

Review

Deep Learning Aided Data-Driven Fault Diagnosis of Rotatory Machine: A Comprehensive Review

Shiza Mushtaq ¹, M. M. Manjurul Islam ²  and Muhammad Sohaib ^{1,*}

¹ Department of Computer Science & Engineering, Lahore Garrison University, Lahore 54000, Pakistan; shiz25musht@gmail.com

² Information, Communication and Technology Center, Fondazione Bruno Kessler, 38123 Trento, Italy; m.m.manjurul@gmail.com

* Correspondence: md.sohaib@lgu.edu.pk

Abstract: This paper presents a comprehensive review of the developments made in rotating bearing fault diagnosis, a crucial component of a rotatory machine, during the past decade. A data-driven fault diagnosis framework consists of data acquisition, feature extraction/feature learning, and decision making based on shallow/deep learning algorithms. In this review paper, various signal processing techniques, classical machine learning approaches, and deep learning algorithms used for bearing fault diagnosis have been discussed. Moreover, highlights of the available public datasets that have been widely used in bearing fault diagnosis experiments, such as Case Western Reserve University (CWRU), Paderborn University Bearing, PRONOSTIA, and Intelligent Maintenance Systems (IMS), are discussed in this paper. A comparison of machine learning techniques, such as support vector machines, k-nearest neighbors, artificial neural networks, etc., deep learning algorithms such as a deep convolutional network (CNN), auto-encoder-based deep neural network (AE-DNN), deep belief network (DBN), deep recurrent neural network (RNN), and other deep learning methods that have been utilized for the diagnosis of rotary machines bearing fault, is presented.

Keywords: auto-encoders; bearing; condition monitoring; convolutional neural network; deep belief network; deep learning; fault diagnosis; machine learning; recurrent neural network



Citation: Mushtaq, S.; Islam, M.M.M.; Sohaib, M. Deep Learning Aided Data-Driven Fault Diagnosis of Rotatory Machine: A Comprehensive Review. *Energies* **2021**, *14*, 5150. <https://doi.org/10.3390/en14165150>

Academic Editor: Fernando Sánchez Lasheras

Received: 16 July 2021

Accepted: 18 August 2021

Published: 20 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Motion is powered by electromechanical systems, which account for around 70% of the gross energy consumption in industrialized economies [1]. By 2017, the global market was at the size of USD 96,967.9 million, and is expected to reach USD 136,496.1 by the year 2025 [2]. One of the basic components that is used in industries is an electrical motor that converts electrical energy into mechanical energy.

Specifically, based on motor types, the global market is divided into DC, or AC or hermetic motors, which in turn are further subdivided as:

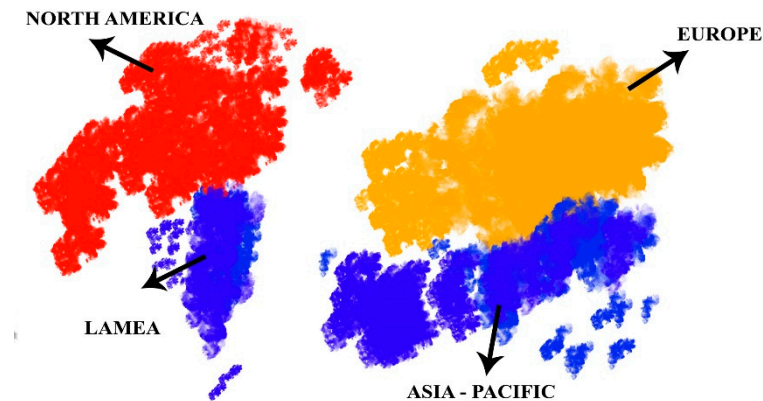
- Alternating current (AC) motors, synchronous AC motors, and induction AC motors;
- Direct current (DC) motors and brushless DC motors;
- Hermetic motors.

The global market of electric motors can be further classified based on operating industries such as automotive vehicles, industrial machinery, aerospace, household, and commercial applications. In the manufacturing and automotive industries, due to an increase in demand for compressor systems, the industrial segment contributed the largest share in the year 2017, which is even estimated to increase by 2025 [3].

Figure 1 provides the continent-wise market shares of electric motors' global usages and Figure 2 presents application-wise usages of electric motors in 2017 and their forecasts by 2025.

An electric motor consists of different apparatus such as a rotor, bearings, stator, air gap, commutator, and windings. Among these parts, a bearing is the core of the rotating

motor as it supports and locates the rotor to keep the air gap small and consistent, and it transfers the load from the motor to shaft. It is one of the most important mechanical components to diminish the friction between the rotating and stationary elements [4]. If the equipment fails to work during the use, it will affect the systems operations and can even cause serious economic losses and casualties. According to the literature review, around 50–60% of the failure of induction rotating machines is caused by bearing [4]. Therefore, fault diagnosis of a rotating machine bearing is inevitable to avoid the unexpected breakdown. An effective fault diagnosis of the bearing can ensure the efficient operation of the systems, and it detects and identifies the bearing faults during the operation of the motor.



GLOBAL ELECTRIC MOTOR MARKET - BY REGION

Figure 1. Global electric motor market record.

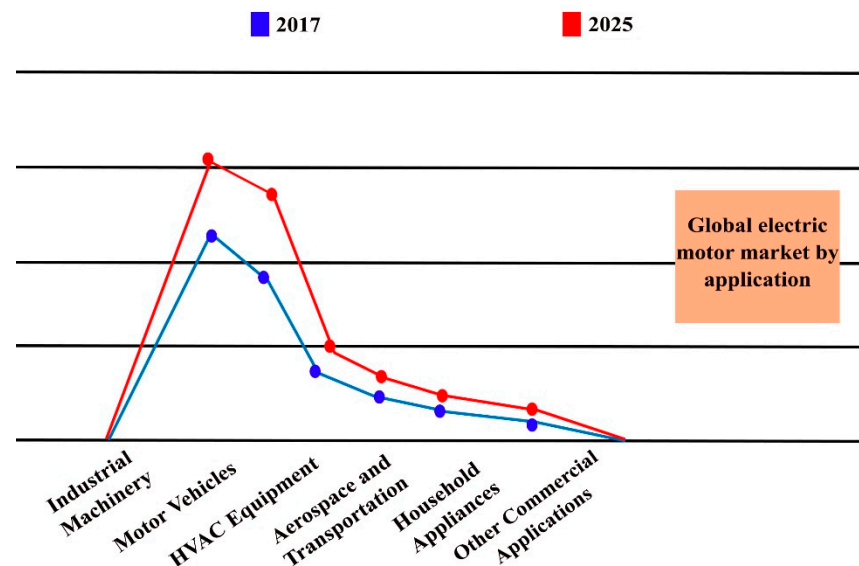


Figure 2. Global electric motor market by application [2].

Over the past few decades, researchers have carried out extensive research on bearing fault diagnosis. Additionally, new approaches and research are emerging in this field with the advancement of technology and industrial techniques. The work consists of various techniques that focus on different domains of the bearing fault diagnosis pipeline. For example, some researchers focused on the effective classification mechanism consisting of machine learning and deep learning techniques, while others dedicated themselves to

signal processing techniques to handle complex and nonlinear signals which are normally encountered during the fault diagnosis process.

This paper reviews the machine learning and deep learning algorithms used for bearing fault bearing diagnosis and discusses the future direction in this field. The main contributions are as follows:

1. A detailed analysis of a standard bearing fault diagnosis pipeline is given;
2. An overview of shallow machine learning techniques used in the field of bearing fault diagnosis and their limitations;
3. A systematic review of the literature available on bearing fault diagnosis in the last decade mainly focusing on the application of deep learning algorithms;
4. Discussion on the future directions in the field of bearing fault diagnosis.

The rest of the paper is organized in the following manner. Section 2 consists of the common public datasets available for bearing motor fault diagnosis experiments. Section 3 covers the classical and deep learning algorithm-based research on fault bearing diagnosis. Finally, Section 4 shows deep-learning-based fault bearing diagnosis research and their comparison.

2. A Standard Pipeline of Bearing Fault Diagnosis

Bearings are essential elements in rotating machines which ensure smooth operation by reducing friction among different components of the machine. Bearings are the main contributor to the failure of rotary machines, accounting for around 50–60%, since they have to operate in a harsh working environment [5]. An unexpected failure of the bearing can cause sudden breakdown to the machine or result in the entire system collapse, which could lead to huge financial loss and time wastage. Therefore, this sector receives significant attention from researchers in an effort to find more efficient solutions. The general diagnosis methodology consists of four steps, i.e., data acquisition, feature extraction, feature selection, and fault classification, as shown in Figure 3.

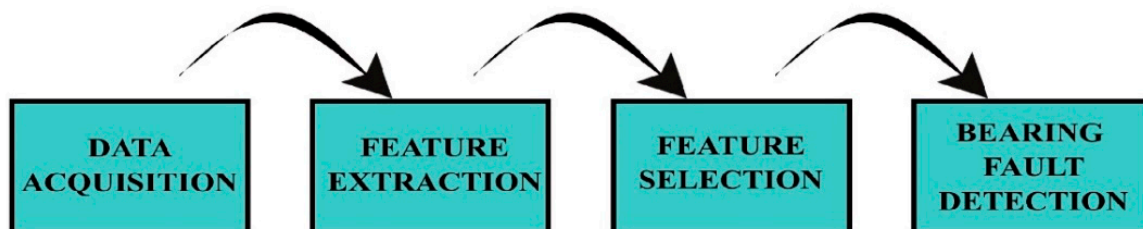


Figure 3. Generic steps of a data-driven fault diagnosis methodology.

2.1. Data Acquisition

It is the process of sampling signals that calculates real-world physical conditions and converts the result samples into the digital numeric values that a computer manipulates. In the first step of diagnosis, it collects the vibration signals, acoustic emission signals or electric motor current signals, that reflect the health status of bearings from the sensor systems.

2.2. Feature Extraction

Feature extraction begins with a set of measured data and creates derived values (features) that are intended to be useful and non-redundant, easing the learning and generalization phases and, in some situations, resulting in superior human interpretations. It converts the raw signals into statistical characteristics that convey information about the machine's status, which is known as feature extraction. In order to obtain high-accuracy recognition outcomes, the feature extractor design plays a vital part in the pattern recognition challenge. The actual bearing failure signals gathered from rotary machines are in the

time domain, and we may extract characteristics from signals in the time domain, frequency domain, and time–frequency domain. They can be investigated in the frequency and time–frequency domains using the appropriate transformation tool in the respective domains.

2.3. Feature Selection

Feature selection is the process of minimizing the number of input variables when creating a predictive model; the number of input variables should be reduced to lower the computational cost of modeling and, in some situations, to increase the model's performance. It chooses the most discriminant features, such as feature sets with high dimensions, which means having redundant and irrelevant features, which increases learning time, lowers classifier performance, and necessitates a lot of computation. The feature selection stage improves the classification accuracy while also reducing the calculation time. The two most frequent methods for feature selection are: (1) creating a new feature set with inferior dimensions from the extracted feature set. This might be accomplished using Independent Component Analysis (ICA) and Principal Component Analysis (PCA). (2) Using specific benchmarks, deleting non-sensitive or unneeded features. One of the most prominent approaches for this problem is Sequential Selection (SS).

2.4. Bearing Fault Diagnosis/Classification

After selecting the features, they must be passed into a learning-based classifier such as the k-nearest neighbor (KNN), artificial neural network (ANN), or support vector machine (SVM), one-dimension convolutional neural network (1D-CNN), strongly regularized deep convolutional neural network (SRDCNN), etc. to detect the bearing defect.

3. Dataset for Fault Bearing Experiment

A reliable and accessible dataset is required to develop data-driven bearing fault diagnosis methodologies. However, to collect data from a naturally degraded bearing is a time-consuming process. Therefore, most researchers prefer datasets with artificially induced faults on bearing. Some organizations and research centers have made efforts to create datasets and provide open access to researchers across the globe, which helps researchers to implement and evaluate bearing fault diagnosis algorithms. Some of the available popular public bearing datasets are discussed as follows.

3.1. Case Western Reserve University Bearing Dataset

Case Western Reserve University (CWRU) bearing data is a public dataset that is collected from a test rig, shown in Figure 4 [6]. The testbed contains (a) two HP motors, (b) a torque transducer/encoder, (c) a dynamometer and control electronics. According to the description available for the data, different single-point faults were introduced on both the bearings, i.e., the driven end as well as fan end, with an electro-discharge machine having fault diameters of 7, 14, 21, 28, and 40 mils on the rolling elements, inner and outer raceways. Moreover, the dataset consists of vibration signals collected under a sampling frequency of 12 kHz, where the motor speed varies from 1720–1797 revolutions per minute (RPM) and load variations of 0 to 3 HP by using accelerometers (sensors) installed on the fan- and drive-end bearings of the motor.

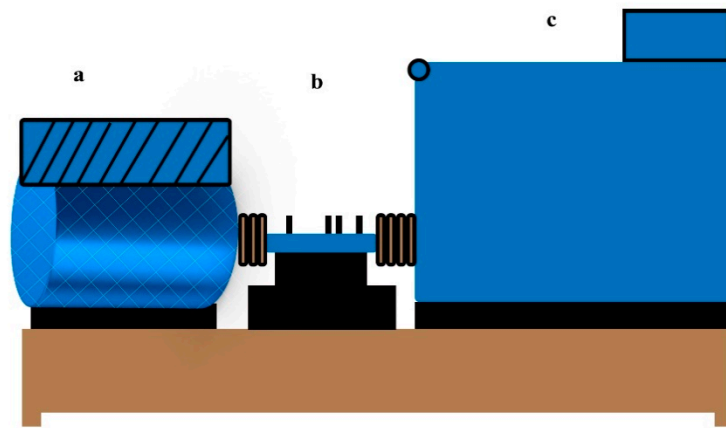


Figure 4. An illustration of Case Western Reserve University bearing test rig. Where (a) two HP motors, (b) a torque transducer/encoder, (c) dynamometer and control electronics.

3.2. Paderborn University Bearing Dataset

The Paderborn University dataset is another public archive for bearing datasets [7]. The testbed used for data collection consists of (a) a test motor, (b) a measuring shaft, (c) a bearing module, (d) a flywheel, and (e) a load motor, as depicted in Figure 5. The collected dataset contains synchronous vibration measurements in addition to motor current measurements. Sensors used in the equipment are one accelerometer, two current sensors, and 1 thermocouple. Vibration signals are under high resolution and 64 kHz sampling frequency. Experiments are carried out on perfectly working 6 perfectly working bearings and 26 damaged bearings, out of which, 12 are artificially damaged and the rest contain real damages triggered by accelerated tests.

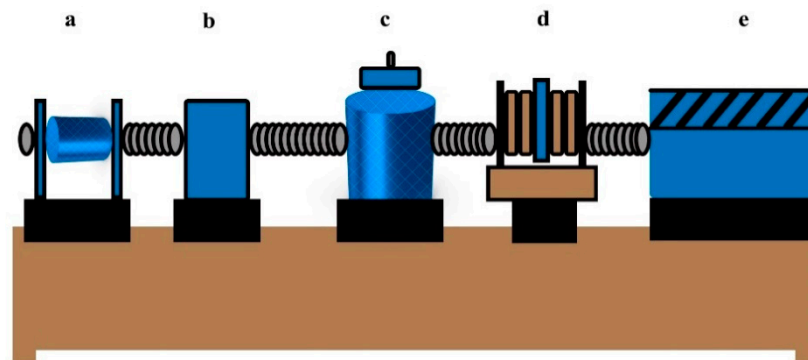


Figure 5. An illustration of Paderborn University experiment setup of bearing data acquisition and fault diagnosis. In the figure, (a) a test motor, (b) a measuring shaft, (c) a bearing module, (d) a flywheel, and (e) a load motor.

3.3. PRONOSTIA Dataset

PRONOSTIA is a useful dataset that contains a real portrayal of the real-time degradation of bearings under different conditions [8]. Its testing equipment consists of the following components: (a) NI CDA Q cards, (b) a pressure regulator, (c) cylinder pressure, (d) a force sensor, (e) the bearing tested, (f) accelerometers, (g) platinum RTD, (h) coupling, (i) a torque meter, (j) a speed reducer, (k) a speed sensor, and (l) an AC motor. Two uni-axis accelerometers of 25.6 kHz sampling frequency which are installed in horizontal and vertical positions, as can be seen in Figure 6. Equipment is categorized as a rotating speed sensor and force sensor.

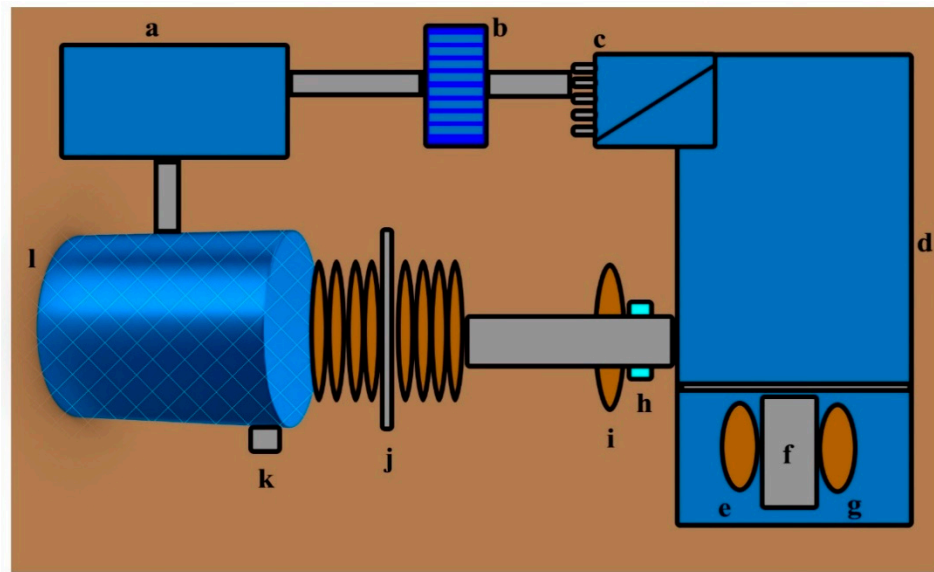


Figure 6. An illustration of PRONOSTIA testbed for bearing run-to-failure dataset. In the figure, (a) NI CDA Q cards, (b) a pressure regulator, (c) cylinder pressure, (d) a force sensor, (e) the bearing tested, (f) accelerometers, (g) platinum RTD, (h) coupling, (i) a torque meter, (j) a speed reducer, (k) a speed sensor, and (l) an AC motor.

3.4. IMS Dataset

This dataset is by the Intelligent Maintenance Systems (IMS) industry of the University of Cincinnati [9]. This dataset contains the natural bearing defect evolution, and contains a complete set of vibration signals from initial state to the failure with explicit time stamps, for which the bearing was kept running for 30 consecutive days on a fixed speed of 2000 rpm, which covers 86.4 million cycles before the confirmation of the defect [10]. The equipment contains (a) two accelerometers, four bearings as b1, b2, b3 and b4, (c) a radial load, and (d) four thermocouples which are attached to the outer race of each bearing. The collected vibration data are recorded repeatedly after 5 and 10 min for 1 sec with a 20 kHz sampling rate. The structure is illustrated below in Figure 7.

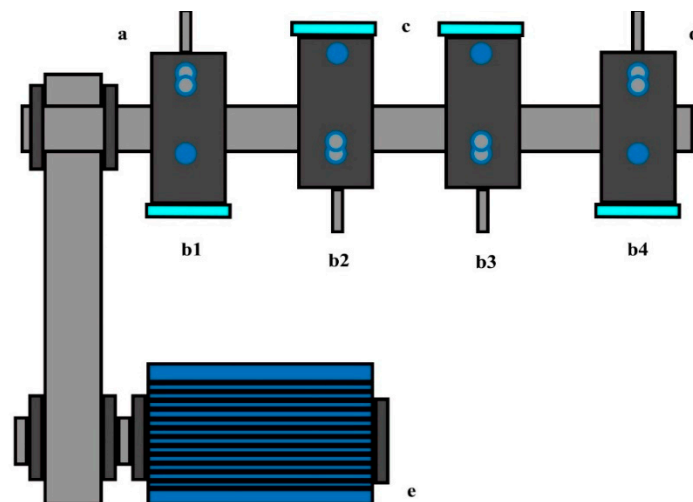


Figure 7. An illustration of the Intelligent Maintenance Systems (IMS) test rig for bearing run-to-failure dataset. In the figure, (a) two accelerometers, four bearings as b1, b2, b3 and b4, (c) a radial load, and (d) four thermocouples which are attached to the outer race of each bearing.

Table 1 provides a summary of different bearing datasets mentioned above.

Table 1. Comparison of motor bearing datasets.

SL	Dataset	Total Sensors	Sensors Type	Sample Frequency
1	Case Western Reserve University	2	Accelerometer	12 and 48 kHz
2	Paderborn University Dataset	1/2/1	Accelerometer, Current sensor, and thermocouple	64 kHz
3	PRONOSTIA Dataset	2/1	Accelerometer and thermocouple	25.6 kHz
4	Intelligent Maintenance Systems Dataset	2	Accelerometer	20 Hz

3.5. Highlights of the Datasets

In Section 3 of this article, titled “Dataset for Fault Bearing Experiment”, details four public datasets for fault bearing diagnosis have been elaborated upon. We conclude that the dataset Case Western Reserve University (CWRU) is one of the most used datasets for fault bearing diagnosis, as well as detection. The Paderborn University Dataset is considered as most efficient one that contains a real portrayal of the real-time degradation of bearings under different conditions. It has four sensors and a sampling frequency of 66 kHz. Moreover, the Intelligent Maintenance Systems Dataset contains the natural bearing defect evolution; both can be used for fault detection and the prediction of remaining useful life (RUL).

3.6. Effects of the Datasets

There are various conditions that effect the datasets; different datasets give different results on same training models.

A fault bearing dataset is composed of a signal, and a signal is made up of three components: (1) frequency, (2) amplitude, (3) phase. These properties of the signals vary from one fault to another, and variations in the signals can be observed within the signals of same health type if they are collected under different working conditions. So, the type of data used in the bearing fault diagnosis process has significance, as it affects the performance of the developed model. Figure 8 [11] illustrates vibration signals for a healthy bearing (HB), a bearing with an outer race crack (BORC), a bearing with a rough inner surface (BRIRC), a ball with corrosion pitting (BCP), and combined bearing components defects (CBD) at 2000 rpm with no loader [10]. It is evident from the figures that all the signals possess different waveforms; furthermore, these waveforms can undergo significant variations if the working conditions of the machinery during the data collection process changes. Table 2 shows the parameters/conditions that effect the signals.

Table 2. Parameters that effect the signals of datasets.

S. NO.	Parameters
1	Bearing specification (brand/model)
2	Outer race diameter
3	Inner race diameter
4	Ball diameter
5	Ball number
6	Contact angle
7	Clearance
8	Noise
9	Phase angle
10	Change in amplitude
11	Change in sampling frequency

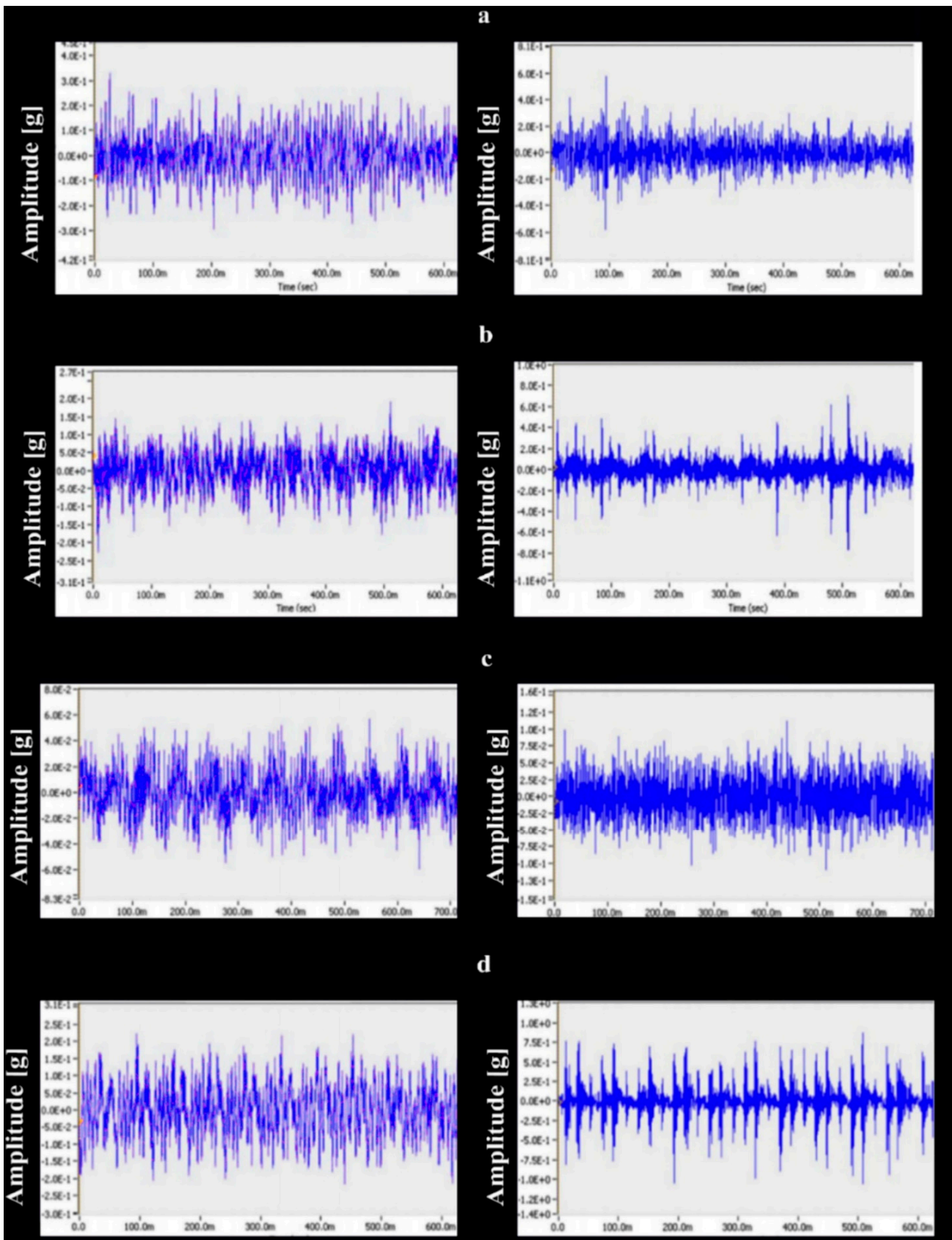


Figure 8. Time Waveform of Case Western Reserve University Bearing Dataset (a) vibration signals for healthy bearing (b) bearing with outer race crack, (c) bearing with rough inner race surface and (d) ball with corrosion pitting all at 2000 rpm with no loader. Right side time waveforms are vertical response and left side waveform are horizontal response.

4. Shallow Learning for Bearing Fault Diagnosis

In a standard bearing fault diagnosis framework, the fault classification is normally performed through traditional machine learning (ML) algorithms. The classical machine learning algorithms are considered to be shallow, as they do not follow the concept of deep networks. In return, the learning capability of the shallow networks is limited, and hence fails to extract salient information from the complex, nonlinear, and high dimension data. Therefore, to apply shallow learning for bearing fault diagnosis, researchers mostly rely on the standard bearing fault diagnosis pipeline that includes feature extraction and selection steps.

Classical ML Algorithms for Fault Bearing Diagnosis

Table 3 presents the classical machine learning algorithms used for bearing fault diagnosis-related research, including the k-mean singular value decomposition (K-SVD) dictionary algorithm for feature extraction [12], which can extract the fault frequency of every band, and then the back propagation neural network (BP NN) can be applied for the detection of failure type and to obtain the accurate fault bearing diagnosis. Similarly, [13] proposed an ANN method for fault bearing diagnosis using a Local Binary Pattern (LBP) histogram. It is based on the micro-texture analysis of vibration images with the local binary patterns. In [14], the author proposed the use of infrared thermography (IRT) for bearing fault diagnosis. For the decomposition of the thermal image, a two-dimensional discrete wavelet transform (2D-DWT) was used. The dimensionality of extracted data was reduced using principal component analysis (PCA), and then the most important characteristics were determined. The support vector machine (SVM), linear discriminant analysis (LDA), and k-nearest neighbor (KNN) were also evaluated as classifiers for fault classification and performance evaluation. The results show that the SVM outperformed both the LDA and the KNN. Furthermore, authors in [15] proposed a method that is based on sensing theory which can collect and compress raw data effectively concurrently. In [16], authors proposed the Energy Fluctuated Multiscale Feature (EFMF) mining method with the Deep ConvNet model for spindle bearing fault diagnosis. W. Zhang et al. [17] proposed a novel DL method using the residual learning algorithm for fault bearing diagnosis. P. Luo et al. proposed LSTM (long short-term memory) for fault recognition, and a neural network to exploit the fault detection for fault bearing diagnosis [18]. C. Wu et al. [19] proposed KMCSVC based on the kernel matrix to find the fault locations and identify severities. In [20], authors proposed a method to identify the bearing condition with statistical central moments time-domain vibration and five maximum peaks and power spectral density, with the help of ANN/SVM classifiers to obtain the high accuracy of bearing fault diagnosis. An automatic method for bearing fault diagnosis based on pattern recognition and signal processing technology with the combination of v-SVM for the detection of a fault was proposed in [21], whereas in [22], authors proposed a technique based on the voltage, speed and stator current of a machine for the diagnosis of bearing fault. This technique also detects lubricant problems and is perfect for classification. An improved Ant Colony Optimization (ACO) algorithm based on adaptive control parameters and the SVM (support vector machine) model for correct fault bearing diagnosis was proposed [23]. Furthermore, an overview of more classical ML-based research is presented in Table 3.

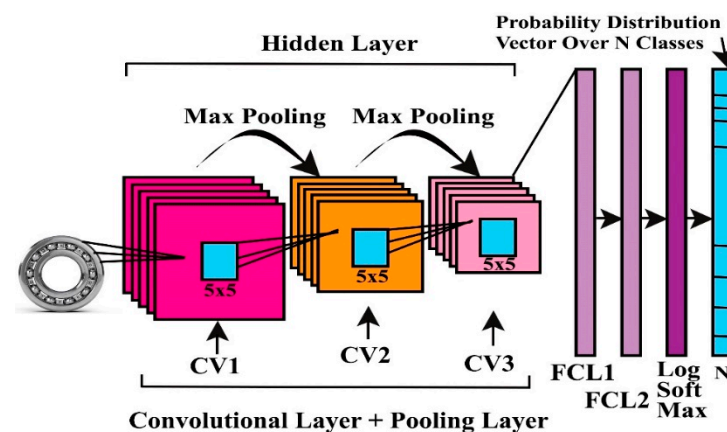
Table 3. Comparison of classical algorithms for fault bearing diagnosis.

Author	Year	Learning Method	Average Accuracy	Data Set
R. Zhang et al. [12]	2019	KSVD	70%	ABLT—a bearing life enhancement test bench
Khan and Kim [13]	2016	ANN-LBP histogram	100%	CRWU
Ankush Mehta et al. [14]	2021	KNN-SVM-LDA	90%	Experimental Setup
Zihan Chen. [15]	2017	2-stage matching pursuit	99.69%	CWRU
Ding and He. [16]	2017	EFMF-ConvNet	98.8%	CWRU
W. Zhang et al. [17]	2018	Residual learning algorithm	99.99%	CWRU
Luo and Hu [18]	2019	LSTM-NN	98%	CWRU
Wu et al. [19]	2017	KMC SVM	99.1%	CWRU
Tyagi and Panigrahi [20]	2017	ANN-SVM	97.9%	Experimental Setup
Fernández-Francos et al. [21]	2013	SVM	99% 100%	ISM and CWRU
Yadav et al. [22]	2013	LS-SVM	87%	3-Phase Squirrel cage induction
Deng et al. [23]	2016	IMASFD	97.67%	CWRU

5. Deep Learning Algorithms Used for Fault Bearing Diagnosis

5.1. Convolutional Neural Network (CNN)-Based Bearing Fault Diagnosis

The CNN is inspired by the animal cortex and was introduced in 1994 for detecting patterns from the input image to form a complex features map in a hierarchical way (Fukushima, 1980). It has an advantage over other learning algorithms when dealing with two-dimensional data, as it can autonomously learn the input data approximation through their layered architecture. Therefore, it is considered as an efficient and end-to-end learning system in which only a single objective function of a given model is to be optimized. The basic architecture of the CNN is given in Figure 9.

**Figure 9.** An architecture of convolutional neural network for rotating bearing fault diagnosis.

A group of researchers proposed an algorithm based on CNN, proposed in [24]. This work aimed to automate the feature extraction from the bearing signals using a CNN so that the overhead of feature extraction and selection from bearing data could be avoided. In [25], an adaptive hierarchical CNN equipped with a SoftMax classifier which can automatically learn salient information from the vibration acceleration signals was used for bearing fault diagnosis. The developed hierarchical adaptive network was composed of two layers, i.e., the first layer was to identify the type of bearing faults

and the second layer was to predict the severity of bearing faults. This model could adaptively vary the learning rate of the model during the training phase, which enhances the learning capability of the network significantly. Hence, the proposed model delivered high-classification accuracy when tested with the unseen data in the testing phase and was able to predict the fault severity effectively. Chen Lu et al. [26] proposed a health state classification-based intelligent fault bearing diagnosis method, which was proposed to use a hierarchical CNN, and it extracts features automatically from vibration signals. Meanwhile, in [27], researchers proposed an idea of feature extraction from the bearing data acquired through multiple sensors. The results of the proposed method suggest an enhancement in the classification accuracy, because it is believed that data from multiple sensors is enriched in salient information about the health status of the bearing as compared to the single sensor data. Thus, the above-discussed CNN-based methods present that they can achieve high and more reliable diagnosis performance. In [28], a multi-scale convolutional neural network (MS-DCNN) was proposed, and researchers proved that a multi-scale convolutional layer can expand and deepen the neural network for better learning, robust feature representation which reduces training time and network parameters and a reduced processing time. Furthermore, in [29], researchers proposed a method for bearing fault diagnosis that ends the manual feature extraction by deep CNN for automatic feature extraction and for adapting signal characteristics using the swarm optimization method. Ince et al. proposed a monitoring system with implementation on the CNN [30]. This method achieves high-level generalization and avoids the need for manual parameter tuning and hand-crafted feature extraction. They claimed that their proposed method does not need any form of transformation, feature extraction, and preprocessing. Their proposed method can directly access the raw data to evaluate the bearing fault diagnosis effectively. In [31], a method for monitoring bearing health was proposed. The proposed system fuses the feature extraction and classification blocks of the common fault detection approach into a single body at this state: the one-dimension convolutional neural network (1D-CNN) learns exact optimized features from the raw data with BP training when classification is performed by MLP layers. Wen et al., 2018 proposed a different method of jointed signal analysis and DNN for bearing fault diagnosis, and applied the S transform technique to obtain the time–frequency formulation of signals and developed a modified CNN network [32]. In 2019 [33], proposed a deep CNN method that combines the detailed convolution, the input gate structure of LSTM and the residual network for fault bearing diagnosis, which shows higher denoising ability. In [34], a scheme based on the CNN and the bi-spectrum analysis of the vibration signals was proposed. It is proposed that this method can be used for bearing diagnosis under variable speed conditions. In [35], a proposed method works on raw signals without any time-consuming hand-crafted feature extraction process, and it works well when working load changes and working under noisy environments. In [36], a new fault bearing diagnosis method was proposed by developing a signal-to-image conversion method by using a famous motor bearing dataset, a self-priming centrifugal pump dataset, and an axial piston hydraulic pump dataset. In [37], a method of using the CNN structure using 2D images for fault bearing diagnosis was proposed. In [38], an approach for fault bearing diagnosis using the 1D-CNN technique was proposed, and it added a preprocessing step in the diagnosis pipeline which calculates the frequency spectrum of vibration signals. Hao et al. proposed an end-to-end solution for fault bearing diagnosis with one-dimensional convolutional long short-term memory (1D-CLSTM) networks [39].

Further comparison of articles is discussed in Table 4. In these articles, the targeted faults of the bearings are outer raceways, inner raceways, ball/roller element fault, B fault, normal, damaged gear bearing, damaged bearing output shaft, motor current signals, and vibration signals with different levels of efficiency and better utilization of deep learning methods have been proposed.

Table 4. CNN methods used for fault bearing diagnosis research comparison.

Refs.	Year	Learning Method	Average Accuracy	Dataset
Guo et al. [25]	2016	ADCNN	99.3%	CWRU
Lu et al. [26]	2017	CNN	90%	QPZ-II
Xia et al. [27]	2017	CNN	99.89%	CWRU
Zilong and Wei [28]	2018	MS-DCNN	99.27%	CWRU
Fuan et al. [29]	2017	DCNN	100%	CWRU
Ince et al. [30]	2016	1D-CNN	97.4%	Real-time motor data
Eren et al. [31]	2017	1D-CNN	97%	IMS
Wen, Gao et al. [32]	2018	JCNN	99.94%	CWRU
Zhuang et al. [33]	2019	SRDCNN	95%	CWRU
Sohaib and Kim [34]	2019	CNN	90%	CWRU
W. Zhang et al. [35]	2018	TICNN	95.5%	CRWU
Wen et al. [36]	2017	LeNet-5 CNN	99.79% 99.481% 100%	Famous motor bearing dataset Self-priming centrifugal pump dataset Axial piston hydraulic pump dataset
Oh and Jeong [37]	2019	SRDCNN	95%	CWRU
Hasan et al. [38]	2019	CNN	90%	CWRU
Hao et al. [39]	2018	TICNN	95.5%	CRWU

5.2. Auto-Encoders-Based Bearing Fault Diagnosis

The unsupervised method of auto-encoders was first proposed in the year 1980 for the pre-training of an ANN [40,41]. It is defined as a broadly implemented greedy layer-wise neural network pre-training method. It is a unique neural network, since both its input and output are the same. This network learns itself. ANN trains an auto-encoder which consists of the encoder, bottleneck, decoder, and reconstruction loss. The encoder produces the new features representation from the old feature's representation. The bottleneck is a layer that contains the compressed representation of the input data, which is the lowest dimension of the input data. The decoder is the reverse of the encoder process and reconstruction loss is the method that measures the performance of the decoder and how close the output is to the original input. The output of the encoder is the input of the decoder. For imitating the input as a final output, the ANN takes the mean square error among the original input and output as the loss function and the decoder is released, while the encoder part remains. Classifiers can employ the output of the encoder in the feature representation stage. The general architecture of the encoder is illustrated in Figure 10.

AEs are trained by ANNs which comprise two parts, i.e., the encoder and decoder. Diverse research has been accomplished using AEs, including the first article which proposed a tool that diagnosed bearing faults with massive data, using five layers of auto-encoder from the frequency spectrum and effectively performed the classification of machines' health, in which the accuracy of 99.6% was achieved [42]. Furthermore, authors proposed the effective usage of the Gaussian kernel function and a deep auto-encoder network, resulting in effective bearing fault diagnosis [43]. In [44], two-layered faults bearing diagnosis was proposed: one is for the identification of the fault pattern in the rotatory bearing machine, and the second is for identification of the crack size in certain faults.

Article [45] states that fault bearing diagnosis using the capability of AEs and the high training speed of an Extreme Learning Machine (ELM) provided a better classification performance without explicit feature extraction. In [46], the feasibility of the Stacked Denoising Auto-encoder (SDAE)-based fault bearing diagnosis with the use of health state classification datasets from rolling bearings was proposed. In [47], a study on a fault recognizer based on the SDAE to denoise and extract features from the raw vibration

signals by stacking several denoising auto-encoders was proposed. In [48], a novel deep AE feature learning method for rotating fault diagnosis was developed. In [49], a multi-sensor feature fusion method for fault bearing diagnosis with SAE and DBN methods combination was proposed. A bearing fault diagnosis solution comparison with multiple techniques, proving that sparse auto-encoders are better and can be deployed in health, motors, and air compressors, was proposed in [50]. In [51], a new method was proposed that works on temporal vibration signals. WTA is used during the training stage to learn sparse features that are suitable for fault bearing diagnosis. Additionally, to obtain improvement in diagnosis result accuracy, soft voting method was applied. In [52], a method that uses AE sensors with big data simplifies the signals using STFT to transform raw signals from the time domain to the frequency domain for the generation of the spectrum matrix. This spectrum generates sub-patterns to obtain the optimized DL structure, and then the Large Memory Storage and Retrieval (LAMSTAR) network diagnoses the bearing fault as proposed. This also presents an effective use of the deep auto-encoders network in classification and feature extraction in fault bearing diagnosis, which minimizes the time consumption rate and maximizes the accuracy rate. The average highest accuracy is achieved in [46,48] of 100%, whereas other proposed solutions are also accurate and the best as per their strategies. The targeted faults of the bearing are the inner raceway, outer raceway, roller fault, normal, cage fault, vibration signals, eccentric fault, spalling fault, misalignment fault, and abrasion fault. A survey of the results achieved through deep auto-encoders used in fault bearing diagnosis of previous research is presented in Table 5.

Table 5. Auto-encoder methods used for fault bearing diagnosis research comparison.

Refs.	Year	Learning Method	Average Accuracy	Dataset
Feng Jia et al. [42]	2016	DNN-AE	99.6%	Rolling element bearing and Planetary Gearbox
Wang et al. [43]	2018	DNN-Gaussian radial basis kernel function and AE	86.75%	The aeroengine of aircraft
Sohaib et al. [44]	2017	SSAE-DNN	99.1%	CWRU
Mao et al. [45]	2017	AE-ELM	100%	CWRU
Lu et al. [46]	2017	SDA	84.01%	CWRU
Guo et al. [47]	2017	SDAE	100%	CWRU
Shao et al. [48]	2017	AFSA-SDAE	87.8%	Gearbox Electrical Locomotive roller bearing
Zhuyun Chen et al. [49]	2017	SAE-DBN	91.76%	Rolling element bearing
Verma et al. [50]	2013	SAE	97.22%	Air compressor
Chuanhao Li et al. [51]	2017	FC-WTA-AE	98.47%	CWRU
Fischer and Igel [52]	2017	LAMSTAR	96%	Bearing seeded

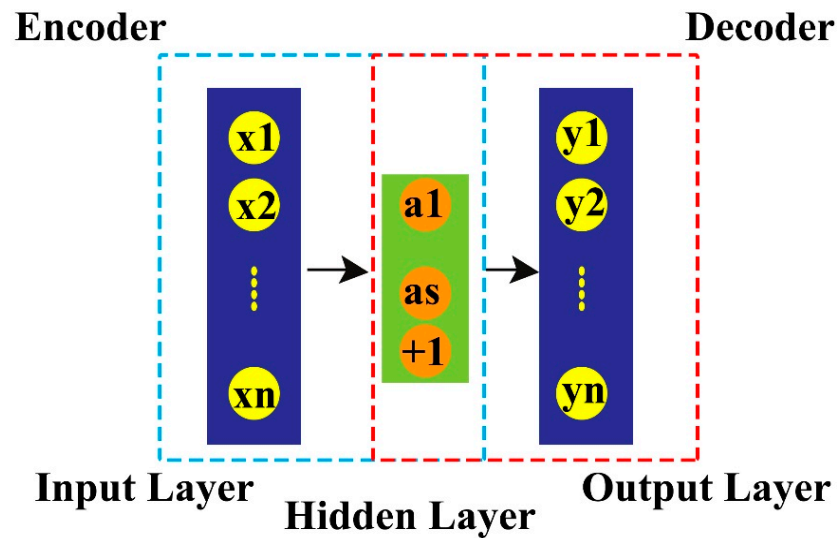


Figure 10. An architecture of an auto-encoder.

5.3. Deep Belief Network (DBN)-Based Methods for Bearing Fault Diagnosis

The DBN is a deep neural network that is constructed from various layers of RBMs—Restricted Boltzmann Machines [53]. Every RBM has layers of visible and hidden unit layers, and there is a connection between visible and hidden layers. The generic structure of the DBN is illustrated in Figure 11. There are multiple independent neurons in every layer. (h_1, h_2, h_3) are hidden layers, visible layer y , hidden layer x .

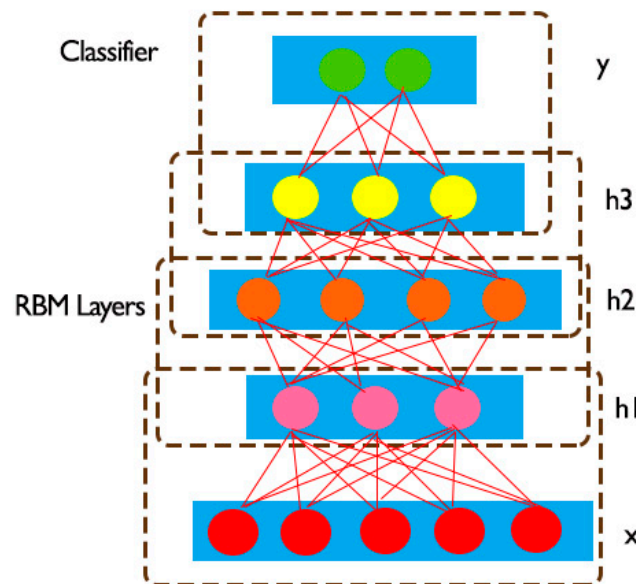


Figure 11. An architecture of deep belief network (DBN).

The process of DBN learning begins from the lowest visible layer. The process comprises two stages: firstly, the RBM layers are pre-trained in a greedy method step by step. In the second stage, fine-tuning of the complete network takes place for the parameter adjustment of the network so that better performance can be achieved. The input data approximation learned through the unsupervised training of the first RBM is inputted to train the next RBM, and this training continues until the last RBM has been trained and has learned the approximations successfully. Much research has been carried out in previous years using the DBN in the field of bearing fault diagnosis [54]. In this paper, a

new hierarchical bearing fault diagnosis method, NM, is adapted to the training process of the DBN to directly extract deep data features from signals of the frequency domain. In [55], an adaptive DBN with a dual-tree complex wavelet packet, which refines the measured vibration signals to design the original set of features, was proposed. Additionally, this method can recognize the different bearing faults. In [56], a bearing fault and severity diagnosis framework was proposed by binding many techniques together to obtain more accurate and capable bearing fault diagnostic algorithms. In [57], a method for rolling bearing fault diagnosis is proposed that has three steps: DBNs are constructed according to different hyperparameters, then IWV is used to determine every DBN's weight matrix, and then DBNs vote together to their respective matrix to obtain the final diagnosis result. In [58], a hierarchical diagnosis network for conducting the rolling bearing fault diagnosis was proposed, and researchers employed a wavelet packet transform representation of fault features and a DBN to classify/detect the type of fault. In [59], an ADBN was proposed that identifies the different conditions of bearing with DTCWPT, which measures vibration signals to design real feature sets.

The comparison of articles containing DBN-based methods is shown in Table 6. The targeted faults in these approaches were: health, inner raceway, outer raceways, normal, ball raceway, gear teeth breakage, broken bar, bowed rotor, stator winding defect, unbalanced motor, defecting bearing and roller fault.

Table 6. DBN-based methods used for fault bearing diagnosis research comparison.

Refs.	Year	Learning Method	Average Accuracy	Dataset
Shen et al. [54]	2019	HA-DBN	99.96%	Bearing Test rig
Tao et al. [55]	2016	DBN	94.73%	QPZ-II
Yu et al. [56]	2020	DBN-DS	99.69%	Qingdao University of Technology Bearing Fault Test rig
Liang et al. [57]	2018	DBN	84.2%	CWRU
Gan et al. [58]	2016	HDN-DBN	99.78%	CWRU
Shao et al. [59]	2017	Adaptive-DBN	96.89%	CWRU

5.4. Recurrent Neural Network (RNN)-Based Methodologies

An RNN processes input data in a recurrent manner. The architecture of an RNN is illustrated in Figure 12. The recurrent model can capture and model the sequential data or time-series data as the path goes from its hidden to the output layer. It is a generalized form of a Feed Forward Neural Network (FNN) that has internal memory. It gets to train with back propagation. An RNN is recurrent, as they perform a similar function for every input of data, whereas the output of the current input depends on previously considered computation. Through Time (BPTT) and a notorious gradient vanishing issue stemmed from its nature. To tackle this issue, LSTM is augmented by adding recurrent forget gates. LSTM is capable of modeling long-term dependency in data so it wins a dominant role in time series and text analysis and achieves success in natural language processing, video analysis, speech recognition, etc. In [60], a model based on the LSTM neural network was proposed. In [61], a data-driven method was proposed, with long-term time dependencies handled by this method; spatial and temporal dependencies can be utilized to detect faults based on the available sensor measurement signals for bearing fault detection. In [62], a technique for BLDC fault detection and diagnosis is presented. Additionally, the applications of these techniques to detect and accurately classify under non-stationary operating conditions is presented. Some of the work using this method is discussed below in Table 7.

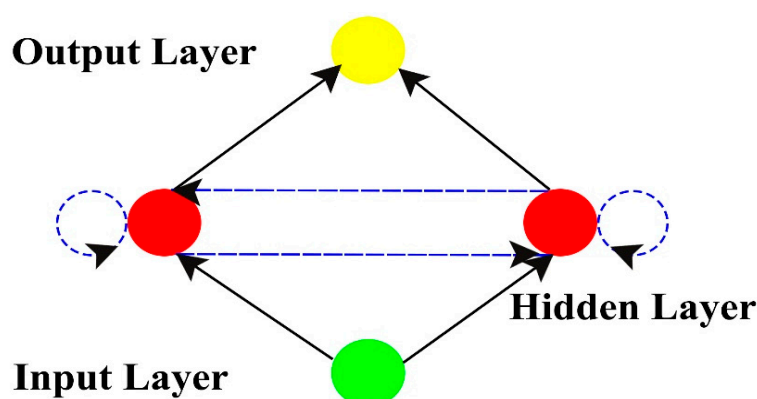


Figure 12. A recurrent neural network architecture.

Table 7. Six recurrent neural network-based methods used for bearing fault diagnosis research comparison.

Refs.	Year	Learning Method	Average Accuracy	Data Set
M. Li et al. [60]	2019	RNN-LSTM	98%	CWRU
Yang et al. [61]	2018	LSTM-RNN	99.9%	Wind Turbine Driven Train Diagnostic Simulator
Abed et al. [62]	2019	RNN	97%	Experimental Setup

5.5. Other Methods

There are various deeper learning methods for fault bearing diagnosis. Some of them are new approaches, and some are a mixture of previously discussed methods. In [63], an ensemble stack sparse auto-encoder space for fault bearing diagnosis was proposed. In [64], multiple wavelet fusion in a deep residual network with the help of two techniques, i.e., concatenation and maximization, is used to effectively capture useful information for bearing fault diagnosis. In [65], authors proposed a method of mapping original sound signals into time–frequency in the first step by STFT; then, SAE extracts the intrinsic fault features automatically. After that, SoftMax regression is used to recognize the fault modes of the feature vectors. In [66], a fault bearing diagnosis method was proposed while using all of the above mentioned methods, with four different preprocessing schemes. Similarly, in [67], a method of Dilated Residual Networks and DWWC to find a good set of features in fault diagnosis was proposed using the Planetary Gearbox dataset. Table 8 compares these methods.

Table 8. Other DL-based methods used for fault bearing diagnosis research comparison.

Refs.	Year	Learning Method	Average Accuracy	Data Set
J. He et al. [63]	2020	ESSAE	99.71%	CWRU
M. Zhao et al. [64]	2019	Multiple Wavelet Coefficients Fusion and deep residual network	96.29%	The rolling bearing test stand
Liu et al. [65]	2018	STFT-DL and Sound Signals	99.82	CWRU
Zhiqiang Chen et al. [66]	2017	DBM-DBN-Stack Auto-encoder	99%	Experimental Setup fabricated by Universidad Politecnica Salesiana Ecuador
M. Zhao et al. [67]	2018	DRN-DWWC	99.60%	Planetary Gearbox

6. Discussion

The considered studies for deep-learning-based fault diagnosis framework were mainly surveyed by considering the proposed methodology, and the reliability of the diagnostic performance. Almost all of the considered methods were developed based on publicly available datasets, which gives the freedom of reproducibility and scope of further analysis for further research in this field. It can be inferred that during the early stages of research in this field, researchers relied mostly upon engineered features and classical machine learning algorithms. However, as the research progressed in the field, more realistic methods have been considered by the researchers during fault diagnosis of the bearing. One of the realistic assumptions is the use of data that are collected under working conditions that are similar to the real-time environment of the industry. A few of the practicalities that can be considered during data collection are variable motor speed, variable motor load, presence of compound faults, and the presence of multiple fault severities. These variations constitute erratic working conditions of the machinery under examination, which makes the fault diagnosis process a challenging task. Based on the literature review, it is safe to say that the bearing fault diagnosis models developed using classical machine learning algorithms encounter deterioration in the fault diagnosis performance under erratic working conditions of the machinery. Therefore, under such circumstances, rather than the classical domain-dependent statistical feature analysis-based frameworks, deep-learning-based approaches establish the diagnosis approach as a general framework by improving the performance accuracies. Among the deep-learning-based approaches, the most popular techniques, such as the CNN, AE, DBN, RNN, DNN, SAE, etc., are efficiently utilized in rotatory machine fault diagnosis, which achieves higher accuracy than classical methods. By our survey, while conducting these experiments, CWR is nominated as the most considered dataset. However, for a real-world scenario, where the dataset is not acquired from the ideal conditions, there is still great opportunity to explore these established methods to make a more generalized and robust model for diagnosis.

6.1. Limitations

Classical Machine Learning

Despite the fact that machine learning algorithms have been widely used in the construction of a predictive maintenance mechanism, there are certain drawbacks. The purpose of developing predictive maintenance algorithms is to automatically detect and diagnose any issue in the equipment under observation. It is also necessary to detect faults in order to adopt an efficient equipment prognosis approach. The following are some of the limits of machine learning in the context of predictive maintenance [68–70].

1. Generalizability

Machine learning has a domain-specific implementation methodology. This means that the algorithm must be trained and fine-tuned separately for each type of application.

2. Domain-Related Knowledge

Expert knowledge of the problem domain is necessary when utilizing machine learning algorithms in predictive maintenance activities. In the machine-learning-based fault detection, diagnostic, and prognostic procedure, a feature engineering step is required. Feature engineering is a challenging process that necessitates a great deal of experience to develop handcrafted features that can structure the dataset. It can also detect a growth in fault.

3. Learning Ability, Reliability and Performance

Because machine learning methods require a simple network topology, such networks have limited learning capability. Shallow networks are the term used to describe these types of networks. In practice, the data used in data-driven predictive maintenance is noisy, nonlinear, and complicated. Machine learning algorithms cannot manage data with abnormalities, non-stationarity, or non-linearity, which is common with data from

industrial equipment. As a result, shallow networks are limited in their ability to provide data abstraction in the form of failure prediction features. As a result, when using real-time datasets for predictive maintenance, the overall performance of machine learning algorithms degrades.

4. Cross-Domain Analysis

In cross-domain applications, there is a lack of performance. Satisfactory performance is not guaranteed if the nature of the application becomes complex. The failure prediction data are used to guide maintenance operations.

6.2. Advantages and Disadvantages of Deep Learning

6.2.1. Advantages

1. The automated learning of structures from new data is the main benefit of using a deep learning system. The hierarchical order of nonlinear transformations makes it simple to extrapolate information from coarse data without the requirement for feature extraction and selection.
2. Because the overhead of feature engineering and selection is not required, developing condition monitoring, fault detection and diagnosis, and prognosis strategies for predictive maintenance is quite simple.
3. Transfer learning is better served by deep learning algorithms. It paves the way for cross-domain data-driven predictive maintenance solutions to be developed.
4. When compared to machine-learning-based predictive maintenance strategies, deep-learning-based predictive maintenance strategies have a higher generalization potential.
5. The bigger the number of layers and neurons in a deep learning network, the more complicated the problems can be that are conceived, resulting in a performance improvement.
6. The most appealing aspect of using deep learning in predictive maintenance is that these networks can automatically extract the relevant feature from data, obviating the need for manual feature engineering.
7. When deep learning is up to date, it can predict failures and cover every new event or behavior.

6.2.2. Disadvantages

1. To perform better than other strategies, it necessitates a big volume of data.
2. Because of the complicated data models, training is exceedingly costly. Deep learning also necessitates the use of pricey GPUs and hundreds of workstations. The users' costs will rise as a result of this.
3. Because it necessitates knowledge of topology, the training method, and other characteristics, there is no standard theory to aid you in choosing the correct deep learning tools. As a result, it is difficult for less skilled people to adopt it.
4. It is difficult to grasp output based just on learning, and therefore, this necessitates the use of classifiers. Such tasks are carried out using algorithms based on convolutional neural networks.

6.3. Comparison of Deep Learning Models

Table 9 presents the detailed comparison of deep learning-based models for fault diagnosis.

Table 9. Details of Deep Learning Models.

SL#	Model	Description	Pros	Cons
1	Deep Neural Network (DNN)	More than two layers are present. This allows for sophisticated non-linear relationships to be created. It is utilized for both classification and regression.	It is frequently utilized and has a high level of accuracy.	Because the error is propagated back to the previous one layer, the training process is not straightforward. The model's learning process is likewise far too slow.
2	Convolutional Neural Network (CNN)	With two-dimensional data, this network performs well. It is made up of convolutional filters that turn two-dimensional data into three-dimensional data.	Very good performance, and the model learns quickly.	For categorization, it requires a large amount of labeled data.
3	Recurrent Neural Network (RNN)	It has the ability to learn and remember sequences. All of the weights are shared throughout all of the stages and neurons.	LSTM, BLSTM, MDLSTM, and HLSTM are some of the versions that can learn sequential events and reflect time dependencies. These provide cutting-edge accuracy in speech recognition, character recognition, and a number of other natural language processing applications.	Due of gradient vanishing and the necessity for large datasets, there are numerous difficulties.
4	Deep Belief Network (DBN)	DBNs are probabilistic generative models that give a combined probability distribution across observable data and labels.	It addresses the problem of parameter selection, which can lead to poor local optima in some circumstances, and ensures that the network is properly established. Because the procedure is unsupervised, no tagged data are required. However, DBNs have a number of flaws, such as the high computational cost of training a DBN and the lack of clarity surrounding the processes for further network optimization based on maximum likelihood training approximation.	They do not account for the two-dimensional structure of an input image, which may significantly affect their performance and applicability in computer vision and multimedia analysis problems.
5	Auto-Encoders	Auto-encoders are a type of unsupervised learning technology in which neural networks are used to learn representations. We will create a neural network architecture in such a way that we force a compressed knowledge representation of the original input due to a bottleneck in the network.	They are particularly useful in feature extraction, since they can represent data as nonlinear representations.	An auto-encoder must be trained. Before you even start developing the real model, that is a lot of data, processing time, hyper parameter adjustment, and model validation. Instead of capturing as much information as possible, an auto-encoder learns to capture as much relevant information as feasible.
6	Deep Boltzmann Machine (DBMs)	The DBM has entirely undirected connections, whereas the top two layers constitute an undirected graphical model and the lower layers form a directed generative model. Units in odd-numbered levels are conditionally independent on units in even-numbered layers, and vice versa, in DBMs with several layers of hidden units.	They can capture multiple layers of complicated input data representations and are suitable for unsupervised learning, since they can be trained on unlabeled data, but they can also be fine-tuned for a specific job in a supervised manner.	One of the most significant is the high computing cost of inference, which makes collaborative optimization on large datasets nearly impossible. Several strategies for improving the effectiveness of DBMs have been presented. These include employing distinct models to initialize the values of the hidden units in all layers to speed up inference.

6.4. Future Perspectives of Deep Learning

Deep-learning-based predictive maintenance still has room for improvement. In the next subsections, some of the limitations of deep learning algorithms in terms of predictive maintenance are discussed.

6.4.1. Enhanced Generalization

Although advanced deep learning approaches such as fine-tune transfer learning and multitask learning have added a feeling of generality to data-driven predictive maintenance tactics, these concepts still need to be investigated further. Domain-independent data-driven predictive maintenance can be implemented using notions such as these.

6.4.2. Explain-Ability

Deep learning's data processing and exploration capabilities are unquestionably superior to machine learning's. Its application in predictive maintenance has eliminated a lot of the overhead and difficulties that traditional machine learning techniques had. To name a few advantages, it can readily handle large amounts of data and can learn important information from inputs without the need for a domain-specific feature engineering process. Deep learning algorithms, on the other hand, are more like a black box than expanded capability. There is currently no comprehensive explanation for how deep learning algorithms correctly simulate complicated, nonlinear, and nonstationary data in an abstract manner. Furthermore, it is not known how the estimated codes, also known as features, perform better than their predecessors in terms of predictive maintenance. There is a need for explainable deep-learning-based predictive maintenance strategies.

6.4.3. Multimodal and Multisensor Data Fusion

Data fusion from numerous sensors and modalities is an intriguing and viable extension of data-driven predictive maintenance based on deep learning. Data fusion can offer detailed information about bearing faults, which can help improve bearing fault detection models. Data fusion from many sensors is also a practical aspect, as multiple sensors are typically mounted on the concerned component to collect data for better performance.

7. Conclusions

In this paper, we investigated the applications of deep learning algorithms for bearing fault diagnosis. In most of the studies, researchers like to rely on the publicly available datasets due to the easier availability, and ideal working conditions. From the performance analysis of the considered studies, we saw that the deep learning algorithms are highly capable of learning the health characteristics automatically, and the diagnostic performance has significantly been improved. Furthermore, the analysis indicates that the accuracy of many improved deep-learning-based methods can improve comparatively through more training, which gives an idea for the exploration and new work to be carried out for intelligent fault bearing diagnosis. However, it should be considered that the successes of deep learning-based diagnosis models still rely on some kind of domain-based analysis, and are subject to sufficient labeled samples. Therefore, this review is anticipated to scientifically present the development and progress of a deep-learning-based bearing fault diagnosis framework and deliver valuable guidelines for future research.

Author Contributions: Conceptualization, S.M. and M.S.; methodology, S.M., M.S. and M.M.M.I.; validation, S.M., M.S. and M.M.M.I.; formal analysis, S.M., M.S. and M.M.M.I.; investigation, M.S. and M.M.M.I.; data curation, S.M. and M.S.; writing—original draft preparation, S.M., M.S. and M.M.M.I.; writing—review and editing, M.S. and M.M.M.I.; visualization, S.M. and M.S.; supervision, M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

List of Symbols

AE	Auto-encoder
AE-DNN	Auto-encoder Deep Neural Network
AFSA	Artificial Fish Swarm Algorithm
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DCNN	Deep Convolutional Neural Network
DL	Deep Learning
DRN	Deep Residual Networks
DT-CWT	Dual tree- Complex Wavelet Transform
DWWC	Dynamical Weighted Wavelet Connected
DWT	Discrete Wavelet Transform
ELM	Extreme Learning Machine
FC	Fully Connected Layer
FFT	Fast Fourier Transform
FT	Fourier Transform
GA-PSO	Genetic Algorithms-Practical Swarm Optimization
ICA	Independent Component Analysis
IMS	Intelligent Maintenance System
KSVD	K means singular value decomposition
KNN	k-Nearest Neighbor
LAMSTAR	Large Memory Storage and Retrieval
LSTM	Long Short-Term Memory
ML	Machine Learning
PCA	Principal Component Analysis
PSD	Photoshop
RBM	Restricted Boltzmann Machine
REB's	Rolling Elements Bearings
RNN	Recurrent Neural Network
RMS	Root Mean Square Value
RPM	Revolutions per minute
RTD	Resistance Temperature Detector
SDAE	Stacked Denoising Auto-encoder
STFT	Short -time Fourier Transform
S-Transform	Stock well Transform
SVM	Support Vector Machine
WPT	Wavelet Packet Transform
WTA	Winner Take All Auto-encoder

References

1. Saidur, R. A review on electrical motors energy use and energy savings. *Renew. Sustain. Energy Rev.* **2010**, *14*, 877–898. [CrossRef]
2. Allied Market Research, Electric Motors Market Overview. Available online: <https://www.alliedmarketresearch.com/electric-motor-market> (accessed on 16 July 2021).
3. Jian, Y.; Qing, X.; He, L.; Zhao, Y.; Qi, X.; Du, M. Fault diagnosis of motor bearing based on deep learning. *Adv. Mech. Eng.* **2019**, *11*, 1687814019875620. [CrossRef]
4. Shao, H.; Jiang, H.; Zhang, X.; Niu, M. Rolling bearing fault diagnosis using an optimization deep belief network. *Meas. Sci. Technol.* **2015**, *26*, 115002. [CrossRef]
5. Nandi, S.; Toliyat, H.A.; Li, X. Condition monitoring and fault diagnosis of electrical motors—A review. *IEEE Trans. Energy Convers.* **2005**, *20*, 719–729. [CrossRef]
6. Case Western Reserve University (CWRU) Bearing Data Center. Available online: <https://csegroups.case.edu/bearingdatacenter/pages/download-data-file> (accessed on 16 July 2021).
7. Lessmeier, C.; Kimotho, J.K.; Zimmer, D.; Sextro, W. KAT-DataCenter. Available online: <https://mb.uni-paderborn.de/kat/forschung/datacenter/bearing-datacenter> (accessed on 16 July 2021).

8. Nectoux, P.; Gouriveau, R.; Medjaher, K.; Ramasso, E.; Chebel-Morello, B.; Nouredine Zerhouni, C.V. Pronostia: An experimental platform for bearings accelerated life test. In Proceedings of the IEEE Conference on Prognostics and Health Management, Denver, CO, USA, 18–21 June 2012; p. 18.
9. Qiu, H.; Lee, J.; Lin, J. Wavelet Filter-based Weak Signature Detection Method and its Application on Roller Bearing Prognostics. *J. Sound Vib.* **2006**, *289*, 1066–1090. [[CrossRef](#)]
10. Kankar, P.; Sharma, S.C.; Harsha, S. Fault diagnosis of ball bearings using machine learning methods. *Expert Syst. Appl.* **2011**, *38*, 1876–1886. [[CrossRef](#)]
11. Qiu, H.; Lee, J.; Lin, J.; Yu, G. Robust performance degradation assessment methods for enhanced rolling element bearing prognostics. *Adv. Eng. Inform.* **2003**, *17*, 127–140. [[CrossRef](#)]
12. Zhang, R.; Fang, Y.; Zhou, Z. Fault diagnosis of rolling bearing based on k-SVD dictionary learning algorithm and BP Neural Network. In Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence, Shanghai, China, 20–22 September 2019.
13. Khan, S.A.; Kim, J.M. Automated Bearing Fault Diagnosis Using 2D Analysis of Vibration Acceleration Signals under Variable Speed Conditions. *Shock Vib.* **2016**, *2016*, 1–11. [[CrossRef](#)]
14. Mehta, A.; Goyal, D.; Choudhary, A.; Pabla, B.S.; Belghith, S. Machine Learning-Based Fault Diagnosis of Self-Aligning Bearings for Rotating Machinery Using Infrared Thermography. *Math. Probl. Eng.* **2021**, *2021*, 9947300. [[CrossRef](#)]
15. Chen, Z. Bearing fault diagnosis with compressed data based on two-stage matching pursuit. In Proceedings of the 2017 Prognostics and System Health Management Conference, Harbin, China, 9–12 July 2017.
16. Ding, X.; He, Q. Energy-Fluctuated Multiscale Feature Learning with Deep ConvNet for Intelligent Spindle Bearing Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 1926–1935. [[CrossRef](#)]
17. Zhang, W.; Li, X.; Ding, Q. Deep residual learning-based fault diagnosis method for rotating machinery. *ISA Trans.* **2019**, *95*, 295–305. [[CrossRef](#)]
18. Luo, P.; Hu, Y. Research on Rolling Bearing Fault Identification Method Based on LSTM Neural Network. In Proceedings of the 2018 the 6th International Conference on Mechanical Engineering, Materials Science and Civil Engineering, Xiamen, China, 21–22 December 2018.
19. Wu, C.; Chen, T.; Jiang, R. Bearing fault diagnosis via kernel matrix construction based support vector machine. *J. Vibroeng.* **2017**, *19*, 3445–3461. [[CrossRef](#)]
20. Tyagi, S.; Panigrahi, S.K. A DWT and SVM based method for rolling element bearing fault diagnosis and its comparison with Artificial Neural Networks. *J. Appl. Comput. Mech.* **2017**, *3*, 80–91. [[CrossRef](#)]
21. Fernández-Francos, D.; Martínez-Rego, D.; Fontenla-Romero, O.; Alonso-Betanzos, A. Automatic bearing fault diagnosis based on one-class m-SVM. *Comput. Ind. Eng.* **2013**, *64*, 357–365. [[CrossRef](#)]
22. Yadav, O.P.; Joshi, D.; Pahuja, G.L. Support Vector Machine based Bearing Fault Detection of Induction Motor. *Indian J. Adv. Electron. Eng.* **2013**, *1*, 34–39.
23. Deng, W.; Li, X.; Zhao, H. Study on A Fault Diagnosis Method of Rolling Element Bearing Based on Improved ACO and SVM Model. *Int. J. Future Gener. Commun. Netw.* **2016**, *9*, 167–180. [[CrossRef](#)]
24. Janssens, O.; Slavković, V.; Vervisch, B.; Stockman, K.; Loccufer, M.; Verstockt, S.; van de Walle, R.; van Hoecke, S. Convolutional Neural Network Based Fault Detection for Rotating Machinery. *J. Sound Vib.* **2016**, *377*, 331–345. [[CrossRef](#)]
25. Guo, X.; Chen, L.; Shen, C. Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Meas. J. Int. Meas. Confed.* **2016**, *93*, 490–502. [[CrossRef](#)]
26. Lu, C.; Wang, Z.; Zhou, B. Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. *Adv. Eng. Inform.* **2017**, *32*, 139–151. [[CrossRef](#)]
27. Xia, M.; Li, T.; Xu, L.; Liu, L.; de Silva, C.W. Fault Diagnosis for Rotating Machinery Using Multiple Sensors and Convolutional Neural Networks. *IEEE/ASME Trans. Mechatron.* **2018**, *23*, 101–110. [[CrossRef](#)]
28. Zilong, Z.; Wei, Q. Intelligent fault diagnosis of rolling bearing using one-dimensional multi-scale deep convolutional neural network based health state classification. In Proceedings of the ICNSC 2018—15th IEEE International Conference on Networking, Sensing and Control, Zhuhai, China, 27–29 March 2018.
29. Fuan, W.; Hongkai, J.; Haidong, S.; Wenjing, D.; Shuaipeng, W. An adaptive deep convolutional neural network for rolling bearing fault diagnosis. *Meas. Sci. Technol.* **2017**, *28*, 095005. [[CrossRef](#)]
30. Ince, T.; Kiranyaz, S.; Eren, L.; Askar, M.; Gabbouj, M. Real-Time Motor Fault Detection by 1-D Convolutional Neural Networks. *IEEE Trans. Ind. Electron.* **2016**, *63*, 7067–7075. [[CrossRef](#)]
31. Eren, L. Bearing fault detection by one-dimensional convolutional neural networks. *Math. Probl. Eng.* **2017**, *2017*, 1–9. [[CrossRef](#)]
32. Wen, L.; Gao, L.; Li, X.; Wang, L.; Zhu, J. A Jointed Signal Analysis and Convolutional Neural Network Method for Fault Diagnosis. *Procedia CIRP* **2018**, *72*, 1084–1087. [[CrossRef](#)]
33. Zhuang, Z.; Lv, H.; Xu, J.; Huang, Z.; Qin, W. A deep learning method for bearing fault diagnosis through stacked residual dilated convolutions. *Appl. Sci.* **2019**, *9*, 1823. [[CrossRef](#)]
34. Sohaib, M.; Kim, J.-M. Fault diagnosis of rotary machine bearings under inconsistent working conditions. *IEEE Trans. Instrum. Meas.* **2019**, *69*, 3334–3347. [[CrossRef](#)]
35. Zhang, W.; Li, C.; Peng, G.; Chen, Y.; Zhang, Z. A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mech. Syst. Signal Process.* **2018**, *100*, 439–453. [[CrossRef](#)]

36. Wen, L.; Li, X.; Gao, L.; Zhang, Y. A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method. *IEEE Trans. Ind. Electron.* **2018**, *65*, 5990–5998. [[CrossRef](#)]
37. Oh, J.W.; Jeong, J. Convolutional neural network and 2-D image based fault diagnosis of bearing without retraining. In Proceedings of the 2019 the 3rd International Conference on Compute and Data Analysis, Kahului, HI, USA, 14–17 March 2019.
38. Zhao, C.; Sun, J.; Lin, S.; Peng, Y. Fault Diagnosis Method for Rolling Mill Multi Row Bearings Based on AMVMD-MC1DCNN under Unbalanced Dataset. *Sensors* **2021**, *21*, 5494. [[CrossRef](#)]
39. Hao, S.; Ge, F.X.; Li, Y.; Jiang, J. Multisensor bearing fault diagnosis based on one-dimensional convolutional long short-term memory networks. *Meas. J. Int. Meas. Confed.* **2020**, *159*, 107802. [[CrossRef](#)]
40. Schmidhuber, J. Deep Learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)] [[PubMed](#)]
41. Ballard, D.H. Modular Learning in Neural Networks. *AAAI* **1987**, *647*, 279–284.
42. Jia, F.; Lei, Y.; Lin, J.; Zhou, X.; Lu, N. Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. *Mech. Syst. Signal Process.* **2016**, *72–73*, 303–315. [[CrossRef](#)]
43. Wang, F.; Dun, B.; Deng, G.; Li, H.; Han, Q. A deep neural network based on kernel function and auto-encoder for bearing fault diagnosis. In Proceedings of the I2MTC 2018—2018 IEEE International Instrumentation and Measurement Technology Conference: Discovering New Horizons in Instrumentation and Measurement, Houston, TX, USA, 14–17 May 2018.
44. Islam, M.M.M.; Kim, J.-M. Automated bearing fault diagnosis scheme using 2D representation of wavelet packet transform and deep convolutional neural network. *Comput. Ind.* **2019**, *106*, 142–153. [[CrossRef](#)]
45. Mao, W.; He, J.; Li, Y.; Yan, Y. Bearing fault diagnosis with auto-encoder extreme learning machine: A comparative study. *Proc. Inst. Mech. Eng. Part C J. Mech. Eng. Sci.* **2016**, *231*, 1560–1578. [[CrossRef](#)]
46. Lu, C.; Wang, Z.Y.; Qin, W.L.; Ma, J. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. *Signal Process* **2017**, *130*, 377–388. [[CrossRef](#)]
47. Guo, X.; Shen, C.; Chen, L. Deep fault recognizer: An integrated model to denoise and extract features for fault diagnosis in rotating machinery. *Appl. Sci.* **2017**, *7*, 41. [[CrossRef](#)]
48. Shao, H.; Jiang, H.; Zhao, H.; Wang, F. A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mech. Syst. Signal Process* **2017**, *95*, 187–204. [[CrossRef](#)]
49. Chen, Z.; Li, W. Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 1693–1702. [[CrossRef](#)]
50. Verma, N.K.; Gupta, V.K.; Sharma, M.; Sevakula, R.K. Intelligent condition based monitoring of rotating machines using sparse auto-encoders. In Proceedings of the PHM 2013—2013 IEEE International Conference on Prognostics and Health Management, Gaithersburg, MD, USA, 24–27 June 2013.
51. Li, C.; Zhang, W.; Peng, G.; Liu, S. Bearing Fault Diagnosis Using Fully-Connected Winner-Take-All Autoencoder. *IEEE Access* **2017**, *6*, 6103–6115. [[CrossRef](#)]
52. Fischer, A.; Igel, C. An introduction to restricted Boltzmann machines. In *Iberoamerican Congress on Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2012.
53. Sohaib, M.; Kim, C.-H.; Kim, J.-M. A hybrid feature model and deep-learning-based bearing fault diagnosis. *Sensors* **2017**, *17*, 2876. [[CrossRef](#)]
54. Shen, C.; Xie, J.; Wang, D.; Jiang, X.; Shi, J.; Zhu, Z. Improved hierarchical adaptive deep belief network for bearing fault diagnosis. *Appl. Sci.* **2019**, *9*, 3374. [[CrossRef](#)]
55. Tao, J.; Liu, Y.; Yang, D. Bearing Fault Diagnosis Based on Deep Belief Network and Multisensor Information Fusion. *Shock Vib.* **2016**, *2016*, 1–9. [[CrossRef](#)]
56. Yu, K.; Lin, T.R.; Tan, J. A bearing fault and severity diagnostic technique using adaptive deep belief networks and Dempster-Shafer theory. *Struct. Health Monit.* **2020**, *19*, 240–261. [[CrossRef](#)]
57. Liang, T.; Wu, S.; Duan, W.; Zhang, R. Bearing fault diagnosis based on improved ensemble learning and deep belief network. *J. Phys. Conf. Ser.* **2018**, *1074*, 012154. [[CrossRef](#)]
58. Gan, M.; Wang, C.; Zhu, C. Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings. *Mech. Syst. Signal Process.* **2016**, *72–73*, 92–104. [[CrossRef](#)]
59. Shao, H.; Jiang, H.; Wang, F.; Wang, Y. Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet. *ISA Trans.* **2017**, *69*, 187–201. [[CrossRef](#)]
60. Li, M.; Wei, Q.; Wang, H.; Zhang, X. Research on fault diagnosis of time-domain vibration signal based on convolutional neural networks. *Syst. Sci. Control Eng.* **2019**, *7*, 73–81. [[CrossRef](#)]
61. Yang, R.; Huang, M.; Lu, Q.; Zhong, M. Rotating Machinery Fault Diagnosis Using Long-short-term Memory Recurrent Neural Network. *IFAC-PapersOnLine* **2018**, *51*, 228–232. [[CrossRef](#)]
62. Abed, W.; Sharma, S.; Sutton, R.; Motwani, A. A Robust Bearing Fault Detection and Diagnosis Technique for Brushless DC Motors Under Non-stationary Operating Conditions. *J. Control. Autom. Electr. Syst.* **2015**, *26*, 241–254. [[CrossRef](#)]
63. He, J.; Ouyang, M.; Yong, C.; Chen, D.; Guo, J.; Zhou, Y. A novel intelligent fault diagnosis method for rolling bearing based on integrated weight strategy features learning. *Sensors* **2020**, *20*, 1774. [[CrossRef](#)] [[PubMed](#)]
64. Zhao, M.; Kang, M.; Tang, B.; Pecht, M. Multiple Wavelet Coefficients Fusion in Deep Residual Networks for Fault Diagnosis. *IEEE Trans. Ind. Electron.* **2019**, *66*, 4696–4706. [[CrossRef](#)]

65. Liu, H.; Li, L.; Ma, J. Rolling Bearing Fault Diagnosis Based on STFT-Deep Learning and Sound Signals. *Shock Vib.* **2016**, *2016*, 1–12. [[CrossRef](#)]
66. Chen, Z.; Deng, S.; Chen, X.; Li, C.; Sanchez, R.V.; Qin, H. Deep neural networks-based rolling bearing fault diagnosis. *Microelectron. Reliab.* **2017**, *75*, 327–333. [[CrossRef](#)]
67. Zhao, M.; Kang, M.; Tang, B.; Pecht, M. Deep Residual Networks with Dynamically Weighted Wavelet Coefficients for Fault Diagnosis of Planetary Gearboxes. *IEEE Trans. Ind. Electron.* **2018**, *65*, 4290–4300. [[CrossRef](#)]
68. Çınar, Z.M.; Nuhu, A.A.; Zeeshan, Q.; Korhan, O.; Asmael, M.; Safaei, B. Machine Learning in Predictive Maintenance towards Sustainable Smart Manufacturing in Industry 4.0. *Sustainability* **2020**, *12*, 8211. [[CrossRef](#)]
69. Lv, F.; Wen, C.; Bao, Z.; Liu, M. Fault diagnosis based on deep learning. In Proceedings of the 2016 American Control Conference (ACC), Boston, MA, USA, 6–8 July 2016. [[CrossRef](#)]
70. Hasan, M.J.; Sohaib, M.; Kim, J.M. 1D CNN-based transfer learning model for bearing fault diagnosis under variable working conditions. In *International Conference on Computational Intelligence in Information System*; Springer: Cham, Switzerland, 2019. [[CrossRef](#)]