bechhoefer. *Acoustics and Vibration Database*. Accessed: Jun. 25, 2018. [Online]. Available: http://data-acoustics.com/measurements/gear-faults/gear-1/

[220] F. O. Heimes, "Recurrent neural networks for remaining useful life estimation," in *Proc. Int. Conf. Prognostics Health Manage. (PHM)*, Oct. 2008, pp. 1–6.

[221] G. L. Forbes and R. B. Randall, "Estimation of turbine blade natural frequencies from casing pressure and vibration measurements," *Mech. Syst. Signal Process.*, vol. 36, no. 2, pp. 549–561, Apr. 2013.

[222] A. Saxena and K. Goebel, "Turbofan engine degradation simulation data set," NASA Ames Prognostics Data Repository, NASA Ames Res. Center, Moffett Field, CA, USA, 2008. [Online]. Available: https://tiarcnasagov/tech/dash/groups/pcoe/prognostic-data-repository/

[223] E. Ramasso and A. Saxena, "Performance benchmarking and analysis of prognostic methods for CMAPSS datasets," *Int. J. Prognostics Health Manage.*, vol. 5, p. 15, 2014.

ZAIR ASRAR BIN AHMAD received the Ph.D. degree from the Faculty of Mechanical Engineering, Otto-von-Guericke-University Magdeburg, Magdeburg. He is currently a Senior Lecturer with the School of Mechanical Engineering, Faculty of Engineering, Universiti Teknologi Malaysia, Skudai, Malaysia. His current research interests include civil engineering, acoustic engineering, and structural engineering.

MOHD SALMAN LEONG has more than 35 years professional engineering consulting experience, and is acknowledged by the industry and government agencies as the leading authority in acoustics, noise, and vibration in the country. He has been involved in many of the mega-projects and high impact consulting and investigation projects in oil and gas, power generation, infrastructure and construction industries. He is currently a Professor and a Principal Consultant, the Founding Director of the Institute of Noise and Vibration, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.

SYAHRIL RAMADHAN SAUFI is currently pursuing the Ph.D. degree with the Institute of Noise and Vibration, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia. His current research interests fault diagnosis artificial intelligence and structural assessment.

MENG HEE LIM has more than 13 years professional consulting experience in acoustics and vibration, with specialization in machinery diagnostics. He has extensive experience in power generation gas and steam turbines, blade failures diagnostics and experimental modal analysis. He is currently a Senior Lecturer and a Consultant with the Institute of Noise and Vibration, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia.

• • •