

Received April 1, 2020, accepted April 18, 2020, date of publication April 27, 2020, date of current version June 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2990528

Bearing Fault Detection and Diagnosis Using Case Western Reserve University Dataset With Deep Learning Approaches: A Review

DHIRAJ NEUPANE^{ID} AND **JONGWON SEOK**^{ID}

Department of Information and Communication Engineering, Changwon National University, Changwon 51140, South Korea

Corresponding author: Jongwon Seok (jwseok@changwon.ac.kr)

This work was supported by the Changwon National University, in 2019–2020.

ABSTRACT A smart factory is a highly digitized and connected production facility that relies on smart manufacturing. Additionally, artificial intelligence is the core technology of smart factories. The use of machine learning and deep learning algorithms has produced fruitful results in many fields like image processing, speech recognition, fault detection, object detection, or medical sciences. With the increment in the use of smart machinery, the faults in the machinery equipment are expected to increase. Machinery fault detection and diagnosis through various deep learning algorithms has increased day by day. Many types of research have been done and published using both open-source and closed-source datasets, implementing the deep learning algorithms. Out of many publicly available datasets, Case Western Reserve University (CWRU) bearing dataset has been widely used to detect and diagnose machinery bearing fault and is accepted as a standard reference for validating the models. This paper summarizes the recent works which use the CWRU bearing dataset in machinery fault detection and diagnosis employing deep learning algorithms. We have reviewed the published works and presented the working algorithm, result, and other necessary details in this paper. This paper, we believe, can be of good help for future researchers to start their work on machinery fault detection and diagnosis using the CWRU dataset.

INDEX TERMS Bearing, deep learning, machine learning, machinery fault detection and diagnosis, CWRU dataset.

I. INTRODUCTION

Nowadays, electric machines are used ubiquitously in manufacturing applications. With the rapid growth and improvement in science and technology, and the development of modern industries, machinery equipment is operated in daily basis and for almost all applications which, sometimes, make these machines work under unfavorable conditions, humidity and excessive loads resulting in motor breakdowns leading to huge maintenance expenses, degradation in production level, severe monetary losses, and potential risk of loss of lives. The rotating machines and induction engines play a crucial role in the industrial systems. These rotating machines are composed of numerous elements, such as stator, rotor, shaft, and bearings [1]. Rolling element bearings, also commonly known as bearings, are the core and vulnerable components in the machinery whose health condition, i.e., the crack or

faults in different places when operated under varying load, directly affects in the performance and efficiency, stability, and lifespan of the machines [2]. The rolling element bearing (REB) consists of four components: inner-race, outer-race, ball and cage. Fig. 1 shows the experimental platform of the CWRU bearing test rig for the ball bearing system [3], [4], the bearing components, and its cross-sectional view.

Numerous studies [5], [6], regarding the possibility of induction engine failures, reveal that a bearing fault is the top common fault category that accounts for one-third of the entire defects, and the failure of these REBs is one of the most common causes of machine breakdown resulting in severe loss of safety and property and even the crash of the machine or loss of the lives in some cases [7]. For these reasons, fault detection and diagnosis of these REBs have become an essential part of development and engineering research. Condition monitoring and fault detection mechanism of REBs are expected to provide information about the real working state of the machinery at each moment

The associate editor coordinating the review of this manuscript and approving it for publication was Inês Domingues^{ID}.

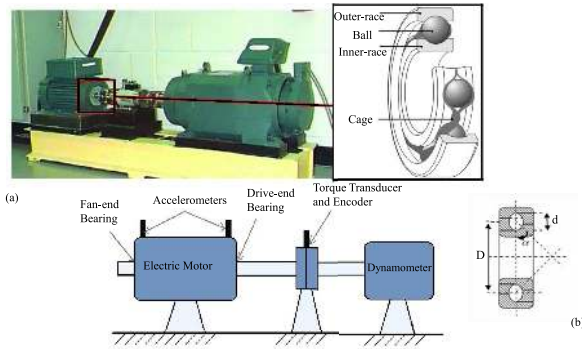


FIGURE 1. (a) An Experimental platform of the CWRU bearing test rig for ball bearing system [3], [4], the components of REB, and (b) its cross-sectional view.

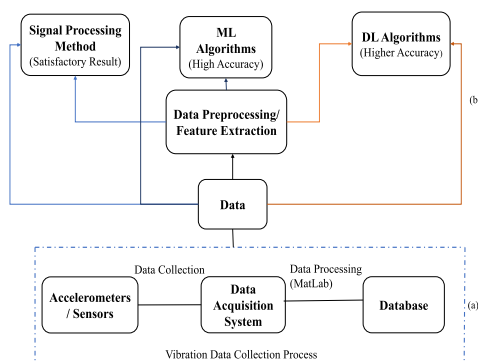


FIGURE 2. (a) Vibration data collection process, (b) general flow diagram of the implementation of different techniques and their respective performance for bearing fault detection using vibration data.

without stopping the production line. Moreover, for the proper understanding of processes associated with bearing faults, the mechanical vibration signals, which can detect, locate and distinguish different types of faults, are considered as one of the most useful and productive sources of information [1], [8].

The working procedure for the bearing fault detection consists of sensors placed at different locations of the machine through which these signals are sent to the data acquisition system for further processing [9]. Fig. 2 (a) shows the vibration data collection process for the CWRU dataset. The performance of fault detection methods does not only depend on the quality of the vibration signals collected, but also on the effectiveness of the applied signal processing and feature extraction techniques [1]. Traditionally, the maintenance of these types of machinery used to be performed after the machine fault had happened. Such a posterior maintenance approach generally leads to unexpected machine breakdown resulting in financial losses and casualties [10]. Thus, it is significant to monitor the bearing health condition on the working state of the machine. For detecting and diagnosing the faults in machinery and REBs, many signal processing approaches, machine learning (ML)-based approaches, and deep learning (DL)-based approaches have been proposed and implemented.

The conventional signal processing methods can be analyzed in the time-domain, frequency-domain, and time-frequency-domain. The methods like fast Fourier transform (FFT) [11], wavelet transformation (WT) [12], empirical mode decomposition (EMD) [13], ensemble empirical mode decomposition (EEMD) [14], empirical wavelet transform (EWT), wavelet packet transform (WPT) [15], variational mode decomposition (VMD), stochastic resonance, sparse decomposition, etc., have been suggested and implemented for the fault signal analysis and classification of the collected vibration data. These approaches belong to either of time-domain or frequency-domain or time-frequency domain analysis.

Time-domain analysis, which usually involves the scalar indices to determine bearing condition, is the most straightforward technique for detecting and diagnosing faults in bearings [16]. In this analysis, the signal is studied through temporal vibrational signal data, and based on its value, the bearing condition can be estimated. Some of the approaches in this method are measuring the peak value, peak-to-peak value, root-mean-square (RMS), and crest factor [17], skewness, kurtosis [18], and spectral kurtosis [19], impulse factor, shape factor, and clearance factor [20]. Similarly, frequency-domain or spectral analysis is the most broadly used method for fault diagnosis in REBs. Here, the time-domain vibration signals are converted to discrete frequency components using FFT. The FFT and discrete Fourier transform (DFT) algorithms are used to analyze the raw vibration signal in the frequency-domain. The advantage of frequency-domain over the time-domain analysis is that the frequency-domain method can detect the required specific frequency components effortlessly [16]. The bearing vibration signals are the combination of periodic components, which are cyclic, time invariance and non-stationary, leading to the development of time-frequency analysis approaches. The frequency-domain technique has the advantage that it can handle both the stationary and non-stationary vibration signals. Many time-frequency strategies like short time Fourier transform (STFT) [21], Wigner-Ville distribution [22], and wavelet transform [12] have been used in machinery fault detection and diagnosis (MFDD).

Even though a suitable height of anomaly diagnosis accuracies was stated, the signal processing methods have some kinds of disadvantages. Temporal analysis cannot determine the defective component of a machine. The frequency peak of bearing fault is not strong enough to be distinguished in FFT analysis of bearing fault. Similarly, cepstrum analysis produces many unwanted large peaks near the zero point, which makes the output difficult to interpret. Moreover, it uses several FFTs and IFFTs on each frame, which can be computationally expensive. Envelope analysis requires the knowledge of the resonance frequency and filtering band in advance, which demands some experience. Wavelet transform, on the other hand, comes up with problems like difficulty in the selection of the suitable mother wavelet, the choice of decomposition level, and its frequency band,

which contains the necessary information for fault detection [1]. Again, the maximum of these methods needs different features for different types of vibration data. The conventional methods entirely depend on the values at the fault characteristics frequencies to determine the existence of a bearing fault [2]. Features such as mean, median, kurtosis, peak-to-peak, minimum, maximum, standard deviation, absolute mean, skewness, and RMS are the general characteristics of the vibrating signal, which help describe the exact condition of the bearing. From this, one can easily guess how troublesome and crucial it is to choose the right features to characterize the specific signals that are used for the classification. Thus, the manually selected or hand-crafted features may not best describe the motor bearing data, and the data cannot achieve the fundamental solution for the detection and diagnosis. Putting it simply, what features extraction is needed for the optimal choice for a particular signal are remained unanswered up to date [23], [24]. There may occur some exceptional patterns or relationships unseen in the data themselves that can expose a bearing fault, and these distinctive features can be almost impossible for human beings to identify at the initial stage. Thus, many scholars have applied several machine learning procedures for the detection of bearing faults from those vibration data.

Machine learning is a branch of artificial intelligence, which is used to teach machines how to handle data more efficiently. Its working algorithm can be summarized in three steps: It takes some data, finds patterns in data, and predicts new patterns from the data. ML algorithms are used extensively in almost every sector, like speech or text recognition, industrial applications, medical diagnosis, social networking sites, and finance [25]. In MFDD, the use of these ML algorithms for developing a knowledge base system has become useful for early diagnosis of a defect to prevent catastrophic failure and reduce operating costs [26]. In bearing fault detection some of the broadly practiced ML algorithms are artificial neural networks (ANN) [27], principal component analysis (PCA) [28], support vector machines (SVM) [29], [30], k-Nearest Neighbors (k-NN) [31], singular value decomposition (SVD) [32]. These algorithms better analyze the data, learn from them, and then apply what they have learned to make smart decisions concerning the occurrence of bearing faults. Furthermore, these machine learning approaches have yielded a satisfactory result in this field.

The use of deep learning has been increased these days rapidly. Deep learning is a subfield of machine learning, which defines both higher level and lower level categories with higher accuracy. It works excellently with a massive amount of data. Deep learning techniques provide better efficiency and accuracy than ML as the techniques of deep learning tend to solve the problem end-to-end, whereas the ML techniques firstly need the problem statements to break down into different parts and finally combine their results [33]. Fig. 3 shows the working algorithm of both ML and DL, and their performance with the amount of data [34]. Looking at the significant advantages of deep learning,

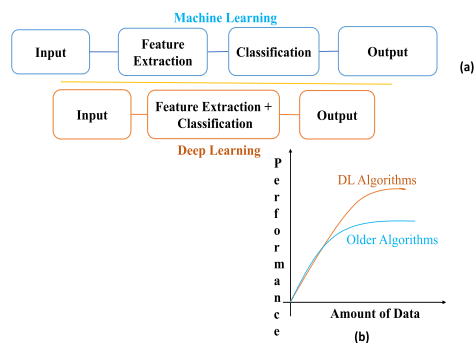


FIGURE 3. (a) DL/ML working algorithm, (b) performance of DL algorithms vs. older algorithms with the amount of data [34].

DL-based methods have been proposed and are being increasingly used for the MFDD too. Fig. 2 (b) shows the general flow diagram of the implementation of different techniques and their respective performance for bearing fault detection using vibration data.

The methodology implemented in this review paper is like this: the research works which implement deep learning-based approaches using the CWRU dataset, more specifically Drive-End (DE) defects data and normal-baseline data, for bearing fault detection and diagnosis are overviewed here. Except for [143], all the papers summarized here use DE (12k or 48k) faulty data and normal-baseline data as healthy bearing data in their deep learning models, whereas [143] uses all the records (drive-end, normal-baseline and fan-end data) of the CWRU bearing dataset.

This paper presents a detailed overview of the recent works and popular DL-based approaches for the MFDD. The outline of this paper is as follows: Section I introduces the machinery fault, conventional signal processing, and ML/DL approach used in bearing fault detection. Section II contains a brief introduction of the publicly available datasets used in bearing fault detection and diagnosis. Section III is the central part of this paper, which describes the popular DL-algorithms used in machinery fault detection and diagnosis. In this section, we present a detailed overview of the most recent works using DL-based methods like auto-encoders (AE), convolutional neural networks (CNN), deep belief networks (DBN), generative adversarial networks (GAN), recurrent neural networks (RNN) and long short-term memory (LSTM), and reinforcement learning (RL). Section IV is about some of the works which use transfer learning and domain adaptation-based DL methods. Since this paper mainly focuses on the works using the CWRU dataset, section V points out some of the weaknesses of the CWRU dataset. Section VI lists some of the limitations and challenges using DL models for MFDD, and the next section presents some recommendations from the authors. Finally, the paper is concluded in section VIII.

II. DATASET

Data is the fundamental unit and the foundation for all ML or DL architectures. Deep networks and DL algorithms

are influential ML structures that work most excellent if trained on vast amounts of data [35]. With the small training dataset, we get limited sample variations, and the network efficiency decreases. So, the quantity and the frequency of data availability has a vital role in DL applications. It can be said that an adequate amount of data for training in each class leads to a better prediction of that class. Generally, the more the data, the better the accuracy. The data structures with different labels can be better learned using abundant variations, and the model will recognize the invariant features compared to such differences [36]. Usually, a dataset can be categorized into two groups depending on its complexity: Simple dataset and a Complex dataset. A simple dataset, which is also called a good dataset in general terminology, is much easier to use, tends to allow for effective and easy data manipulation and calculations for some sort of meaningful statistical analysis [37]. Again, a well-labeled and balanced dataset which does not contain the aberrant data or missing values is also considered as a good one.

On the other hand, a complex set of data can be defined as a big dataset, which has a large volume, wide variety, large velocity and large veracity [38]. The real-world data are mostly highly skewed, and learning from these datasets is a challenging task for standard classification algorithms. When presented with complex imbalanced datasets, these algorithms fail to represent the distributive characteristics of the data accurately and resultantly provide unfavorable accuracies across the classes of the data. Though it is quite challenging to deal with the complex or imbalanced data, the imbalanced data are the characteristics of multiple real-world applications such as medical diagnosis, fraud detection, machinery fault detection, pricing catastrophe [39], [40]. The problem of imbalanced datasets has been approached from two main directions. The first approach is to preprocess the data by under-sampling the majority instances or oversampling minority instances. The latter is a cost-sensitive classification approach. In this approach, a specified learner is changed to incorporate a fluctuating penalty for each of the considered groups of samples [41].

Similarly, data obtained from multiple sources often mean messy data or leads to the data that follow a different internal logic or structure, which increases its complexity. The other factor that makes the data complex is its size. Generally, the bigger the data, the more complicated it is [42]. Traditional techniques simply could not identify all the features of imbalanced data because they assume the data is balanced, and the prediction is biased to the class where the sample number is relatively high. A better approach would be to develop a more generic organizing principle that can accommodate all possible types, rather than individual approaches that deal with specific types one by one [41].

The bearing is the broadly applied topic for machinery fault detection or anomaly detection. The reason may be the readily available public dataset [43]. Some of the accessible open-source bearing dataset used for the bearing fault detection and diagnosis are as follows:

- CWRU Dataset
- Paderborn University Dataset
- FEMTO Dataset
- MFPT Dataset
- IMS Dataset

A. CWRU DATASET

The CWRU dataset is a popular, open-source, and easily accessible dataset. The generated dataset is recorded and available on the CWRU website, which provides access to the bearing data for normal and faulty bearings. In this database, the data were collected for normal bearings, single-point drive-end (DE), and fan-end (FE) defects. The CWRU bearing dataset serves as the standard reference [23] and the fundamental dataset [2] to authenticate the performance of different ML and DL algorithms.

The bearing test rig arrangement used to obtain the CWRU dataset is shown in Fig. 1, which consists of a 2 hp Reliance electric induction motor, a torque transducer, a dynamometer, and control electronics, which is not shown in the figure. The test bearings support the motor shaft. Torque is applied to the shaft through a dynamometer and electronic control system. The faults were seeded on the REBs, the inner-race (IR) and outer-race (OR), and each faulty bearing was reinstalled on the test rig. Electro-discharge machining was used to introduce the single point faults to the test bearings with fault diameters of 7 mils, 14 mils, 21 mils, 28 mils, and 40 mils. One mil is equal to 0.001 inches. SKF bearings were used for the 7, 14, and 21 mils diameter faults, whereas for the 28 mils and 40 mils faults, NTN equivalent bearings were used. The depth of the fault was 0.011 inches for all the bearings except for inner-race faulty bearing of diameter 0.028 inches, an outer-race faulty bearing of diameter 0.040 inches, and a ball bearing fault of diameter 0.028 inches. The depth of the fault is 0.050 inches for both inner-race faulty bearing of diameter 0.028 inches and outer-race faulty bearing of diameter 0.040 inches. Moreover, for the ball bearing fault of 0.028 inches diameter, the depth of the fault is found to be 0.150 inches [4].

Acceleration data was measured at locations near to and far-off the motor bearings. The data is collected from multiple sensors placed at different places. Accelerometers, which were attached to the housing with magnetic bases and placed at the 12 o'clock position at both the DE and FE of the motor bearing, were used for collecting vibration data. Additionally, for some experiments, an accelerometer was attached to the motor supporting base plate too. Once the data was collected using a 16 channel DAT recorder, it was processed in a MATLAB environment, and all the data files were stored in MATLAB (.mat) format. Each file contains one or more of the recorded DE, FE, and base plate acceleration (BA) data. The sampling frequency of 12 kHz and 48 kHz were used for the collection of data. For the drive-end bearing experiments, data was collected at 12k and 48k samples per second. Fan-end data was collected at 12k samples per second. For the normal baseline, the data collection rate was 48k samples per second.

Vibration data was recorded for motor loads of 0 to 3 horsepower, with motor speeds of 1720 to 1797 rpm after the faulted bearings were reinstalled into the test motor. The dataset consists of data files for different torque loads applied by the dynamometer. However, ‘load’ is virtually meaningless in bearing fault detection and diagnosis because of a lack of a mechanism to convert the torque to a radial load-borne by the bearings. The main consequence of the motor load is on the shaft speed, which is declined by approximately 4% in the maximum load (3 hp) case, and this would have minimal effect on the diagnosable of the datasets [4], [23].

The dataset consists of 161 records, which are grouped into four classes: 48k normal-baseline, 48k drive-end fault, 12k drive-end fault, and 12k fan-end fault. Each group, again, consists of datasets for ball bearing (B) fault, inner-race fault, and outer-race faults. According to the fault location relative to the load zone, the outer race faults are further classified into three categories: ‘centered’ (fault in the 6.00 o’clock position), ‘orthogonal’ (3.00 o’clock), and ‘opposite’ (12.00 o’clock). For the names of the data files, the first letter represents fault position, the next three numbers signify fault diameters, and the last number denotes bearing loads. For example, the data file ‘B007_0’ contains the ball bearing fault data, which has the fault of diameter 0.007 inches operated under the motor load of 0 hp. Similarly, the data file ‘OR014@6_1’ contains the fault data of outer-race fault of diameter 0.014 inches when the load was centered (fault in 6 o’clock position) and operated under the motor load of 1 hp [4].

“Defective bearings produce vibration equal to the rotational speed of each component bearing frequencies. They relate notably to the rotation of the balls, the cage, and the passage of the balls on the inner and outer races” [1]. The bearing fault frequencies associated with the defective inner-race, outer-race, cage, and ball are as follows:

$$BPFI = \frac{nf_r}{2} \left(1 + \frac{d}{D} \cos\right) \quad (1)$$

$$BPFO = \frac{nf_r}{2} \left(1 - \frac{d}{D} \cos\right) \quad (2)$$

$$FTF = \frac{f_r}{2} \left(1 - \frac{d}{D} \cos\right) \quad (3)$$

$$BSF = \frac{Df_r}{2d} \left(1 - \left[\frac{d}{D} \cos\right]^2\right) \quad (4)$$

where, BPFI is inner-race ball pass frequency, BPFO is outer-race ball pass frequency, FTF is fundamental train frequency (cage speed), BSF is ball (roller) spin frequency, Furthermore, f_r is the shaft speed, n is the number of rolling elements, d is the rolling element diameter, D is the bearing pitch diameter, and α is the angle of the load from the radial plane. Fig. 1 (b) shows the cross-sectional view of REB, where the parameters Dd and α can be visualized. Furthermore, Table 1 shows the information of the bearings used in the CWRU dataset, whereas Table 2 tabulates different bearing fault frequencies. These frequencies are helpful when detecting the faults through signal processing methods.

TABLE 1. Bearing information¹.

Bearing Type	I.D. (inch)	O.D. (inch)	Thickness (inch)	B.D. (inch)	P.D. (inch)
6205-2RS JEM SKF (DE)	0.9843	2.0472	0.5906	0.3126	1.537
6203-2RS JEM SKF (FE)	0.6693	1.5748	0.4724	0.2656	1.122

TABLE 2. Information of bearing fault frequencies.

Bearing Type	Fault frequencies in Hz (multiple of shaft speed)			
	BPFI	BPFO	FTF	BSF
6205-2RS JEM SKF (DE bearing)	5.4152	3.5848	0.39828	4.7135
6203-2RS JEM SKF (FE bearing)	4.9469	3.0530	0.3817	3.9874

TABLE 3. Dataset length information.

Dataset	Length of Dataset			
	Maximum	Minimum	Standard Deviation	Mean
12k DE	IR007_3: 122917	B028_0: 120801	442.59	121895.96
48k DE	IR021_2: 491446	IR014_0: 63788	131280	411861.71
12k FE	B014_2: 122269	OR021@3_1: 120617	354.77	121306.51
Normal baseline	Normal_3: 485643	Normal_0: 243938	120468	424,636.75

Again, CWRU bearing dataset is long, varied, and complex. Each data file consists of data of different lengths, which is not an integer multiple of 2. The Table 3 shows the bearing length information for 4 classes of CWRU dataset.

As mentioned earlier, the CWRU bearing dataset is widely used and is taken as the standard reference for validating many ML and DL algorithms. The dataset does not contain the masking sources, which make it easier to use [22]. The review of the research employing DL algorithms using this dataset is presented in section III.

B. PADERBORN UNIVERSITY DATASET

This dataset [44] is also for bearing fault diagnosis and is provided by the KAT datacenter in Paderborn University. The essential components of the test rig are a drive motor, a torque measurement shaft, a test module and a load motor. Fig. 4 (b) shows the mechanical setup of the test rig for the Paderborn University dataset. The Paderborn university bearing dataset consists of the high-resolution vibration data, which are collected from experiments performed on six healthy bearings, and 26 damaged bearing sets. Out of the 26 damaged bearing sets, 12 were artificially damaged, and 14 were damaged using accelerated life tests to simulate real damage [45]. It provides the basis for the development, validation, and

¹I.D.: Inside Diameter O.D.: Outside Diameter B.D.: Ball Diameter P.D.: Pitch Diameter

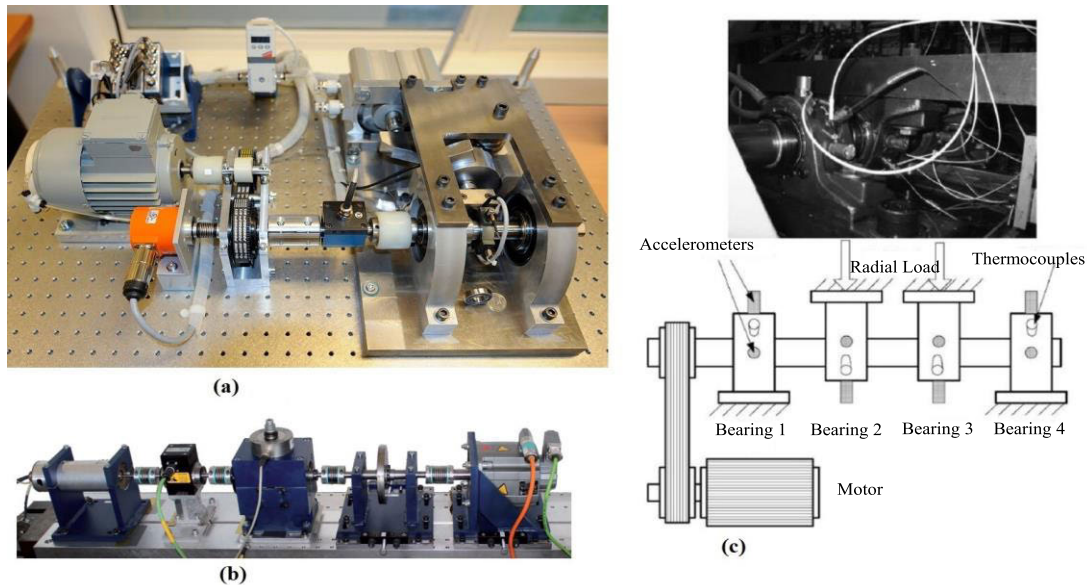


FIGURE 4. Overview of (a) PRONOSTIA experimental platform [53] (b) Mechanical setup of the test rig for Paderborn University dataset [44], and (c) experimental setup for IMS dataset [62].

training of the diagnostic algorithms for condition monitoring (CM) of rolling bearings. The test rig was operated under different operating conditions to ensure the robustness of CM methods at different operating conditions [44]. In total, experiments with 32 different bearing damages in ball bearings of type 6203 were performed. This dataset consists of the synchronously measured motor currents and vibration signals, which enable more accurate testing and implementation of the ML algorithms in practical applications, where the real defects are generated through aging and the gradual loss of lubrication [2]. References [45]–[48] use this dataset in their research.

C. FEMTO DATASET

FEMTO dataset [49], [50] is provided by FEMTO-ST Institute, Besancon, France. The real experiments on bearing's accelerated life tests, which are generated using an experimental platform called PRONOSTIA, are provided in this dataset. PRONOSTIA is an experimentation platform dedicated to test and validate bearings fault detection, diagnostic and prognostic approaches [51]. The main objective of PRONOSTIA is to provide real data related to accelerated degradation of bearings performed under constant and/or variable operating conditions, which are controlled online. The experimental platform, which is shown in Fig. 4 (a), allows to conduct bearings' degradations in only a few hours, and thus it is possible to get a significant number of experiments within a week. PRONOSTIA testbed is composed of three main parts: a rotating part, a degradation generation, and a measurement part. The operating conditions are characterized by two sensors: a rotating speed sensor and a force sensor. In the PRONOSTIA platform, the bearing's health monitoring is ensured online by gathering two types of signals: temperature and vibration signals with the

help of horizontal and vertical accelerometers. Furthermore, the vibration signals were recorded every 10 seconds with sampling frequency 25.6 kHz, which allows catching all the frequency spectrum of the bearing during its whole degradation. The vibration signals related to the degraded bearings are then compared to a nominal vibration signal of a non-degraded bearing or nominal bearing. Finally, the monitoring data provided by the sensors can be used for further processing in order to extract relevant features and continuously assess the health condition of the bearing. In total, 17 run-to-failure data under three different operating conditions were included in the FEMTO dataset, but the specific faulty mode of the failed bearing under each test is not declared [52], [53]. Some of the publications using this dataset are: [54]–[56].

D. MFPT DATASET

Another dataset for fault detection and diagnosis of REBs is the MFPT dataset [57], which is provided by the Society for Machinery Failure Prevention Technology. The goal of the Condition Based Maintenance (CBM) fault database is to provide various datasets of known good and faulted conditions for both bearings and gears. A bearing fault dataset has been provided to facilitate research into bearing analysis. The dataset comprises data from a bearing test rig (nominal bearing data, an outer race fault at various loads, and inner race fault at various loads), and three real-world faults [57]. Three real-world example files include an intermediate shaft bearing from a wind turbine, an oil pump shaft bearing from a wind turbine, and a real-world planet bearing fault. The data made available by the MFPT uses a NICE bearing. The defect seeding processes is not described. Three measurements with a load of 1201 N on the bearing are provided for the baseline condition and an outer race fault, as well as seven measurements for both the outer and inner race faults over a range

of 0–1334 N bearing load [58]. This dataset, thus, is freely distributed with example processing code with the hope that researchers and CBM practitioners will improve upon the techniques, and consequently, sophisticated CBM systems will be developed. The research which use this dataset for rolling bearing fault detection are [58]–[61].

E. IMS DATASET

This bearing dataset [62], [50] is provided by the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati, and can be downloaded from the Prognostic data repository, which is hosted by NASA. The Prognostics data repository is a collection of datasets that have been donated by various universities, agencies, or companies. The data repository focuses exclusively on prognostic datasets, i.e., datasets that can be used for the development of prognostic algorithms [50]. For the collection of the IMS bearing dataset, four test bearings were mounted on one shaft driven by an AC motor and coupled by rub belts. The experimental setup for this dataset is shown in Fig. 4 (c). The rotational speed was kept constant at 2000 rpm [43]. This database consists of three different datasets. In set one, 2 high precision accelerometers have been installed on each bearing, whereas in datasets two and three, only one accelerometer has been used. Each dataset is formatted in individual files, each containing a 1-second vibration signal snapshot, recorded at specific intervals. Each file consists of 20,480 points with a sampling rate set of 20 kHz. The file name indicates when the data was collected [62].

Some of the research works which employ this dataset are [63], [64].

Among all the datasets mentioned above, Case Western Reserve University (CWRU) dataset is the most widely used dataset [23] for the classification and detection of fault diagnostics of machinery bearings; thus, it serves as the fundamental dataset to validate the performance of the ML and DL algorithms [23]. This paper focuses on the works carried out for the bearing faults detection using the CWRU dataset with DL-based approaches. The following section describes the DL-based approaches used for bearing fault detection and diagnosis.

III. DEEP LEARNING-BASED APPROACHES FOR BEARING FAULT DIAGNOSIS

Deep learning is a branch of ML that is based on ANNs and inspired by the functionality of human brain cells called neurons. It is not mandatory to program everything explicitly in deep learning. It allows computational models, which are composed of multiple processing layers, to learn representations of data with multiple levels of abstraction. These methods have dramatically improved the state-of-the-art in speech recognition, visual object recognition, object detection, anomaly detection, and many other domains. DL discovers complicated structure in large datasets by using the back-propagation algorithm to indicate how a machine should change its internal parameters that are used to compute the

representation in each layer from the representation in the previous layer [65]. After analyzing some simple features in the early levels, the network sends this information to the following level, which takes this simple data, combines it into something more complex, and transfers it on the third level. This process lasts until the final layer, where the classification or output is obtained. Deep convolutional nets have brought about breakthroughs in processing images, video, speech, and audio. On working with both structured and unstructured data, DL networks perform automatic feature extraction without human involvement [66], [67].

Deep learning approaches can further be classified into supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. A supervised learning algorithm is based on training a model with the data that are correctly classified or labeled. In this algorithm, the model learns to map the input variables to the output variables. The goal is to approximate the mapping function so well that when the new data is introduced to the model, it can easily predict the output variables. Logistic regression, classification, and back-propagation neural networks are some examples of using supervised learning algorithms. However, most of the real-world data are found to be unlabeled. Labeling each datum is a complicated and expensive process and needs expert supervision. So, there is an algorithm that works for unsupervised data, and it is called unsupervised learning. This learning algorithm can learn and organize information without providing error signals to evaluate the potential solution [68]. Clustering, dimensionality reduction, association, anomaly detection, etc. use this algorithm.

In almost all real-world applications, the data is sparsely labeled. Semi-supervised learning lies in between supervised and unsupervised learning in which only the small set of data has corresponding labels. Data provided is with a mixture of labeled and unlabeled data. The semi-supervised learning algorithm is the one in which one can use both supervised learning techniques to make a best-guess prediction for the unlabeled data with the back-propagation algorithm and unsupervised learning technique to discover and learn the structure in the input variables [69], [70]. This algorithm overcomes the drawbacks of supervised and unsupervised learning and is being implemented these days widely. Another deep learning algorithm is reinforcement learning (RL), in which the algorithm is presented with examples that lack labels. This algorithm, which is the combination of ANNs and RL architecture, allows software-defined agents to learn the finest actions possible to achieve their goals. In this type of learning, the algorithm must make decisions, and these decisions relate to a consequence. It deals with delayed rewards and trains the structures to learn from the consequences of their own decision [71]. It is analogous to trial and error in human learning.

Again, common ML systems conventionally address isolated tasks. Transfer learning tries to change this by developing methods to transfer knowledge learned in one or more source tasks and use it to improve learning in

a related target task. Transfer learning is a machine learning with an additional source of information apart from the standard training data: knowledge from one or more related tasks [72]. The ML and DL algorithms work efficiently when the data is massive. However, having a vast amount of data is not always favorable in real-world applications. So, to overcome this issue, transfer learning approaches were proposed. It solves the problem of insufficient training data [73]. In this algorithm, the information flows in one direction only, from source task to the target task. Among all the transfer learning approaches, domain adaptation is one of the most popular methods [2]. Domain adaptation deals with scenarios in which a model trained on a source distribution is used in the context of a different target distribution. In general, domain adaptation uses labeled data in one or more source domains to solve new tasks in a target domain [74].

Because of the above mentioned and other numerous advantages of DL-based algorithms, they have been used in multiple studies. Image recognition, pattern recognition, object detection, text recognition, speech recognition, fault detection, and anomaly detection are some of the fields in which DL algorithms have performed well. In MFDD, both the supervised and unsupervised methods have been implemented. The DL models like auto-encoders, convolutional neural networks, deep belief networks, generative adversarial networks, recurrent neural network and long short-term memory, and reinforcement learning have shown their excellent performance with higher accuracy and reliability.

The following section describes the deep learning algorithms that have been applied for the task.

A. AUTO-ENCODERS AND MODIFIED AUTO-ENCODERS

Auto-encoders are widely adopted unsupervised learning procedure which is trained to copy its input to its output. Inside, it has a hidden layer h , which defines a code that is used to denote the input. The network can be regarded as containing two parts: an encoder function $h = f(x)$ and a decoder that produces a reconstruction $r = g(h)$ [67]. Fig. 5 shows the common structure of an auto-encoder where x denotes the input data vector $\{x_1, x_2, \dots, x_n\}$, h represents the vector in the hidden layer $\{h_1, h_2, \dots, h_m\}$, r is the outcomes of the output layer $\{r_1, r_2, \dots, r_n\}$, and simultaneously, it also signifies the reconstructed data vector. Here, W_1 and W_2 are the weight matrices that complete the connection between the adjacent layers. The ANN takes the mean-square-error between the original input and output as a loss function, which intends to imitate the input as the final output. After this, the structure is trained, and the decoder part is removed, and only the encoder part is kept. Hence, the output of the encoder is the feature representation that can be introduced in next-stage classifiers.

Many studies have been carried out for the bearing fault detection and diagnosis with an auto-encoder and modified auto-encoders. F. Jia *et al.* did the pioneering work in [75], where they utilized a 5-layer auto-encoder based deep neural network (DNN) and pre-trained it layer by layer with the unsupervised technique and further fine-tuned the DNN

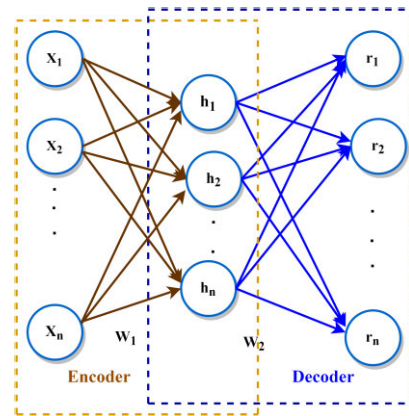


FIGURE 5. Architecture of an auto-encoder.

with a back-propagation algorithm for classification of the fast Fourier transformed faults and normal bearing data. The merit of this architecture is that it can adaptively obtain fault characteristics from the measured signals for many diagnosis problems, and the technique is good at establishing the non-linear mapping relationship between the different health conditions of machinery and the corresponding measured signals. The number of classes used was 10 and, the classification accuracy for the CWRU bearing data ranges from 99.68 to 99.95 percent, which is much better than the accuracy obtained by using the BPNN-based method. They implemented the PCA method to verify the ability of the proposed method.

Similarly, in [76], the authors have used a stacked auto-encoder to construct a DNN. Their algorithm involves two steps of first training the auto-encoder with the input set to get the feature vectors (say h) and then removing the output layer of the auto-encoder, which was trained before, and setting the feature vectors obtained, h , as the input set for the next auto-encoder. So, they stacked the auto-encoder for the classification of the faulty bearing data. In the beginning, they implemented several data preprocessing procedures where they divided the data from various groups into segments of 600 points with 80% overlap, and then FFT was applied on every section to get the amplitude, which was so small that it needed to be multiplied by a certain coefficient. The number of neurons in hidden layers was 800 and 400, respectively, and the number of fault classes was 6. They also used the PCA method to reduce the data dimensions for visualizing.

Some other works which employ auto-encoder in bearing fault diagnosis using the CWRU dataset are listed and described in Table 4. The references [77], [78] are some of the works carried out on MFDD with an auto-encoder using a dataset other than the CWRU dataset.

B. CONVOLUTIONAL NEURAL NETWORKS

After first proposed by LeCun for image processing [86], we can see how the CNNs have massively improved the performance in computer vision, object detection, natural language processing and speech recognition with the

TABLE 4. Some published works employing auto-encoders in MFDD².

A. R.	N. C.	Method	Classifier	Result	Remark
[79]	4	<ul style="list-style-type: none"> Stacked denoising auto-encoder (SDAE) based fault diagnosis method 3 Consecutive steps: train-test split of data, the establishment of deep hierarchical structure, validating the model The original CWRU bearing data are injected with a 15 dB random noise to mimic the noisy condition, and multiple operating condition datasets are used as test sets to examine its fault identification capability under speed and load variance. The input signal was reconstructed into high-order representations via self-learning with multiple hidden layers, where the destruction process was employed to improve the learning robustness 	SoftMax	For 12 dB SNR, accuracy = 91.67%, for 20 dB, accuracy = 99.83%	<ul style="list-style-type: none"> Noise-resilient model Reconstruction error shows little variation when the level of data destruction continues to increase
[80]	4	<ul style="list-style-type: none"> An integrated deep fault recognizer model based on the SDAE, which, inherently considers spatial and temporal correlations, is applied to both denoise random noises in the raw signals and represent fault features in fault pattern diagnosis for both bearing rolling fault and gearbox fault, and trained in a greedy layer-wise fashion SDAE with three hidden layers, 500 units in each layer 	SoftMax	99.25% accuracy for 0 hp load and 99.62% for 1 hp load	<ul style="list-style-type: none"> Noise-resilient Learning rate (L.R.) for pre-training and fine-tuning are 0.8 and 0.001, respectively.
[81]	10	<ul style="list-style-type: none"> A bearing fault diagnosis method based on fully connected winner-take-all auto-encoder in which the model explicitly imposes lifetime sparsity on the encoded features by keeping only k% most significant activations of each neuron across all the samples in a mini-batch. A soft voting method is applied to aggregate prediction results of signal segments sliced by a sliding window to increase the accuracy and stability 	Gath-Geva (GG)	83.4% accuracy for -4 dB SNR 97.2% for 0 dB SNR 99.97% for 10 dB SNR	<ul style="list-style-type: none"> Noise-resilient model The simulated dataset is generated by adding white Gaussian noise
[82]	12	<ul style="list-style-type: none"> Deep wavelet auto-encoder (DWAE) with ELM Firstly, the wavelet function is employed as the nonlinear activation function to design wavelet auto-encoder (WAE), which can effectively capture the signal characteristics. Secondly, a DWAE is constructed with multiple WAEs to enhance the unsupervised feature learning ability. Finally, ELM is adopted as the classifier 	Extreme Learning Machine (ELM)	Accuracy= 95.20%	<ul style="list-style-type: none"> Independence on manual feature extraction Reduced training time
[83]	7	<ul style="list-style-type: none"> A nonlinear projection is applied to achieve the compressed acquisition, which not only reduces the amount of measured data that contained all the information of faults but also realizes the automatic feature extraction in the transform domain For exploring the discrimination hidden in the acquired data, a four-layer stacked sparse auto-encoders-based DNN is established and performed with an unsupervised learning procedure followed by a supervised fine-tuning process 	Sigmoid	Accuracy= 97.47%	<ul style="list-style-type: none"> A compression ratio of 70% is applied only 30% of the original data are used to train the model
[84]	10	<ul style="list-style-type: none"> A semi-supervised learning approach for bearing anomaly detection using variational auto-encoder (VAE) based deep generative models, which allows for effective utilization of dataset when only a small subset of data have labels The latent-feature discriminative model performs better than other benchmark unsupervised models 	SoftMax	98.06% accuracy for N=2150 data segment	<ul style="list-style-type: none"> Data segmentation is done for data pre-processing
[85]	12	<ul style="list-style-type: none"> An ensemble deep auto-encoders (EDAe) is implemented with chains of auto-encoders with separate activation functions for unsupervised feature learning from the measured vibration signals Further, a combination policy is designed to ensure correct and stable diagnosis outcomes The input of the proposed method is always the raw vibration data (400-dimension). 	SoftMax	99.15% training accuracy 97.18% testing accuracy	<ul style="list-style-type: none"> No signal pre-processing is needed

increment in the production and memory of GPU. No doubt, CNNs are the most representative model of deep learning.

²A. R.: Article Reference N.C.: Number of Classes H.P.: Hyper-Parameters LReLU: Leaky ReLU

Thus, the use of CNNs has proliferated within computer vision. After the fruitful overview of AlexNet [87], the period of deep 2-D CNNs has started and immediately exchanged the outdated classification and acknowledgement approaches within a short period.

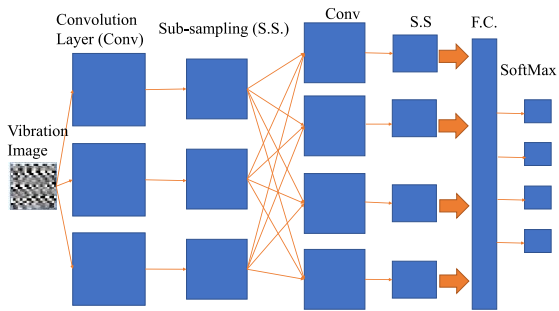


FIGURE 6. 1-D CNN architecture.

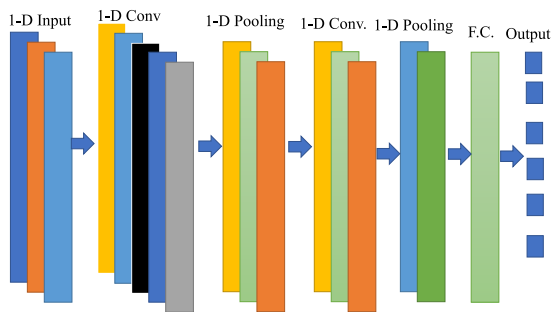


FIGURE 7. CNN architecture for vibration image used in [93].

Again, 1-D CNN is the modified version of 2-D CNN in which forward propagation (FP) and back-propagation (BP) are simple array operations rather than matrix operations, which makes them more efficient for specific applications in dealing with 1-D signals. Also, 1-D CNN is a relatively shallow structure, which makes them able to learn challenging tasks involving 1-D signal, and they are also suited for real-time applications [24]. Fig. 6 and Fig. 7 show the basic 2-D and 1-D CNN structures, respectively. The use of CNNs in bearing fault detection has been practiced in recent years. Some of the published works for detecting and diagnosing machinery bearing faults with a CNN architecture using the CWRU dataset have been described briefly in the following section.

A hierarchical adaptive deep convolutional neural network (ADCNN) is applied on the CWRU dataset in [88], which has two hierarchically organized components: a fault determination layer and a fault size evaluation layer. Here, the learning rate is changed dynamically so that a better trade-off is maintained between the training speed and accuracy. The ADCNN model in the first layer is based on the classical LeNet5 models proposed by LeCun. The number of layers used is 9, and the batch size is 100. The number of classes used is 4, among which one is the ‘healthy’ bearing class, and the other three are for fault bearings. The overall accuracy is 97.7%. SoftMax is used in the classification layer.

In [89], a CNN-based approach with multiple sensor fusion is proposed for fault diagnosis of rotating machinery in which both spatial and temporal raw data from the drive-end and fan-end are stacked by converting 1-D vibration data into 2-D input matrix and sent to CNN as input. Representative

features were learned directly from raw signals by training the CNN model, where no hand-crafted features were needed. Out of the total samples, 70% were used for training, 15% for validation, and 15% for testing. The average accuracy with two sensors was found to be 99.41%, whereas that with only one sensor is 98.35%. Minibatch stochastic gradient descent and dropout were utilized in the training process with increased efficiency and prevention of overfitting when the size of available data was small. The number of classes used was 9, and SoftMax is used in the classification layer.

An intelligent rotating machinery fault diagnosis based on DL using data augmentation technique is proposed in [90]. The authors use two augmentation methods: sample-based and dataset-based, and five augmentation techniques in general: additional Gaussian noise, masking noise, signal translation, amplitude shifting, and time stretching. Raw vibration signals are directly used as input without any preprocessing in this research where CNN structure and residual learning algorithm are further applied to train the network. A 1-D convolutional layer is first used with F_N filter kernels of F_L length window size. The raw data are processed for feature extraction initially and transformed into multiple feature maps. Next, residual blocks are stacked to learn high-level abstract representations. In each block, two convolutional layers are adopted for residual learning. For the convolutional layers in the proposed network, the zeros-padding operation is implemented to keep the feature map dimension from changing. They also use the batch normalization technique after each convolution layer to accelerate the training process and dropout for reducing the data overfitting. ReLU is used as an activation function, whereas SoftMax is used as a classifier. The number of labels used is 10. The best accuracy obtained on the CWRU dataset is 99.91% for 400 training samples.

Some other works employing CNN for MFDD using the CWRU bearing dataset are tabulated in Table 5. Again, [24], [91]–[93], and [94] are some of the published works which use CNN architecture for MFDD using a dataset other than the CWRU bearing dataset

C. DEEP BELIEF NETWORKS

Deep Belief Networks, first proposed in [100], are the network architectures consisting of many intermediate layers of Restricted Boltzmann Machine (RBM) in which each RBM layer communicates with the earlier and the following layers. Also, there is no intra-layer communication. The final layer is for the classification. In contrast to other models, each layer in deep belief networks learns the entire input. The deep architecture of DBN addresses the issues with Multi-Layer Perceptrons (MLPs) getting stuck at local optima by representing multiple features of input patterns hierarchically. Furthermore, they also benefit in optimizing the weights at each layer. This problem-solving approach involves making the optimal choice at each layer in the sequence, ultimately finding a global optimum.

TABLE 5. Some published works employing CNN in MFDD.

A.R.	N.C.	Method	H.P.	Result	Remark
[95]	10	<ul style="list-style-type: none"> Raw 1-D signals are divided into n equal parts, and each section is aligned as the row of the signal image (2-D) in sequence 2-D representation of 1-D signals is used to analyze and compress the power-quality event data using 2-D as the feature extractor Used 5*5 kernel in 2-D CNN whereas 25*1 kernels were used in 1-D CNN 	BN MaxPool ReLU SoftMax	99.95% accuracy for dataset A and 98.17% for dataset B	<ul style="list-style-type: none"> Compared the model with 1-D signal and 2-D representation as input CNN using 2-D representation is more efficient
[96]	10	<ul style="list-style-type: none"> A method of converting raw signals into 2-D images which can extract the features of the converted 2-D images through CNN Time-domain messages fill the pixels of the image A segmentation sample of length K_2 is randomly extracted from the time-domain signals. Then, a picture with a pixel point of $K \times K$ is obtained by processing the sample. The intercepted signal segments are normalized from 0 to 255-pixel value of the grey image. 	BN MaxPool ReLU Dropout SoftMax	100% accuracy in dataset A, B, C and D whereas 99.8% in dataset E	<ul style="list-style-type: none"> performance decreases rapidly when the workload changes
[97]	2/4/ 10	<ul style="list-style-type: none"> Fully automated framework and closed-loop system in which raw data is processed feed to the CNN model The output from the output layer with $1 \times N$ size (N is an integer for classification or equals to 1 for regression) from the CNN model is fed back to the machinery model for decision making. 	ReLU MaxPool SoftMax	100% classification in all the classes	<ul style="list-style-type: none"> Difficult to interpret the mechanism of the learned model due to the intrinsic complexity of the hidden layers
[98]	4	<ul style="list-style-type: none"> Vibration signals are used directly as input data for the deep structure of CNN The 1-D vibration signals are split into segments with the same length which are then processed for vibration image The amplitude of each sample in the vibration signal is normalized into the range [-1,1] which becomes the intensity of the corresponding pixel in the corresponding image 	ReLU SoftMax sub-sampling Layer	With 10 and 20 filters in 1 st and 2 nd layer resp., accuracy =96.75% With 30 and 60 resp., accuracy =100%	<ul style="list-style-type: none"> The number of pixels in the vibration image equals to the number of samples in the vibration signal
[59]	10	<ul style="list-style-type: none"> A sensor data-driven fault diagnosis method by fusing S-transform (ST) algorithm and CNN in which an ST layer is designed based on the ST algorithm In the ST layer, sensor data is automatically converted into 2-D time-frequency matrix, and then a new ST-CNN model is constructed, and the time-frequency coefficient matrices are inputted into the constructed ST-CNN model for the classification 	ReLU MaxPool	Average accuracy = 99.97%	<ul style="list-style-type: none"> Data samples are obtained by a sampling window from sensor data which are then randomly divided for training, testing validation and testing
[99]	10	<ul style="list-style-type: none"> Multi-scale deep CNN (MS-DCNN) is adopted to save extensive training time and network parameters required for CNN algorithms 1*1 convolution is used for dimensionality reduction and increasing the depth and width of the network. The convolution kernel of different sizes is used to extract different-scale features in parallel The number of parameters to be determined is only 52,172 in MS-DCNN, which is significantly fewer than that of 1-D CNN (171,606) and 2-D CNN (213,206) 	SoftMax MaxPool	Accuracy= 99.27%	<ul style="list-style-type: none"> 9 Layers network Ran more smoothly in the training process compared to 1-D CNN and 2-D CNN Sliding stride technique is used for data augmentation Used the concept of "Inception" module

These greedy learning algorithms start from the bottom-most layer and move up, finetuning the generative weights. Here, the learning takes place on a layer-by-layer basis training layers of the DBN one at a time. Consequently, each layer obtains a different type of data. Except for the first and last layers, each layer in a DBN has a dual role, i.e., it serves as the hidden layer to the earlier nodes, and as the

input or a visible layer to the following nodes. It can be called a network built of single-layer-networks. These networks can solve many problems like low velocity and the overfitting phenomenon in deep learning. An architecture of DBN is shown in Fig. 8. Besides being used widely in the field of recognizing, clustering, and generating images, video sequences and motion-capture data, they have also been used

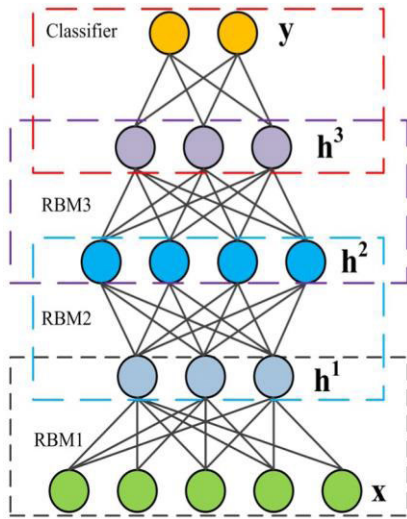


FIGURE 8. Architecture of DBN [106].

in detecting and diagnosing rotary machinery fault. Some of the published works in MFDD using the CWRU dataset are described below.

A rolling bearing fault diagnosis using adaptive DBN with a dual-tree complex wavelet packet (DTCWPT) is proposed in [101]. Here, an adaptive DBN is constructed to improve the convergence rate and identification accuracy with multiple stacked adaptive restricted Boltzmann machines (RBMs). The authors used three-layer DTCWPT to preprocess the raw vibration signals to improve fault features. Three experiments were conducted. In experiment 1, first, DTCWPT, EMD, EEMD, and WPT are used to preprocess the raw vibration signal, respectively. Then, nine feature parameters are, respectively, extracted from each frequency-band signal of DTCWPT and WPT and effective IMFs of EMD and EEMD. Finally, all the normalized feature parameters are directly fed into the adaptive DBN. The decomposition level of DTCWPT is set as 3, and the architecture of the adaptive DBN is 72-400-250-100-16. The initial learning rate and momentum are selected as 0.03 and 0.86, respectively. The increasing factor, decreasing factor and threshold are 1.25, 0.6, and 0.2, respectively. The pre-training of each adaptive RBM is completed using 200 iterations. For DBN and EEMD, the ensemble number is chosen as 100, and the added noise amplitude is 0.005 times the standard deviation of the analyzed signal. The first 10 IMFs containing almost all valid information are selected, and the DBN architecture is 90-500-300-100-16. In experiment 2, DTCWPT is applied to preprocess the raw vibration signal, and then, the nine features are extracted from each frequency-band signal of DTCWPT. Finally, without any manual feature selection, all the normalized feature parameters are directly fed into the adaptive DBN, BP, general regression neural network (GRNN), and SVM, respectively. The testing accuracy using the proposed method is 94.37%. In Experiment 3, sixteen sensitive features are manually selected from the original feature set, including

eight kurtosis values and eight wavelet packet energy values. Then, the selected 16-dimensional features are fed into the adaptive DBN, BP, GRNN, and SVM, respectively. The architecture of the adaptive DBN is 16-80-45-25-16.

Similarly, [102] talks about DBN and Dempster-Shafer (DS) evidence theory for bearing fault diagnosis. Here, researchers installed multiple sensors where the features from the signals from each sensor are extracted using DBN and classified with SoftMax. Then, the predicted result from the SoftMax is fused by DS evidence theory to generate the final prediction of bearing health status. In experiments, vibration signals from DE and FE sensors, which are of four bearing health conditions, are considered. To obtain enough data for the training and testing model, each single vibration data file is split into segments that have equal length. They used the FFT algorithm as the preprocessor for the raw time-domain data. For the feature learning and feature classification, two DBNs are designed to learn fault features from the bearing signals. Since signals from both sensors are vibration signals, they develop both DBN models with the same configuration. The training processes of DBNs are carried out step by step. After the feature learning process, two feature sets, each from FE and DE sensors is obtained. Two SoftMax classifiers are employed to classify these two feature sets. The classification results of two SoftMax classifiers are combined to achieve a better result. DS evidence theory is exploited to fuse the decisions and generate the final prediction. The proposed method achieved 100% accuracy. Also, the proposed system does not require to be retrained when the working condition (load) changes. They also evaluated the model performance under different noisy conditions, in which the worst-case accuracy was 98.5%.

Some of the other remarkable works regarding MFD which employ DBN using the CWRU dataset are listed in Table 6, and the references of the works using other datasets: [103]–[108].

D. GENERATIVE ADVERSARIAL NETWORKS

Generative adversarial networks (GANs) are deep neural net structures that are comprised of two nets, competing one against the other. They belong to the set of generative models, which means they can generate or create new content. They typically work with random input and produce something on output. They were first introduced in [114] by Ian Goodfellow and other researchers. Here, the discriminator, D and the generator, G play the following two-player mini-max game [114] with value function

$$Min_G Max_D V(D, G) = E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \quad (5)$$

For the provided training dataset, the generator intends to create samples that have an equal probability distribution as the real data or training dataset. The discriminator belongs to the conventional binary classifier and is primarily responsible for dual tasks. In the first one, it is essential to determine

TABLE 6. Some published works employing DBN in MFDD.

A.R.	N.C.	Method	H.P.	Result	Remark
[109]	10	<ul style="list-style-type: none"> A method based on Hilbert envelope spectrum and DBN in which the vibration signals under different test conditions are resampled and the whole Hilbert envelope spectrum of those resampled signals are directly used as eigenvectors Next, the DBN classifier model is established to recognize the fault bearing 	Sigmoid	For 100 hidden units accuracy = 99.55%	<ul style="list-style-type: none"> The more the number of hidden units, better the learning of high complexity non-linear relationship between the feature vector and the bearing fault type
[110]	10	<ul style="list-style-type: none"> An optimization DBN for REB fault diagnosis in which stochastic gradient descent is used to efficiently fine-tune all the connection weights after the pre-training of RBMs based on the energy functions Particle swarm optimization (PSO) technique is then used to decide the optimal structure of the trained DBN, and the optimization DBN is designed. The optimal architecture of the DBN given by PSO is 18–57–38–25–10. 	LR=0.23 Momentum =0.86	Average accuracy of dataset A= 89.2%, and that of dataset B= 90.7%	<ul style="list-style-type: none"> It is necessary to extract time-domain features phase-by-phase from all the data, which not only narrows the length of data but also gets statistical properties in each phase
[111]	4	<ul style="list-style-type: none"> An encoder is constructed using DBN which can minimize the energy between the output and input vibration signal DBN is constructed by stacking RBMs layer by layer which ensures that DBN can be practiced through the CD algorithm 	Sigmoid	Accuracy= 96.67%	<ul style="list-style-type: none"> The computation time of DBN was longer than HHT and FFT, but shorter than Wavelet-3
[112]	10	<ul style="list-style-type: none"> An improved ensemble method in which a series of DBNs with different hyper-parameters are constructed and trained by the Contrastive-Divergence (CD-k) algorithm. Then, it is used to acquire the weights matrix for each DBN which finally uses these matrices for the classification 	Sigmoid up to DBN5 ReLU from DBN6-10	Accuracy= 84.2%	<ul style="list-style-type: none"> No Preprocessing Normalizing the original vibration data
[113]	10	<ul style="list-style-type: none"> The construction of Hierarchical Deep Network (HDN) with DBN in which the two-layer network enables HDN to put forward fewer requirements of features and the classifier with the wavelet packet energy feature The first layer identifies the fault types whereas the other layer recognizes the fault severity ranking 	Sigmoid	Accuracy= 99.03%	<ul style="list-style-type: none"> Three stacked RBM After the train-test split, samples were further divided into small mini-batches

whether the input arrives from the real data distribution or the generator. In the second task, the discriminator directs the generator over the back-propagation gradient to produce a more realistic sample, which is the only method for the generator to improve its performance of generating fake samples. During this two-player min-max game, the generator takes in impulsive noise as input and outputs a fake sample, which should be maximized by the discriminator as the probability of the result from the original training set. During the training period, the discriminator processes the sample of the training set as an input for half of the time duration and takes in the fake sample G_{Sample} generated by the generator as an input for the other half.

The discriminator, in this process, is trained to optimize or maximize the distance between groups, and to differentiate between the actual image from the training set and the fake sample generated from the generator. Thus, the generator should be capable of making the created probability distribution and the actual data distribution as near as possible so that the discriminator cannot decide which one is real and which one is a fake sample. Hence, in this adversarial process, the generator’s capability to study the actual data distribution becomes stronger and stronger, and the discriminators’ feature learning and discriminative capacity also become stronger and stronger. Eventually, the training will reach Nash equilibrium, where the Discriminator will be unable to

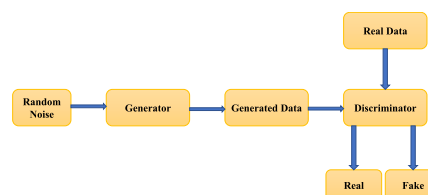


FIGURE 9. Architecture of GAN.

differentiate between two distributions, i.e., $D(x) = 1/2$. Since this equilibrium is very tough to find, there are many research papers [115], [116] for cracking this issue [114] [117]. Fig. 9 shows an overview of GAN architecture. Like in every other field, the use of GAN is increasing in bearing fault detection also. Several works have been published using GAN for the machinery bearing fault detection and classification using the CWRU dataset. Some of them are described briefly in the following section.

Being inspired by the Wasserstein distance of optimal transport, Cheng *et al.* proposed a Wasserstein distance-based deep transfer learning method [13]. The authors created a deep transfer learning (DTL) algorithm to perform learning in the target-domain by leveraging information from an appropriate source-domain to determine the domain feature illustrations to minimize the distributions between the source and target domains through adversarial training. CNN based

feature extractor generates those illustrations. They applied Wasserstein-1 distance to measure the loss of domain discriminator because Wasserstein-1 distance is the optimum transport plan with the bottommost transport cost. In their research, they inspected four types of bearing condition data acquired by CWRU. Simple data preprocessing techniques like dividing the samples equally in source-domain and target-domain for their modification to a stationary process, implementing FFT to compute the power spectrum in frequency-domain of every sample, and clipping the left side of the power spectrum calculated by FFT as the input of WD-DTL are applied to the bearing datasets. They proposed three transfer scenarios, including two unsupervised and one supervised scenario. Unsupervised transfer scenarios are for transfer between motor speeds (US-Speed) and transfer between datasets at two sensor locations (US-Location), whereas a supervised transfer scenario is for transfer between dataset at two sensor locations (S-Locations). The proposed WD-DTL model achieved the best transfer accuracies with a 95.75% average score.

In [118], the authors proposed an unsupervised fault diagnosis of rolling bearings using DNN based on GANs in which they designed categorical adversarial auto-encoder (CatAAE). By adding a classifier on the latent layer of adversarial AE, they introduced a new model named CatAAE for unsupervised clustering. In this work, an auto-encoder is trained through an adversarial training process and carries out an earlier distribution on the latent coding space. The raw data of rolling bearings is in the time-domain. To keep the consistency of vibration signals under the same faulty condition, multiple features, including time-domain and frequency-domain, are extracted to describe the health status through FFT and statistical methods completely. Next, a classifier is trained to differentiate the prior distribution from the fake distribution by harmonizing shared data between instances and their forecasted categorical class distribution. Although the encoder can extract features from the input data, the proposed method still relies on computational time and frequency-domain features of raw data. The number of labels used in this study is 10, and different clustering indicators were obtained when the model was performed under different SNR and across the varying loads. Mixed time-frequency features are employed in the procedure to get better robustness under different environments. For the original signal, the clustering indicator was found to be 98.36, and for 0 dB SNR, the clustering indicator was found to be 94.35. Also, across the different loads, different clustering indicators were obtained. The average was found to be 90.15. SoftMax was used in the last layer as the classifier. For the data preprocessing, multiple features, including time-domain and frequency-domain, were extracted.

In [119], F. Zhou *et al.* addressed the problem of misclassification when the data is unbalanced and proposed a fault diagnosis algorithm based on global optimization GAN in which they combine a robust feature extraction capabilities of DNN with data generation capabilities of GAN for solving

this issue. In this paper, the authors designed a new generator and discriminator of GAN to generate more discriminant fault samples using a scheme of global optimization. The generator is designed to generate those fault features extracted from a few fault samples via auto-encoder instead of fault data sample. The training of the generator is guided by fault feature and fault diagnosis error instead of the statistical coincidence of traditional GAN. The discriminator is designed to filter the unqualified generated samples in the sense that qualified samples are helpful for more accurate fault diagnosis. The adversarial training mechanism is arranged by alternately optimizing the generator and the 2-level discriminator. Here, the network parameters of the generator, G, are fixed, and the authenticity discriminator, D2, is optimized. Then, a DNN-based fault diagnosis discriminator, D1, is established by using the original unbalance dataset. The DNN fault diagnosis model can be updated by the generated samples to get a smaller fault diagnosis error. Finally, the generator, G, is optimized for making the generated samples more qualified for fault diagnosis. The global optimization is accomplished once the discriminator D1, D2, and generator G reach a Nash-equilibrium, which increases the model's efficiency. For preprocessing data, a sliding window is used. The diagnostic accuracy of this method for the three kinds of faults: inner-race, roller, outer-race, for 10:1 unbalanced accuracy ratio is 94.58%, 96.85%, and 93.28%, which is far more than that of using SAE.

Some of the other works which implement GAN for MFDD using the CWRU dataset are as tabulated in Table 7.

Moreover, some of the references for the published article using other datasets than the CWRU dataset are [117], [125], [126].

E. RECURRENT NEURAL NETWORKS

Recurrent neural networks, which are abbreviated as RNNs, are a class of neural networks that permit earlier outputs to be processed as inputs while having hidden states. They remember the past, and their decisions are influenced by what they have learned from the previous state. Thus, RNN recurrently processes the input data [127]. Since RNN has the typical gradient vanishing or gradient exploding issue caused by its nature, the LSTM networks were born to solve the problem. LSTM, an abbreviation of Long Short-Term Memory, is the widespread network of these days. They are also called the cells network, and these cells take the input from the previous state and the current state. Not only in natural language processing, text recognition, or speech recognition, but RNNs are also widely used in machinery fault detection and diagnosis. Fig. 10 shows the basic architecture of the LSTM network.

In [128], an improved bearing fault diagnosis method based on the combined unified structure of 1-D CNN and LSTM was proposed whose input is the raw sampling signal without any pre-processing or traditional feature extraction. Here, the entire architecture is composed of a 1-D CNN layer, a max-pooling layer, an LSTM layer, and a SoftMax layer as the last

TABLE 7. Some published works employing GAN in MFDD.

A.R.	N.C.	Method	H.P.	Result	Remark
[120]	4	<ul style="list-style-type: none"> GAN based MFD approach combining well-designed feature extractor for imbalanced data the three sub-network, encoder-decoder-encoder, is employed based on DCGAN DCGAN is applied without the tanh activation layer 	Sigmoid Adam	Avg. acc. > 99%	<ul style="list-style-type: none"> The model is trained only on normal samples.
[121]	4	<ul style="list-style-type: none"> A deep transfer model in which the WDMAN model exploits complex feature space structures to enable the transfer of different data distributions based on multiple domain critic networks by learning the shared feature representation by minimizing the Wasserstein distance between the source domain and target domain distribution in an adversarial training way by taking advantage of CNN and GAN 	TanH for FC-1 and SoftMax for FC-2	Avg. acc. of 99.39%	<ul style="list-style-type: none"> The t-SNE technology is used to visualize the learned domain invariant feature and investigate the transferability
[122]	10	<ul style="list-style-type: none"> For solving the problem of instability and overfitting, an MFD algorithm based on suitable normalization and semi-supervised GAN is proposed with a 1-D representation of vibration signal as input the input shape of D is 2048*1 and output of G is also 2048*1 In supervised learning, training signals with labels are fed into D, after some convolution, normalization, and activation, the D outputs the probabilities of the signal samples. Then the loss is computed, and D's parameters are optimized In the process of unsupervised learning, random vector z is fed into G, which outputs the fake data, which are taken in D for a probability and loss of batch is computed and then G is optimized by fixing the D. Then, unlabeled training signals are fed into D which then gives the corresponding score 	ReLU for the G, and tanh in the last layer LReLU for the D	99.93% acc.	<ul style="list-style-type: none"> SS-GAN reduces reliance on observational datasets Delete all pooling layers of [49] and replace BN with SN
[124]	6	<ul style="list-style-type: none"> Imbalanced fault diagnosis method in which GAN is used to generate the synthetic sample for minority fault class FFT is implemented to preprocess the original vibration signal, and the spectrum data is used as input of GAN Finally, it puts the synthetic samples generated by GAN into the training set and builds a stacked denoising auto-encoder (SDAE) model for fault diagnosis which directly extracts deep feature from original spectrum data 	Sigmoid SoftMax	Test acc.= 96% F1_score= 97.96%	<ul style="list-style-type: none"> Performance is verified with various imbalance ratios

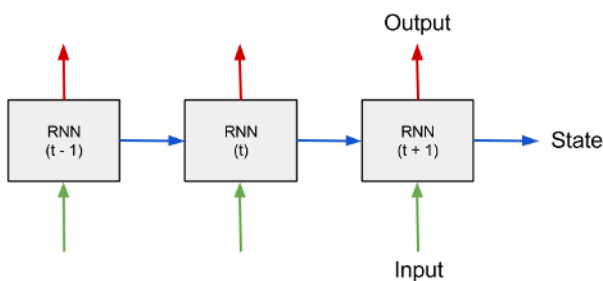


FIGURE 10. Architecture of RNN.

layer classifier. First, a part of raw bearing data is used as the training dataset for the model, and when the number of iterations reaches a specific value, the simulation ends. Then, the rest of the signal data was input in the trained model as the testing dataset to verify the effectiveness of the model. The best testing accuracy ranges up to 99.6%.

Fault diagnosis of rolling bearings with RNN-based auto-encoders is proposed in [129] in which a Gated Recurrent Unit (GRU) based denoising auto-encoder is used to predict the multiple vibration value of REB of the following period are forecasted from the preceding period. These GRU-based non-linear predictive denoising auto-encoders

(GRU-NP-DAEs) are trained with a strong generalization ability for each different fault pattern. The trained GRU-NP-DAEs are received a multiple-input data, and the fault diagnosis result is determined by the relevant GRU-NP-DAE that produces the minimum reconstruction error between the delay of the input and the model output. The concept of classification accuracy is adopted to evaluate the feasibility of the proposed method for health condition detection and fault type classification. During the supervised learning process, the data destruction method is selected, and the length loss method is proposed to enhance the robustness of models. The classification result of GRU-NP-DAE is not lower than 96%, and even SNR decreases to 1 dB.

Some other works on MFDD employing RNN or LSTM using the CWRU bearing dataset are tabulated in Table 8.

Moreover, [132]–[135] are some of the research carried out for MFD using a dataset other than the CWRU dataset.

F. REINFORCEMENT LEARNING

Reinforcement learning is an area of machine learning that deals with delayed reward and trains the systems to learn from the consequences of their own choices. It resolves a kind of problem where decision making is consecutive, and the goal

TABLE 8. Some published work employing RNN in MFDD.

A.R.	N.C.	Method	H.P.	Result	Remark
[130]	12	<ul style="list-style-type: none"> Improved deep recurrent neural network (DRNN) is proposed with stacked recurrent hidden layers, and LSTM units where MSE function is used as the loss function and SGD is used as the optimizer An adaptive learning rate is adopted to improve training performance. 	SGD	Average testing accuracy = 96.53%	<ul style="list-style-type: none"> Frequency spectrum sequences are used as inputs to reduce the input size
[131]	10	<ul style="list-style-type: none"> A stacked residual dilated CNN (SRDCNN) method which is combined of the dilated convolution, the input gate structure of LSTM and the ResNet SRDCNN is used to increase the receptive field of convolution kernel exponentially and extract features from the sample with more points, alleviating the influence of randomness. The input gate structure of LSTM could effectively remove noise and control the entry of information contained the input sample the ResNet is introduced to overcome the problem of vanishing gradients which improves the performance 	SoftMax	99.7%	<ul style="list-style-type: none"> For one example, the length is reduced by half after getting through residual dilated convolution layer Dilated convolution helps in accepting a more extended sample with a large amount of redundant information

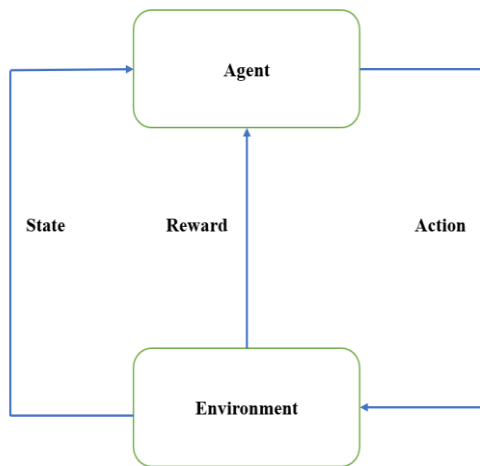


FIGURE 11. RL architecture.

is long-term. Employed by several machines and software to determine the best probable path or action it should take in a situation, reinforcement learning is all about taking favorable action to augment reward in that situation. With the help of simple reward feedback, the machine learns which action is the best to perform, and this is known as the reinforcement signal. Reinforcement learning and supervised learning are different in a way that the training data has the answer key with it in the supervised learning due to which the model is trained itself with the accurate answers, whereas in the reinforcement learning as there is no answer, the reinforcement agent chooses what to do to complete the given task. In the absence of a training dataset, it is bound to learn from its own experience [136], [137]. An overview of RL is shown in Fig. 11. From teaching the machines to play Atari games to its application in the manufacturing industry, reinforcement learning is widely used these days. This reinforcement learning is being applied in the machinery bearing fault detection and diagnosis also.

In [138], Wang *et al.* proposed a neural network architecture automatic search method based on reinforcement learning for fault diagnosis of rolling bearings. The framework of the proposed method contains two components: a controller model and child models. The controller is a recurrent neural network and generates a series of actions; each action specifies a design choice to construct the child models for fault diagnosis. Then, the controller parameters are updated using the policy gradient method of reinforcement learning by maximizing the accuracy of the child models. They used the CWRU dataset for the experimental verification in which they used 12K sampling frequency data. There are 12 different working states in which one was the healthy state, and the remaining 11 were fault states. The vibration data measured at 1750 rpm in each working state were divided into 400 samples, where each sample contains 300 data points. The first 350 samples are used to construct training samples, and the remaining 50 are test samples.

Other important details are shown in Table 9.

The researchers used the RL system with kurtosis as an index in [139] for MFD. The dataset used for carrying out this research was other than the CWRU dataset.

IV. DEEP TRANSFER LEARNING AND DOMAIN ADAPTATION METHOD

The ML and DL algorithms work efficiently when the data is massive. However, having a massive amount of data is not always favorable in real-world applications. So, to overcome this issue, transfer learning approaches were proposed. Transfer learning is an essential tool in machine learning, which solves the problem of insufficient training data [73]. Cross-domain adaptation, one of the most popular methods among the transfer learning approaches, is implemented with situations in which a model trained on a source distribution is used in the context of a different, but related target distribution. In general, domain adaptation uses labeled data in one or more source domains to solve new tasks in

TABLE 9. Other necessary details of [138].

A. R.	N.C.	Method	Classifier	Result	Remark
[138]	12	<ul style="list-style-type: none"> The framework of the proposed method contains two components: a controller model and child models The controller can be regarded as an agent, the controller generates a set of super parameters as the agent’s action to construct the CNN architecture, and uses the accuracy of the child models as the reward. That can be considered as the Markov Decision Process, and the CNN architecture can be found by maximizing the reward ($J(\theta)$). Since the reward is not differentiable, they updated the parameters with the policy gradient method. 	SoftMax	98.47%	<ul style="list-style-type: none"> The search scope is too broad, and there are too many feasible choices to construct child models. Although the random actions are generated to avoid local optimal solutions, the result is still easy to be trapped in the local optimal solution.

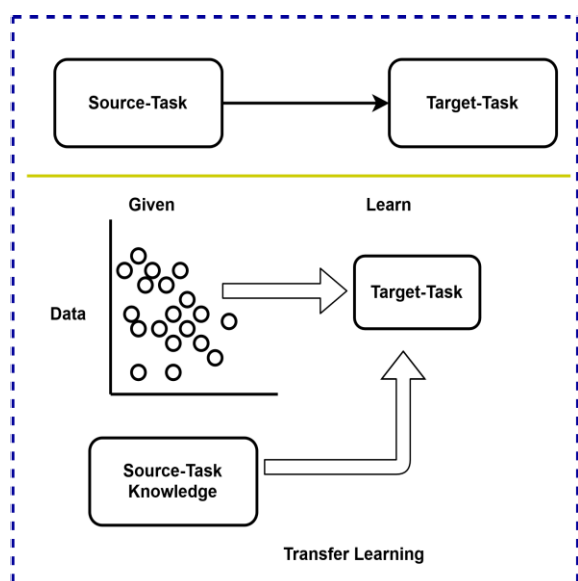


FIGURE 12. Overview of transfer learning algorithm.

a target domain [74]. By discovering the domain-invariant features, domain adaptation creates a knowledge adaptation from the source domain to the target domain, which mitigates the distribution discrepancy between two domains [140]. Fig. 12 shows an overview of the transfer learning method. Recently, the combinations of deep transfer learning with domain adaptation are being widely practiced. Some of the deep transfer learning and cross-domain adaptation strategies used in REB fault detection and diagnosis are briefly summarized and overviewed in this section.

In [141], the authors proposed a novel cross-domain fault diagnosis method based on deep GAN in which artificially generating the fake samples for domain adaptation, and their model provided the reliable cross-domain diagnosis results when testing data are not available for training. In this paper, the authors used the maximum mean discrepancy (MMD) to measure the distribution discrepancy. Here, they designed a two-stage model where the first stage aims at artificially generating different classes of fault data in the target domain,

and the cross-domain classifier is trained in the second stage. In first-stage, three modules are adopted, i.e., feature extractor, generator, and a primary classifier. First, a feature extractor is used to process the data sample, which is the frequency-domain signal of the collected vibration data, from which the high-level representation of the input can be obtained. Next, a primary classifier is used for diagnosis, which ensures the extraction of discriminative features under source supervision. Again, the generators are used for data generation, and each generator is trained explicitly for generating data of each fault class, respectively. After training in Stage 1, the parameters of the feature extractor are fixed and used for obtaining high-level representations of input data. The cross-domain classifier module takes the extracted representations as inputs and outputs the final diagnosis results. Specifically, in the implementation, the zero-padding operation is used to keep the feature map dimension from changing. The dropout technique is preferred with a rate of 0.5 to avoid overfitting. The sigmoid and rectified linear unit activation functions are generally used in the network.

In [142], the authors proposed a model for overcoming the problem of the requirement of a large amount of training data for bearing fault detection and created a model that works efficiently with limited training data. They proposed a based few-shot learning approach, which is based on a siamese neural network model based on deep CNN with wide first-layer kernels (WDCNN). The model works by exploiting sample pairs of the same or different categories, measuring the distance of two WDCNN twins feature vector outputs in terms of whether their outputs are considered quite similar versus dissimilar. In a small training set with 90 training samples, the method can achieve an accuracy of 92.56%. They used the raw data directly without any preprocessing.

Similarly, [143] talks about an automatic classification based on deep learning in which faulty signals are clustered without human knowledge. Here, a dataset, in which each sample is given a random label, is configured after extracting the features of vibration signals from the frequency-domain and used to train DNN to obtain an initial classification. The classification results are assessed by testing the sub-signals

extracted from the raw data, and the sample labels are modified according to the evaluation result. The modified dataset is used to train the DNN a second time. Samples with characteristic faults are clustered in various classes after iterating the DNN training and testing. Here, time-domain signals are converted to the frequency-domain, and the feature extraction and dimensionality reduction are made. A sample dataset is built, giving a random label to each sample. Then, a DNN is constructed and trained. They used the PCA technique to validate their model.

In [144], a domain adaptation method for MFDD based on deep learning in which adversarial training is introduced for marginal domain fusion, and unsupervised parallel data are explored to achieve conditional distribution alignments concerning different machine health conditions. Here, FFT is firstly applied to the temporal signals to obtain the frequency-domain information, which is then fed into the network as inputs. The source-domain supervised data are explored to extract the discriminative features for different machine health conditions. The marginal data distributions of the source and target domains are supposed to be drawn into the same subspace to generalize the fault diagnostic knowledge in the presence of a significant distribution discrepancy. The additional unsupervised parallel data are utilized to align the domain representation structures, and sensors collect the supervised training data and the unsupervised testing data at different places on the machine. Four categories of the input data are considered, i.e., labeled source-domain data, unlabeled target-domain testing data, and additional unlabeled parallel data in both the source and target domains. The number of classes for CWRU bearing data was 10. SoftMax was used for a classification layer. The authors claim that when the unsupervised parallel data cover a more extensive range of classes, higher accuracies can be obtained. Generally, over 80% of accuracies can be achieved if the parallel data contain more than 3 classes. When only a few parallel data are available, higher than 50% accuracies can still be obtained.

Similarly, a transfer CNN for fault diagnosis based on ResNet-50 is proposed for fault diagnosis of bearings in [48] in which they used transfer CNN (TCNN) with the depth of 51 convolutional layers. TCNN applies ResNet-50, which was trained on ImageNet as a feature extractor. First, for the input of ResNet-50, a signal-to-image method was developed for converting time-domain fault signals to RGB images format. Then, a new structure of TCNN(ResNet-50) is designed. Finally, the proposed model was tested on three datasets, including bearing damage dataset provided by KAT data-center, motor bearing dataset provided by CWRU, and self-priming centrifugal pump dataset. The number of classes used for the CWRU dataset was 10, and they used SoftMax in the classification layer. ReLU was used as an activation function, and BN was adopted after each convolution layer. The prediction accuracies of TCNN(ResNet-50) for the CWRU dataset was found to be 99.99%. They compared their model with VGG-16, VGG-19 and Inception-v3,

in which their model performed more efficiently than those all.

Some of the other works using transfer learning or domain adaptation strategy using the CWRU bearing dataset for machinery fault detection are tabulated in Table 10. Some of the works employing DL-based transfer learning and domain-adaptation approaches using the dataset other than the CWRU dataset are [145]–[147].

V. WEAKNESSES OF CWRU BEARING DATASET

CWRU bearing dataset is long, varied, and complex. The data is collected from multiple sensors placed at different places. Each data file consists of data of different lengths, which is not an integer multiple of 2. Again, one can analyze that most of the datasets are exposed to non-classical features of fault recognition; relatively few of the records gave classical characteristics for the specified bearing fault type. Again, many of the datasets exhibit non-stationary characteristics. All the faults cannot be recognized correctly. The data records ranged from very easily diagnosable to not diagnosable [23]. The variance is high because the bands of measurements in the same file are dissimilar. Additionally, not all the frequency components are regular; some are occasionally large or small [143]. So, using preprocessors like FFT or other signal processing techniques for feature extraction can be tough. The main challenges will be in selecting the features-which feature to choose and which feature to ignore.

Again, some records are heavily affected by the data acquisition process, i.e., some records are corrupted with patches of electrical noise, some DE and FE measurements are identical except for a scaling factor, some records appear to be clipped in sections. Moreover, the bearing test rig assembly appeared to affect the detection and diagnosis results more than the fault itself, with an indication of mechanical looseness detected in many of the datasets [23]. Also, the CWRU dataset consists of the data which is operated under fixed speed and load. The significant difference can be seen in the same bearing fault class when operated under varying loads or speeds.

VI. CHALLENGES USING DL MODELS FOR MFDD

The mechanical vibration signal, indeed, is one of the most essential and abundant sources of information for a proper understanding of the phenomena related to bearing effect. Moreover, because of the numerous advantages, the use of DL algorithms for bearing fault detection has been practiced widely. However, DL algorithms have some limitations. 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 [153]. That is why many industrial applications do not generally prefer this black box tool. Some of the challenges using DL techniques for machinery fault detection are listed below:

- Deep learning models perform well when fed with enormous data. So, it is a great challenge to collect vast

TABLE 10. Some published work employing deep transfer learning and domain adaptation in MFDD.

A. R.	N.C.	Method	H.P.	Result	Remark
[123]	10	<ul style="list-style-type: none"> Deep Convolutional Neural Networks with Wide First-layer Kernels (WDCNN) which uses raw vibration signals as input and the wide kernels in the first convolutional layer for extracting features and suppressing high-frequency noise Small convolutional kernels in the earlier layers are used for multilayer nonlinear mapping Adaptive Batch Normalization (AdaBN) is implemented to improve the domain adaptation ability 	SoftMax ReLU AdaBN MaxPool	100% classification accuracy 95.9% average testing accuracy	<ul style="list-style-type: none"> Can perform well in noisy environment conditions too
[148]	10	<ul style="list-style-type: none"> A deep distance metric learning method in which a deep CNN is used as the leading architecture A representation clustering algorithm is proposed to decrease the distance of intra-class variations and maximize the length of inter-class differences simultaneously A domain adaptation method is adopted to reduce the maximum mean discrepancy between training and testing data 	BN LReLU Dropout MaxPool SoftMax	99.34% accuracy when the sample length is 8192	<ul style="list-style-type: none"> average training time is ~ 40 min when the sample length is 8192
[149]	10	<ul style="list-style-type: none"> Machinery fault diagnosis method for domain adaptation in which the maximum mean discrepancies between the source and target domains in multiple layers are minimized to solve the domain shift problem. Multiple kernels of MMD are used to leverage different kernels, and the MMD term is integrated with the classification loss for the network training CNN is used as the leading architecture where four stacked 1-D Convolutional layers are designed 	BN ReLU Max-Pool Dropout	99.76% with an enhanced model 99.17% with default architecture	<ul style="list-style-type: none"> feature space is divided into multiple regions corresponding with different labels the final classification is carried out in a high-dimensional space nonlinearly
[150]	10	<ul style="list-style-type: none"> A deep transfer learning method for MFDD in which the diagnostic knowledge learned from sufficient supervised data of multiple rotating machines is transferred to the weakly supervised target domain is proposed In adversarial training, the domain discriminator is updated to accurately classify the source and target domain data using the high-level features $x_h = G(x)$, while the feature extractor aims to generate domain-invariant features The classifier is trained to diagnose different machine conditions based on x_h under source supervision. The raw measured vibration data, including source-domain and target-domain, are directly fed into the network. All the samples are sent through the feature extractor G, and high-level representation is obtained. Then, the learned representations from the feature extractor are taken as the inputs for the classifier 	ReLU DropOut SoftMax LReLU	In the extreme case where minimal supervised data are available with N_t , train = 10, over 70% accuracies can be obtained	<ul style="list-style-type: none"> Throughout the network, a convolutional filter with a size of 10 is generally employed for feature extraction. ReLU is used for feature extractor and a domain discriminator network Leaky ReLU is used for the classifier network.
[63]	6	<ul style="list-style-type: none"> Firstly, the raw time-series training data collected by sensors are used to train the DNN until the cost function of DNN gets minimal value. Then, test data are used to test the classification accuracy of the DNN on local time-series data Finally, fault diagnosis considering temporal coherence with recent time-series data is implemented. 	Cross-Entropy	Best accuracy on CWRU dataset was 94.4%	<ul style="list-style-type: none"> Taking the benefit of temporal coherence of data, the model can directly recognize raw time-series data without any feature selection or signal processing
[151]	10	<ul style="list-style-type: none"> A deep residual learning architecture is proposed A 1-D convolutional layer is first designed with F_N local filter kernels of F_L length window size where the raw data is processed for feature extraction Then, multiple residual blocks are stacked to learn the high-level representation Two building blocks, in which each block contains two convolutional layers, are employed. 	SoftMax MaxPool BN	99.98% average accuracy was obtained	<ul style="list-style-type: none"> Raw vibration data were used directly as model inputs without preprocessing Zeros-padding operation is implemented
[152]	10	<ul style="list-style-type: none"> A method based on deep CNN and random forest (RF) ensemble learning in which time domain vibration signals are converted into 2-D gray-scale images containing sufficient fault information by continuous wavelet transform (CWT). Then, a CNN model based on LeNet-5 is built to automatically extract multi-level features that are sensitive to the detection of faults from the images. 	SoftMax MaxPool ReLU	99.73%	<ul style="list-style-type: none"> The multi-level features containing both local and global information are utilized to diagnose bearing faults by the ensemble of multiple RF classifiers.

amounts of faulty or healthy machinery data for DL algorithms to work effectively in machinery fault detection and diagnosis.

- Training DL models is extremely computational and expensive. The most sophisticated models take weeks to train using high-performance GPU machines. Also, a huge memory is required to train the model [154].
- Again, DL works poorly when the dataset is unbalanced. It misclassifies the class with a higher number of data samples. Furthermore, having a balanced bearing dataset is itself a significant challenge. So, other different approaches should be implemented to train DL models when the data is unbalanced.
- The working condition of mechanical systems are very complicated and keep on changing from time to time according to production requirements. So, it is quite unrealistic to collect and label enough training samples for DL algorithms to work on bearing fault detection and diagnosis.
- Industry systems are not allowed to run into faulty states due to the consequences, especially for critical systems and failures, and most of the electro-mechanical failures occur slowly and follow a degradation path such that failure degradation of a system might take months or even years, which makes it challenging to collect related datasets.
- Vibration signals need expensive sensors. Again, for the accurate data, the sensors must be mounted tightly on the machine for which the experts may be needed [9]. Otherwise, the data obtained may not be of good quality, which will result in the poor performance of DL models.
- The DL architectures like auto-encoders require a pre-training stage and become ineffective if errors are present in the first layers. Such errors may cause the network to learn to reconstruct the average of the training data. CNNs, on the other hand, heavily rely on labeled data and may require many layers to find the entire hierarchy. Moreover, DBNs do not account for the 2-D structure of an input image, which may significantly affect their performance and applicability when 2-D images are used as input. Again, the steps towards further optimization of the network based on maximum likelihood training approximation are unclear in DBN [155]. Furthermore, finding Nash-equilibrium in GAN training is itself a challenging process. RNNs, on the other hand, encounter the problem of gradient vanishing/ exploding [2]. Moreover, RL needs constant supervision of the subject, and it also needs to know where actions lead to estimate actions and make choices.
- Much deeper networks may have difficulties and challenges such as exploding or vanishing gradients and degradation in the training process, and there is a reduction of accuracy when the depth of the network exceeds maximum [156]. So, network optimization is another challenge with deep network architecture.

- The data quality may not always be good. Sometimes the data is of poor quality and redundant too. Though DL models work great with noisy data, they are still struggling hard to learn from weak quality data, which is a challenging work for DL algorithms in MFDD [38].

VII. RECOMMENDATIONS

For the successful implementation of DL algorithms in machinery fault detection and diagnosis using the CWRU bearing dataset, the authors make the following suggestions:

- Before developing any model or algorithm, we highly recommend future researchers to study and analyze the CWRU dataset and the benchmark study thoroughly.
- CWRU is a varied and complex database. A well-designed classification method and algorithm is needed for classifying such varied and complex dataset. For a DL model to work efficiently, there should be enough data. So, one can apply data augmentation techniques before training the model. The accessible data augmentation techniques like GAN, additional Gaussian noise, masking noise, signal translation, amplitude shifting, time-stretching, overlapping, and so on can be used.
- The CWRU dataset is also used as the validating dataset for the accuracy of the model. When the data from other sources than the CWRU dataset is used, the accuracy may drop down because of the influence of noise or variation in motor speed. So, proper measures should be taken before doing this.
- Much of the research has been carried out in a balanced dataset. Comparatively, it is easier to use a balanced dataset than the imbalanced one. The imbalanced dataset is the characteristic of real-world-applications. There are many approaches to deal with imbalanced data. Authors recommend the readers to have a thorough research about the methods. Traditional techniques simply could not identify all the features of imbalanced data. A better approach would be to develop a more generic organizing principle that can accommodate all possible types, rather than individual approaches that deal with specific types one by one.
- Condition monitoring with vibration signals is itself a challenging task. Vibration signals need expensive sensors. Again, for the accurate data, the sensors must be mounted tightly on the machine for which the experts may be needed [9]. So, if the researchers are interested in working with their data, a proper procedure must be applied when collecting the vibration data.
- Faults in bearings often manifest themselves at high frequencies [23]. If the researchers are interested in generating the fault samples, the use of a high sampling rate is recommended.
- The authors recommend the bearing fault researchers to follow the systematic and comprehensive approach in documentation as well, which will be helpful for future researchers.

VIII. CONCLUSION

In the age of industry 4.0, deep learning algorithms have attracted increasing attention for several research applications. Recently, DL models have been broadly employed in machinery fault detection and diagnosis systems. With continued rapid advances in computer technology, DL models will continue to be robust and attractive tools in machinery fault detection and diagnosis systems. An attempt to summarize and review the recent works and research on MFDD using the CWRU dataset applying deep learning algorithms has been made in this paper. We tried our best to cover the recent works which use DE defects bearing data of the CWRU dataset employing DL-algorithms detailly. The challenges in dealing with vibration data, employing DL-based models, and the recommendations from the authors are listed, which will surely help future researchers.

ACKNOWLEDGMENT

The authors thank Case Western Reserve University for providing free access to the bearing vibration experimental data from their website.

REFERENCES

- [1] A. Boudiaf, A. Moussaoui, A. Dahane, and I. Atoui, "A comparative study of various methods of bearing faults diagnosis using the case western reserve university data," *J. Failure Anal. Prevention*, vol. 16, no. 2, pp. 271–284, Apr. 2016.
- [2] S. Zhang, S. Zhang, B. Wang, and T. G. Habetler, "Deep learning algorithms for bearing fault diagnostics—A comprehensive review," *IEEE Access*, vol. 8, pp. 29857–29881, 2020.
- [3] S.-Y. Li and K.-R. Gu, "Smart fault-detection machine for ball-bearing system with chaotic mapping strategy," *Sensors*, vol. 19, no. 9, p. 2178, 2019.
- [4] *Case Western Reserve University Bearing Data Center*. Accessed: Dec. 22, 2019. [Online]. Available: <https://csegroups.case.edu/bearingdatacenter/home>
- [5] R. N. Bell, D. W. McWilliams, P. O'Donnell, C. Singh, and S. J. Wells, "Report of large motor reliability survey of industrial and commercial installations, part I," *IEEE Trans. Ind. Appl.*, vol. IA-21, no. 4, pp. 853–864, Apr. 1985.
- [6] R. N. Bell, C. R. Heising, P. O'Donnell, S. J. Wells, and C. Singh, "Report of large motor reliability survey of industrial and commercial installations, part II," *IEEE Trans. Ind. Appl.*, vol. IA-21, no. 4, pp. 865–872, Jan. 1985.
- [7] W. Huang, J. Cheng, and Y. Yang, "Rolling bearing fault diagnosis and performance degradation assessment under variable operation conditions based on nuisance attribute projection," *Mech. Syst. Signal Process.*, vol. 114, pp. 165–188, Jan. 2019.
- [8] M. S. Safizadeh and S. K. Latifi, "Using multi-sensor data fusion for vibration fault diagnosis of rolling element bearings by accelerometer and load cell," *Inf. Fusion*, vol. 18, pp. 1–8, Jul. 2014.
- [9] R. A. Patel and B. R. Bhalja, "Condition monitoring and fault diagnosis of induction motor using support vector machine," *Electr. Power Compon. Syst.*, vol. 44, no. 6, pp. 683–692, Apr. 2016.
- [10] C. Sun, Z. Zhang, Z. He, Z. Shen, and B. Chen, "Manifold learning-based subspace distance for machinery damage assessment," *Mech. Syst. Signal Process.*, vols. 70–71, pp. 637–649, Mar. 2016.
- [11] V. K. Rai and A. R. Mohanty, "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2607–2615, Aug. 2007.
- [12] Z. Peng, F. Chu, and Y. He, "Vibration signal analysis and feature extraction based on reassigned wavelet scalogram," *J. Sound Vibrat.*, vol. 253, no. 5, pp. 1087–1100, Jun. 2002.
- [13] C. Cheng, B. Zhou, G. Ma, D. Wu, and Y. Yuan, "Wasserstein distance based deep adversarial transfer learning for intelligent fault diagnosis," *arXiv:1903.06753*, vol. abs/1903.0, pp. 1–11, 2019. [Online]. Available: <https://arxiv.org/abs/1903.06753>
- [14] X. Wang, C. Liu, F. Bi, X. Bi, and K. Shao, "Fault diagnosis of diesel engine based on adaptive wavelet packets and EEMD-fractal dimension," *Mech. Syst. Signal Process.*, vol. 41, nos. 1–2, pp. 581–597, Dec. 2013.
- [15] C. Castejón, M. Jesús, J. J. Gómez, J. Carlos García-Prada, A. J. O. Nez, and H. Rubio, "Automatic selection of the WPT decomposition level for condition monitoring of rotor elements based on the sensitivity analysis of the wavelet energy," *Int. J. Acoust. Vib.*, vol. 20, no. 2, pp. 95–100, Jun. 2015.
- [16] P. Gupta and M. K. Pradhan, "Fault detection analysis in rolling element bearing: A review," *Mater. Today, Proc.*, vol. 4, no. 2, pp. 2085–2094, 2017.
- [17] N. Tandon and B. C. Nakra, "Detection of defects in rolling element bearings by vibration monitoring," *J. Mech. Eng. Division*, vol. 73, pp. 271–282, Jan. 1993.
- [18] D. Dyer and R. M. Stewart, "Detection of rolling element bearing damage by statistical vibration analysis," *J. Mech. Des.*, vol. 100, no. 2, pp. 229–235, Apr. 1978.
- [19] R. Dwyer, "Detection of non-Gaussian signals by frequency domain kurtosis estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Dec. 1983, vol. 2, pp. 607–610.
- [20] A. Shrivastava and S. Wadhvani, "Development of fault detection system for ball bearing of induction motor using vibration signal," *Int. J. Sci. Res.*, vol. 2, no. 5, pp. 256–259, Jun. 2012.
- [21] M. Cocconcelli, R. Zimroz, R. Rubini, and W. Bartelmus, "STFT based approach for ball bearing fault detection in a varying speed motor," in *Proc. Cond. Monit. Mach. Non-Stationary Oper.*, Oct. 2012, pp. 41–50.
- [22] J.-H. Lee, J. Kim, and H.-J. Kim, "Development of enhanced Wigner-Ville distribution function," *Mech. Syst. Signal Process.*, vol. 15, no. 2, pp. 367–398, Mar. 2001.
- [23] W. A. Smith and R. B. Randall, "Rolling element bearing diagnostics using the case western reserve university data: A benchmark study," *Mech. Syst. Signal Process.*, vols. 64–65, pp. 100–131, Dec. 2015.
- [24] 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.
- [25] A. Dey, "Machine learning algorithms: A review," *Int. J. Comput. Sci. Inf. Technol.*, vol. 7, no. 3, pp. 1174–1179, 2016.
- [26] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Fault diagnosis of ball bearings using machine learning methods," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 1876–1886, Mar. 2011.
- [27] M.-Y. Chow, P. M. Mangum, and S. O. Yee, "A neural network approach to real-time condition monitoring of induction motors," *IEEE Trans. Ind. Electron.*, vol. 38, no. 6, pp. 448–453, 1991.
- [28] A. Malhi and R. X. Gao, "PCA-based feature selection scheme for machine defect classification," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 6, pp. 1517–1525, Dec. 2004.
- [29] Y. S. Wang, Q. H. Ma, Q. Zhu, X. T. Liu, and L. H. Zhao, "An intelligent approach for engine fault diagnosis based on Hilbert–Huang transform and support vector machine," *Appl. Acoust.*, vol. 75, pp. 1–9, Jan. 2014.
- [30] S. Bhandari, H. P. Zhao, H. Kim, P. Khan, and S. Ullah, "Packet scheduling using SVM models in wireless communication networks," *J. Internet Technol.*, vol. 20, no. 5, pp. 1505–1512, 2019.
- [31] D. H. Pandya, S. H. Upadhyay, and S. P. Harsha, "Fault diagnosis of rolling element bearing with intrinsic mode function of acoustic emission data using APF-KNN," *Expert Syst. Appl.*, vol. 40, no. 10, pp. 4137–4145, Aug. 2013.
- [32] F. Shen, C. Chen, R. Yan, and R. X. Gao, "Bearing fault diagnosis based on SVD feature extraction and transfer learning classification," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM)*, Oct. 2015, pp. 1–6.
- [33] 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.
- [34] 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.*, vol. 67, no. 10, pp. 7331–7376, Oct. 2019.
- [35] B. Zoph, E. D. Cubuk, G. Ghiasi, T.-Y. Lin, J. Shlens, and Q. V. Le, "Learning data augmentation strategies for object detection," *arXiv:1906.11172*, pp. 1–13, Jun. 2019. [Online]. Available: <https://arxiv.org/abs/1906.11172>
- [36] Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: A big data–AI integration perspective," *IEEE Trans. Knowl. Data Eng.*, early access, Oct. 8, 2019, doi: [10.1109/TKDE.2019.2946162](https://doi.org/10.1109/TKDE.2019.2946162).
- [37] R. Kitchin and G. McArdle, "What makes big data, big data? Exploring the ontological characteristics of 26 datasets," *Big Data Soc.*, vol. 3, no. 1, Jan. 2016, Art. no. 205395171663113.

- [38] 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, Jul. 2018.
- [39] G. Douzas and F. Bacao, "Effective data generation for imbalanced learning using conditional generative adversarial networks," *Expert Syst. Appl.*, vol. 91, pp. 464–471, Jan. 2018.
- [40] X. Ke, J. Dick, D. Quoc, and T. Le Gia, *Learning With Imbalanced Data*. Sydney, NSW, Australia: Univ. New South Wale, 2018.
- [41] B. Krawczyk, "Learning from imbalanced data: Open challenges and future directions," *Prog. Artif. Intell.*, vol. 5, no. 4, pp. 221–232, Nov. 2016.
- [42] R. Akbani, S. Kwek, and N. Japkowicz, "Applying support vector machines to imbalanced datasets to imbalanced datasets," in *Proc. Eur. Conf. Mach. Learn.*, 2004, pp. 39–50.
- [43] H. Zheng, R. Wang, Y. Yang, J. Yin, Y. Li, Y. Li, and M. Xu, "Cross-domain fault diagnosis using knowledge transfer strategy: A review," *IEEE Access*, vol. 7, pp. 129260–129290, 2019.
- [44] *Konstruktions-Und Antriebsstechnik (KAT)—Data Sets and Download (Universität Paderborn)*. Accessed: Dec. 22, 2019. [Online]. Available: <https://mb.uni-paderborn.de/kat/forschung/datacenter/bearing-dataset/>
- [45] V. Pandhare, J. Singh, and J. Lee, "Convolutional neural network based rolling-element bearing fault diagnosis for naturally occurring and progressing defects using time-frequency domain features," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Paris)*, May 2019, pp. 320–326.
- [46] Y. Chen, G. Peng, C. Xie, W. Zhang, C. Li, and S. Liu, "ACDIN: Bridging the gap between artificial and real bearing damages for bearing fault diagnosis," *Neurocomputing*, vol. 294, pp. 61–71, Jun. 2018.
- [47] 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.
- [48] L. Wen, X. Li, and L. Gao, "A transfer convolutional neural network for fault diagnosis based on ResNet-50," *Neural Comput. Appl.*, vol. 32, no. 10, pp. 6111–6124, May 2020.
- [49] *Introduction|Femto-st—Sciences & Technologies*. Accessed: Dec. 28, 2019. [Online]. Available: <https://www.femto-st.fr/en/The-Institute/introduction>
- [50] *NASA Bearing Dataset|Acoustics and Vibration Database*. Accessed: Dec. 22, 2019. [Online]. Available: <http://data-acoustics.com/measurements/bearing-faults/bearing-4/>
- [51] *PRONOSTIA Formatronics*. Accessed: Mar. 30, 2020. [Online]. Available: <http://www.formatronics.fr/en/pronostia-en.html>
- [52] P. Nectoux, "PRONOSTIA: An experimental platform for bearings accelerated degradation tests," in *Proc. IEEE Int. Conf. Prognostics Health Manage.*, 2012, pp. 1–9.
- [53] K. Medjaher, N. Zerhouni, and J. Baklouti, "Data-driven prognostics based on health indicator construction: Application to PRONOSTIA's data," in *Proc. Eur. Control Conf. (ECC)*, Jul. 2013, pp. 1451–1456.
- [54] A. Mosallam, K. Medjaher, N. Zerhouni, A. Mosallam, K. Medjaher, and N. Zerhouni, "Nonparametric time series modelling for industrial prognostics and health management," *Int. J. Adv. Manuf. Technol.*, vol. 69, pp. 1685–1699, Nov. 2013.
- [55] F. Sloukia, M. El Aroussi, H. Medromi, and M. Wahbi, "Bearings prognostic using mixture of Gaussians hidden Markov model and support vector machine," in *Proc. IEEE/ACS Int. Conf. Comput. Syst. Appl. (AICCSA)*, 2013, pp. 1–4.
- [56] H. Li and Y. Wang, "Rolling bearing reliability estimation based on logistic regression model," in *Proc. Int. Conf. Qual., Rel., Risk, Maintenance, Saf. Eng. (QR2MSE)*, Jul. 2013, pp. 1730–1733.
- [57] MFPT Dataset. *Society For Machinery Failure Prevention Technology*. Accessed: Mar. 28, 2020. [Online]. Available: <https://mfpt.org/fault-datasets/>
- [58] C. Sobie, C. Freitas, and M. Nicolai, "Simulation-driven machine learning: Bearing fault classification," *Mech. Syst. Signal Process.*, vol. 99, pp. 403–419, Jan. 2018.
- [59] G. Li, C. Deng, J. Wu, X. Xu, X. Shao, and Y. H. Wang, "Sensor data-driven bearing fault diagnosis based on deep convolutional neural networks," *Sensors*, vol. 19, no. 12, p. 2750, 2019.
- [60] L. Wen, L. Gao, and X. Li, "A new snapshot ensemble convolutional neural network for fault diagnosis," *IEEE Access*, vol. 7, pp. 32037–32047, 2019.
- [61] D. Verstraete, A. Ferrada, E. L. Drogue, V. Meruane, and M. Modarres, "Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings," *Shock Vibrat.*, vol. 2017, Oct. 2017, Art. no. 5067651.
- [62] *IMS Bearing Dataset: Extracting Failure Modes From Vibration Signals*. Accessed: Dec. 22, 2019. [Online]. Available: http://mkalikatzarakis.eu/wp-content/uploads/2018/12/IMS_dset.html
- [63] R. Zhang, Z. Peng, L. Wu, B. Yao, and Y. Guan, "Fault diagnosis from raw sensor data using deep neural networks considering temporal coherence," *Sensors*, vol. 17, no. 3, p. 549, 2017.
- [64] W. Gousseau, J. Antoni, F. Girardin, and J. Griffaton, "Analysis of the rolling element bearing data set of the center for intelligent maintenance systems of the University of Cincinnati," in *Proc. CM*, 2016, pp. 1–14.
- [65] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [66] H. Wang and B. Raj, "On the origin of deep learning," *arXiv:1702.07800*, pp. 1–72, 2017. [Online]. Available: <https://arxiv.org/abs/1702.07800>
- [67] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Accessed: Dec. 22, 2019. [Online]. Available: <https://www.deeplearningbook.org/contents/autoencoders.html>
- [68] R. Sathya and A. Abraham, "Comparison of supervised and unsupervised learning algorithms for pattern classification," *Int. J. Adv. Res. Artif. Intell.*, vol. 2, no. 2, pp. 34–38, 2013.
- [69] Y. C. A. Padmanabha Reddy, P. Viswanath, and B. Eswara Reddy, "Semi-supervised learning: A brief review," *Int. J. Eng. Technol.*, vol. 7, no. 18, p. 81, 2018.
- [70] D. P. Kingma, D. J. Rezende, S. Mohamed, and M. Welling, "Semi-supervised learning with deep generative models," *arXiv:1406.5298*, pp. 1–9, 2014. [Online]. Available: <https://arxiv.org/abs/1406.5298>
- [71] B. J. A. Kröse, "Learning from delayed rewards," *Robot. Auto. Syst.*, vol. 15, no. 4, pp. 233–235, Oct. 1995.
- [72] L. Torrey and J. Shavlik, "Transfer learning," in *Handbook of Research on Machine Learning Applications*, E. S. Olivas, J. D. M. Guerrero, M. M. Sober, J. R. M. Bedito, and A. J. S. Lopez, Eds. Hershey, PA, USA: IGI Global, 2009, pp. 1–22.
- [73] C. Tan, F. Sun, T. Kong, W. Zhang, Y. Chao, and C. Liu, "A survey on deep transfer learning," in *Proc. 27th Int. Conf. Artif. Neural Netw. (ICANN)*, vol. 11141, 2018, pp. 270–279.
- [74] M. Wang and W. Deng, "Deep visual domain adaptation: A survey," *Neurocomputing*, vol. 312, pp. 135–153, Oct. 2018.
- [75] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 303–315, May 2016.
- [76] W. Lu, X. Wang, C. Yang, and T. Zhang, "A novel feature extraction method using deep neural network for rolling bearing fault diagnosis," in *Proc. 27th Chin. Control Decis. Conf. (CCDC)*, May 2015, pp. 2427–2431.
- [77] 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.
- [78] F. Wang, B. Dun, G. Deng, H. Li, and Q. Han, "A deep neural network based on kernel function and auto-encoder for bearing fault diagnosis," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2018, pp. 1–6.
- [79] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, Jan. 2017.
- [80] X. Guo, C. Shen, and L. Chen, "Deep fault recognizer: An integrated model to denoise and extract features for fault diagnosis in rotating machinery," *Appl. Sci.*, vol. 7, no. 1, p. 41, 2017.
- [81] 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, 2018.
- [82] S. Haidong, J. Hongkai, L. Xingqiu, and W. Shuai-peng, "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.
- [83] J. Sun, C. Yan, and J. Wen, "Intelligent bearing fault diagnosis method combining compressed data acquisition and deep learning," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 1, pp. 185–195, Jan. 2018.
- [84] S. Zhang, F. Ye, B. Wang, and T. G. Habetler, "Semi-supervised learning of bearing anomaly detection via deep variational autoencoders," 2019, *arXiv:1912.01096*. [Online]. Available: <http://arxiv.org/abs/1912.01096>
- [85] 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.
- [86] Y. Lecun, L. Bottou, Y. Bengio, and P. Ha, "LeNet," *Proc. IEEE*, vol. 86, pp. 2278–2324, Nov. 1998.

- [87] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
- [88] 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.
- [89] M. Xia, T. Li, L. Xu, L. Liu, and C. W. de Silva, "Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 101–110, Feb. 2018.
- [90] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Intelligent rotating machinery fault diagnosis based on deep learning using data augmentation," *J. Intell. Manuf.*, vol. 31, no. 2, pp. 433–452, Feb. 2020.
- [91] O. Janssens, V. Slavkovič, B. Vervisch, K. Stockman, M. Locquifier, S. Verstockt, R. Van de Walle, and S. Van Hoecke, "Convolutional neural network based fault detection for rotating machinery," *J. Sound Vib.*, vol. 377, pp. 331–345, Sep. 2016.
- [92] L. Eren, "Bearing fault detection by one-dimensional convolutional neural networks," *Math. Probl. Eng.*, vol. 2017, Jul. 2017, Art. no. 8617315.
- [93] 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.
- [94] X. Li, P. Ding, and X. Shi, "Research on bearing fault detection based on convolution neural network," in *Proc. Chin. Autom. Congr. (CAC)*, Oct. 2017, pp. 5130–5134.
- [95] W. Zhang, G. Peng, and C. Li, "Bearings fault diagnosis based on convolutional neural networks with 2-D representation of vibration signals as input," *MATEC Web Conf.*, vol. 95, Oct. 2017, Art. no. 13001.
- [96] J. Zhang, Y. Sun, L. Guo, H. Gao, X. Hong, and H. Song, "A new bearing fault diagnosis method based on modified convolutional neural networks," *Chin. J. Aeronaut.*, vol. 33, no. 2, pp. 439–447, Feb. 2020.
- [97] Y. Yuan, G. Ma, C. Cheng, B. Zhou, H. Zhao, H.-T. Zhang, and H. Ding, "Artificial intelligent diagnosis and monitoring in manufacturing," 2018, *arXiv:1901.02057*. [Online]. Available: <http://arxiv.org/abs/1901.02057>
- [98] D.-T. Hoang and H.-J. Kang, "Rolling element bearing fault diagnosis using convolutional neural network and vibration image," *Cognit. Syst. Res.*, vol. 53, pp. 42–50, Jan. 2019.
- [99] Z. Zilong and Q. Wei, "Intelligent fault diagnosis of rolling bearing using one-dimensional multi-scale deep convolutional neural network based health state classification," in *Proc. IEEE 15th Int. Conf. Netw., Sens. Control (ICNSC)*, Mar. 2018, pp. 1–6.
- [100] 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, Jul. 2006.
- [101] 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.
- [102] D. Tang Hoang and H. Jun Kang, "Deep belief network and dempster-shafer evidence theory for bearing fault diagnosis," in *Proc. IEEE 27th Int. Symp. Ind. Electron. (ISIE)*, Jun. 2018, pp. 841–846.
- [103] C. Shen, J. Xie, D. Wang, X. Jiang, J. Shi, and Z. Zhu, "Improved hierarchical adaptive deep belief network for bearing fault diagnosis," *Appl. Sci.*, vol. 9, no. 16, p. 3374, 2019.
- [104] Z. Chen, X. Zeng, W. Li, and G. Liao, "Machine fault classification using deep belief network," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. Proc.*, May 2016, pp. 1–6.
- [105] J. Tao, Y. Liu, and D. Yang, "Bearing fault diagnosis based on deep belief network and multisensor information fusion," *Shock Vib.*, vol. 37, no. 8, pp. 6077–6085, 2016.
- [106] Z. Chen and W. Li, "Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1693–1702, Jul. 2017.
- [107] S. Dong, Z. Zhang, G. Wen, S. Dong, Z. Zhang, and G. Wen, "Design and application of unsupervised convolutional neural networks integrated with deep belief networks for mechanical fault diagnosis," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Harbin)*, Jul. 2017, Art. no. 51365051.
- [108] 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.
- [109] X. Xin, Y. Liu, and J. Wu, "Bearing fault diagnosis method based on Hilbert envelope spectrum and deep belief network," in *Proc. 54th Annu. Conf. Soc. Instrum. Control Eng. Jpn.*, 2015, pp. 765–770.
- [110] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Meas. Sci. Technol.*, vol. 26, no. 11, Nov. 2015, Art. no. 115002.
- [111] J. Tao, Y. Liu, D. Yang, F. Tang, and C. Liu, "Fault diagnosis of rolling bearing using deep belief networks," in *Proc. Int. Symp. Mater., Energy Environ. Eng.*, 2015, pp. 566–569.
- [112] T. Liang, S. Wu, W. Duan, and R. Zhang, "Bearing fault diagnosis based on improved ensemble learning and deep belief network," *J. Phys., Conf. Ser.*, vol. 1074, Sep. 2018, Art. no. 012154.
- [113] 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.
- [114] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
- [115] Y. Li, N. Xiao, and W. Ouyang, "Improved boundary equilibrium generative adversarial networks," *IEEE Access*, vol. 6, pp. 11342–11348, 2018.
- [116] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," 2017, *arXiv:1701.07875*. [Online]. Available: <http://arxiv.org/abs/1701.07875>
- [117] S. Suh, H. Lee, J. Jo, P. Lukowicz, and Y. Lee, "Generative oversampling method for imbalanced data on bearing fault detection and diagnosis," *Appl. Sci.*, vol. 9, no. 4, p. 746, 2019.
- [118] H. Liu, J. Zhou, Y. Xu, Y. Zheng, X. Peng, and W. Jiang, "Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks," *Neurocomputing*, vol. 315, pp. 412–424, Nov. 2018.
- [119] F. Zhou, S. Yang, H. Fujita, D. Chen, and C. Wen, "Deep learning fault diagnosis method based on global optimization GAN for unbalanced data," *Knowl.-Based Syst.*, vol. 187, Jan. 2020, Art. no. 104837.
- [120] W. Jiang, C. Cheng, B. Zhou, G. Ma, and Y. Yuan, "A novel GAN-based fault diagnosis approach for imbalanced industrial time series," 2019, *arXiv:1904.00575*. [Online]. Available: <http://arxiv.org/abs/1904.00575>
- [121] M. Zhang, D. Wang, W. Lu, J. Yang, Z. Li, and B. Liang, "A deep transfer model with wasserstein distance guided multi-adversarial networks for bearing fault diagnosis under different working conditions," *IEEE Access*, vol. 7, pp. 65303–65318, 2019.
- [122] D. Zhao, F. Liu, and H. Meng, "Bearing fault diagnosis based on the switchable normalization SSGAN with 1-D representation of vibration signals as input," *Sensors*, vol. 19, no. 9, p. 2000, 2019.
- [123] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, 2017.
- [124] W. Mao, Y. Liu, L. Ding, and Y. Li, "Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: A comparative study," *IEEE Access*, vol. 7, pp. 9515–9530, 2019.
- [125] Q. Li, L. Chen, C. Shen, B. Yang, and Z. Zhu, "Enhanced generative adversarial networks for fault diagnosis of rotating machinery with imbalanced data," *Meas. Sci. Technol.*, vol. 30, no. 11, Nov. 2019, Art. no. 115005.
- [126] S. Gao, X. Wang, X. Miao, C. Su, and Y. Li, "ASMID-GAN: An intelligent fault diagnosis method based on assembled 1D convolutional neural network and generative adversarial networks," *J. Signal Process. Syst.*, vol. 91, no. 10, pp. 1237–1247, Oct. 2019.
- [127] *CS 230—Recurrent Neural Networks Cheatsheet*. Accessed: Dec. 29, 2019. [Online]. Available: <https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks>
- [128] 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.
- [129] 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.
- [130] H. Jiang, X. Li, H. Shao, and K. Zhao, "Intelligent fault diagnosis of rolling bearings using an improved deep recurrent neural network," *Meas. Sci. Technol.*, vol. 29, no. 6, Jun. 2018, Art. no. 065107.
- [131] Z. Zhuang, H. Lv, J. Xu, Z. Huang, and W. Qin, "A deep learning method for bearing fault diagnosis through stacked residual dilated convolutions," *Appl. Sci.*, vol. 9, no. 9, p. 1823, 2019.
- [132] B. Purwahyudi, "Recurrent neural network (RNN) based bearing fault classification of induction motor employed in home water pump system," *J. Electr. Eng. Comput. Sci.*, vol. 3, no. 1, pp. 405–412, 2018.
- [133] J. S. Lal Senanayaka, H. Van Khang, and K. G. Robbersmyr, "Autoencoders and recurrent neural networks based algorithm for prognosis of bearing life," in *Proc. 21st Int. Conf. Electr. Mach. Syst. (ICEMS)*, Oct. 2018, pp. 537–542.

- [134] W. Abed, S. Sharma, R. Sutton, and A. Motwani, "A robust bearing fault detection and diagnosis technique for brushless DC motors under non-stationary operating conditions," *J. Control, Autom. Electr. Syst.*, vol. 26, no. 3, pp. 241–254, Jun. 2015.
- [135] L. Guo, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, pp. 98–109, May 2017.
- [136] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
- [137] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-Learning," *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 2094–2100.
- [138] R. Wang, H. Jiang, X. Li, and S. Liu, "A reinforcement neural architecture search method for rolling," *Measurement*, vol. 154, Mar. 2019, Art. no. 107417.
- [139] W. Dai, Z. Mo, C. Luo, J. Jiang, and Q. Miao, "Bearing fault diagnosis based on reinforcement learning and kurtosis," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Qingdao)*, Oct. 2019, pp. 1–5.
- [140] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," 2019, *arXiv:1911.02685*. [Online]. Available: <http://arxiv.org/abs/1911.02685>
- [141] X. Li, W. Zhang, and Q. Ding, "Cross-domain fault diagnosis of rolling element bearings using deep generative neural networks," *IEEE Trans. Ind. Electron.*, vol. 66, no. 7, pp. 5525–5534, Jul. 2019.
- [142] A. Zhang, S. Li, Y. Cui, W. Yang, R. Dong, and J. Hu, "Limited data rolling bearing fault diagnosis with few-shot learning," *IEEE Access*, vol. 7, pp. 110895–110904, 2019.
- [143] Y. Yang, P. Fu, and Y. He, "Bearing fault automatic classification based on deep learning," *IEEE Access*, vol. 6, pp. 71540–71554, 2018.
- [144] X. Li, W. Zhang, N.-X. Xu, and Q. Ding, "Deep learning-based machinery fault diagnostics with domain adaptation across sensors at different places," *IEEE Trans. Ind. Electron.*, vol. 67, no. 8, pp. 6785–6794, Aug. 2020.
- [145] M. He, D. He, and E. Bechhoefer, "Using deep learning based approaches for bearing fault diagnosis with AE sensors," in *Proc. Annu. Conf. Progn. Heal. Manag. Soc.*, Oct. 2016, pp. 282–291.
- [146] M. Zhao and M. Kang, "Weighted wavelet coefficients for fault diagnosis of planetary gearboxes," *IEEE Trans. Ind. Electron.*, vol. 65, no. 5, pp. 4290–4300, Dec. 2018.
- [147] M. 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.
- [148] X. Li, W. Zhang, and Q. Ding, "A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning," *Neurocomputing*, vol. 310, pp. 77–95, Oct. 2018.
- [149] X. Li, W. Zhang, Q. Ding, and J.-Q. Sun, "Multi-layer domain adaptation method for rolling bearing fault diagnosis," *Signal Process.*, vol. 157, pp. 180–197, Apr. 2019.
- [150] X. Li, W. Zhang, Q. Ding, and X. Li, "Diagnosing rotating machines with weakly supervised data using deep transfer learning," *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1688–1697, Mar. 2020.
- [151] W. Zhang, X. Li, and Q. Ding, "Deep residual learning-based fault diagnosis method for rotating machinery," *ISA Trans.*, vol. 95, pp. 295–305, Dec. 2019.
- [152] 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.
- [153] S. R. Saufi, Z. A. B. Ahmad, M. S. Leong, and M. H. Lim, "Challenges and opportunities of deep learning models for machinery fault detection and diagnosis: A review," *IEEE Access*, vol. 7, pp. 122644–122662, 2019.
- [154] N. Westlake, H. Cai, and P. Hall, "Detecting people in artwork with CNNs," in *Computer Vision (Lecture Notes in Computer Science)*, vol. 9913 LNCS. Oct. 2016, pp. 825–841.
- [155] A. Voulozimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, Feb. 2018, Art. no. 7068349.
- [156] E. C. Too, L. Yujian, S. Njuki, and L. Yingchun, "A comparative study of fine-tuning deep learning models for plant disease identification," *Comput. Electron. Agricul.*, vol. 161, pp. 272–279, Jun. 2019.



DHIRAJ NEUPANE received the bachelor's degree in electrical and electronics (communication) engineering from Kathmandu University, Nepal, in 2016. He is currently working as a Graduate Research Assistant with Changwon National University, South Korea. His research interests include machine learning, deep learning, intelligent fault diagnosis method, sonar signal processing, transfer learning, machinery fault detection and diagnosis, object detection, and classification.



JONGWON SEOK received the B.S., M.S., and Ph.D. degrees from Kyungpook National University, Daegu, South Korea, in 1993, 1995, and 1999, respectively. He was a Senior Member of the Research Staff in the Electronics and Telecommunications Research Institute (ETRI) from July 1999 to February 2004, where he was engaged in the development of a multichannel audio codec system, digital broadcasting content protection and management. Since March 2004, he has been with Changwon National University, where he is currently a Professor with the Department of Information and Communication Engineering. His research interests include digital and sonar signal processing, multimedia systems, and deep learning applications.

• • •