Deep Learning Algorithms for Bearing Fault Diagnostics – A Comprehensive Review

Zhang, Shen; Zhang, Shibo; Wang, Bingnan; Habetler, Thomas

Abstract

In this survey paper, we systematically summarize existing literature on bearing fault diagnostics with deep learning (DL) algorithms. While conventional machine learning (ML) methods, including artificial neural network, principal component analysis, support vector machines, etc., have been successfully applied to the detection and categorization of bearing faults for decades, recent developments in DL algorithms in the last five years have sparked renewed interest in both industry and academia for intelligent machine health monitoring. In this paper, we first provide a brief review of conventional ML methods, before taking a deep dive into the state-of-the-art DL algorithms for bearing fault applications. Specifically, the superiority of DL based methods are analyzed in terms of fault feature extraction and classification performances; many new functionalities enabled by DL techniques are also summarized. In addition, to obtain a more intuitive insight, a comparative study is conducted on the classification accuracy of different algorithms utilizing the open source Case Western Reserve University (CWRU) bearing dataset. Finally, to facilitate the transition on applying various DL algorithms to bearing fault diagnostics, detailed recommendations and suggestions are provided for specific application conditions. Future research directions to further enhance the performance of DL algorithms on health monitoring are also discussed.

IEEE Access

Deep Learning Algorithms for Bearing Fault Diagnostics – A Comprehensive Review

SHEN ZHANG, *Student Member, IEEE*, SHIBO ZHANG, *Student Member, IEEE*, BINGNAN WANG, *Senior Member, IEEE*, and THOMAS G. HABETLER, *Fellow, IEEE*

Abstract—In this survey paper, we systematically summarize existing literature on bearing fault diagnostics with deep learning (DL) algorithms. While conventional machine learning (ML) methods, including artificial neural network, principal component analysis, support vector machines, etc., have been successfully applied to the detection and categorization of bearing faults for decades, recent developments in DL algorithms in the last five years have sparked renewed interest in both industry and academia for intelligent machine health monitoring. In this paper, we first provide a brief review of conventional ML methods, before taking a deep dive into the state-of-the-art DL algorithms for bearing fault applications. Specifically, the superiority of DL based methods are analyzed in terms of fault feature extraction and classification performances; many new functionalities enabled by DL techniques are also summarized. In addition, to obtain a more intuitive insight, a comparative study is conducted on the classification accuracy of different algorithms utilizing the open source Case Western Reserve University (CWRU) bearing dataset. Finally, to facilitate the transition on applying various DL algorithms to bearing fault diagnostics, detailed recommendations and suggestions are provided for specific application conditions. Future research directions to further enhance the performance of DL algorithms on health monitoring are also discussed.

Bearing fault, deep learning, diagnostics, feature extraction, machine learning.

I. INTRODUCTION

Electric machines are widely employed in a variety of industry applications and electrified transportation systems. For certain applications these machines may operate under unfavorable conditions, such as high ambient temperature, high moisture and overload, which can eventually result in motor malfunctions that lead to high maintenance costs, severe financial losses, and safety hazards [1]–[3]. The malfunction of electric machines can be generally attributed to various faults of different categories, including drive inverter failures, stator winding insulation breakdown, bearing faults and air gap eccentricity. Several surveys regarding the likelihood of
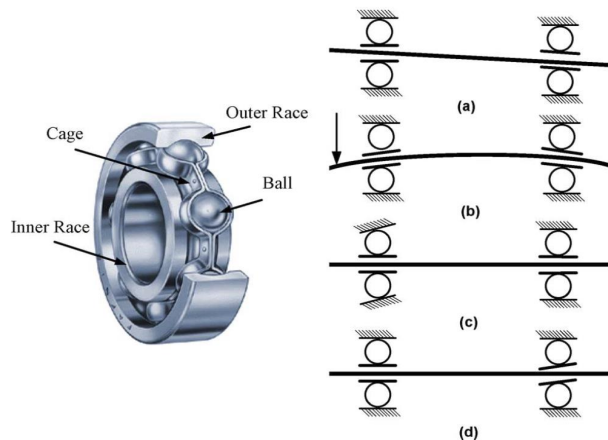
Fig. 1. Structure of a rolling-element bearing with four types of common scenarios of misalignment that are likely to cause bearing failures: (a) misalignment (out-of-line), (b) shaft deflection, (c) crooked or tilted outer race and (d) crooked or tilted inner race [8].

induction machine failures conducted by the IEEE Industry Application Society (IEEE-IAS) [4]–[6] and the Japan Electrical Manufacturers' Association (JEMA) [7] reveal that bearing fault is the most common fault type and is responsible for 30% to 40% of all the machine failures.

The structure of a rolling-element bearing is illustrated in Fig. 1, which contains the outer race typically mounted on the motor cap, the inner race to hold the motor shaft, the balls or the rolling elements, and the cage for restraining the relative distances between adjacent rolling elements [8]. The four common scenarios of misalignment that are likely to cause bearing failures are demonstrated in Fig. 1(a) to (d). Since bearing is the most vulnerable component in a motor drive system, accurate bearing fault diagnostics has been a research frontier for engineers and scientists for the past decades. Specifically, this problem has been approached by developing a physical model of bearing faults, and understanding the relationship between bearing faults and measurable signals, which can be captured by a variety of sensors and analyzed with signal processing techniques. Sensing modalities that have been explored include vibration [9], [10], acoustic noise [11], [12], stator current [13], [14], thermal-imaging [15], and multiple sensor fusion [16], among which vibration analysis is the most dominant. The existence of a bearing fault as well as its specific fault type can be readily determined by performing frequency spectral analysis on the monitored signals and

analyzing their components at characteristic fault frequencies, which can be calculated by a well-defined mechanical model [8] that depends on the motor speed, the bearing geometry and the specific location of the bearing defect.

However, accurately identifying the presence of a bearing fault can be challenging in practice, especially when the fault is still at its incipient stage and the signal-to-noise ratio of the monitored signal is small. In addition, unlike other motor failures (stator inter-turn, broken rotor bar, etc. [3]) that can be accurately determined by electric signals, the uniqueness of a bearing failure lies in its multi-physics nature. It is the primary mechanical vibration due to the bearing defect that triggered the abnormal electric signal, which further influences the output torque, the motor speed, and finally the bearing vibration pattern itself, whose fault frequency is directly proportional to the motor speed. Furthermore, the accuracy of the traditional physical model-based vibration analysis can be further affected by background noise due to external motion and vibration, and its sensitivity is also subject to change with respect to sensor mounting positions and spatial constraints in a highly-compact environment. Therefore, instead of vibration analysis, a popular alternative approach is to analyze the stator current signal [13], [14], which has already been measured in motor drives to regulate the motor's torque and speed, and thus it would not bring extra device or installation costs.

Despite its advantages such as economic savings and simple implementation, the motor current signature analysis (MCSA) can encounter many practical issues. For example, the magnitude of stator currents at the bearing fault frequency can vary at different loads, different speeds, and different power ratings of the motors themselves, thus bringing challenges to identify a universal threshold of the stator current to trigger a fault alarm at an arbitrary operating condition. Therefore, a thorough and systematic commissioning stage is usually required while the motor is still at the healthy condition, and the healthy data would be collected while the target motor is running at different loads and speeds. However, this process, summarized as a "Learning Stage" in patent US5726905 [17], can be tedious and expensive to perform, and needs to be repeated for any new motor with a different power rating.

Most of the challenges described above can be attributed to the fact that all of the conventional model-based methods rely solely upon the threshold value of different signals (data) at the fault frequencies to determine the presence of a bearing fault. These models can only describe the signal features of a few well-defined fault types, while in reality the naturally occurring faults are often more complicated. For example, at the early stage of a fault the signatures can be less well-defined or even not traceable by using the physical models; more than one faults can occur at the same time, which potentially modifies the fault features and creates new features due to the the coupling effect. Therefore, there may exist many unique features or patterns hidden in the data themselves that can potentially reveal a bearing fault, and it is almost impossible for humans to identify these convoluted features through manual observation or interpretation. Therefore, many researchers have applied various machine learning (ML) algorithms, including artificial neural networks (ANN), principal

component analysis (PCA), support vector machines (SVM), etc., to parse the data, learn from them, and apply what they have learned to make intelligent decisions regarding the presence of bearing faults [18]–[21]. Most of the literature applying these ML algorithms report satisfactory results with classification accuracy over 90%.

To achieve an even better performance at versatile operating conditions and noisy environments, deep learning (DL) based methods are becoming increasingly popular to meet this demand [22]–[25]. This literature survey incorporates more than 180 papers dedicated to bearing fault diagnosis, around 80 of which employed some type of DL approaches. The number of papers also grows exponentially over the recent years, indicating a booming interest in employing DL methods for bearing fault diagnostics.

In this context, this paper seeks to present a thorough overview on the recent research work devoted to applying ML and DL techniques on bearing fault diagnostics. The rest of the paper is organized as follows. In Section II, we introduce some of the most popular datasets used for bearing fault detection. Next, in Section III, we look into some traditional ML methods, including ANN, PCA, k-nearest neighbors (k-NN), SVM, etc., with a brief overview of major publications applying each ML algorithm for bearing fault detection. For the main part of this paper, in Section IV, we take a deep dive into the research frontier of DL based bearing fault identification. In this section, we will provide our understanding of the research trend toward DL approaches. Specifically, we will discuss the advantages of DL based methods over the conventional ML methods in terms of fault feature extraction and classifier performance, as well as new functionalities offered by DL techniques that cannot be accomplished before. We will also provide a detailed analysis to each of the major DL techniques, including convolutional neural network (CNN), auto-encoder (AE), deep belief network (DBN), recurrent neural network (RNN), generative adversarial network (GAN), and their applications in bearing fault detection. In Section V, a comparative study is conducted on different DL algorithms to offer a more intuitive insight, which compared the classifier performance utilizing the popular open source "Case Western Reserve University (CWRU) bearing dataset". Finally in Section VI, detailed recommendations and suggestions are provided regrading the selection of specific DL algorithms for specific application scenarios, such as the setup environment, the data size, and the number of sensors and sensor types. Future research directions are also discussed to further improve the classifier accuracy, and facilitate domain adaptation and technology transfer from laboratories to the real-world.

II. POPULAR BEARING FAULT DATASETS

Data is the foundation for all of the ML methods. To develop effective ML and DL algorithms for bearing fault detection, a good collection of datasets is necessary. Since the natural bearing degradation is a gradual process and may take many years, most people conduct experiment and collect data either using bearings with artificially induced faults, or with accelerated life testing methods. While the data collection
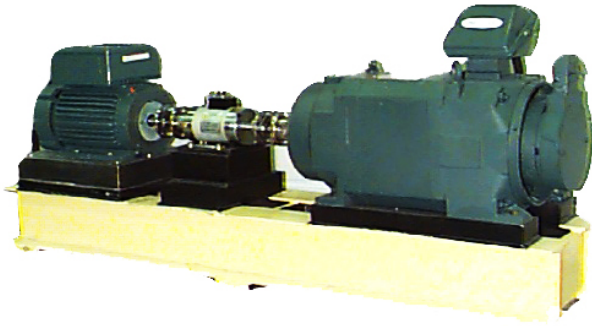
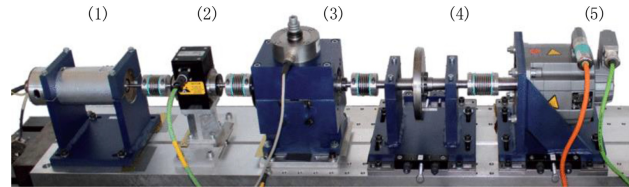Fig. 2. Experimental setup for collecting the CWRU bearing dataset [26].

Fig. 3. Modular test rig collecting the Paderborn bearing dataset consisting of (1) an electric motor, (2) a torque-measurement shaft, (3) a rolling bearing test module, (4) a flywheel, and (5) a load motor [27].
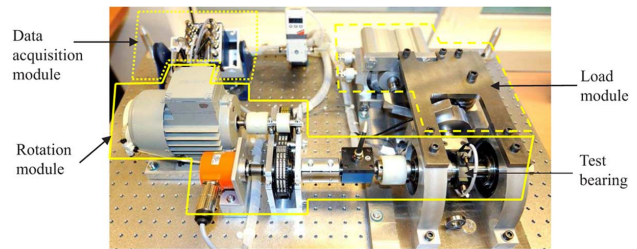

Fig. 4. PRONOSTIA testbed (Department of Automatic Control and Micro-Mechatronic Systems (AS2M), Franche-Comté Electronique Mécanique [29].

is still time consuming, fortunately a few organizations have made the effort and published their bearing fault datasets for engineers and researchers to develop their own ML algorithms. Thanks to their prevalence in the research community, these datasets can also serve as a common ground for the evaluation and comparison of different algorithms.

Before getting into details of various ML and DL developments, in this section, we briefly introduce a few popular datasets used by most papers covered in this review.

A. Case Western Reserve University (CWRU) Dataset

The test stand used to acquire the Case Western Reserve University (CWRU) bearing dataset is illustrated in Fig. 2, in which a 2-hp induction motor is shown on the left, a torque transducer/encoder is in the middle, while a dynamometer is coupled on the right. Single point faults are introduced to the bearings under test using electro-discharge machining with fault diameters of 7 mils, 14 mils, 21 mils, 28 mils, and 40 mils, at the inner raceway, the rolling element and the outer raceway. Vibration data are collected for motor loads from 0 to 3 hp and motor speeds from 1,720 to 1,797 rpm using two accelerometers installed at both the drive end and fan end of the motor housing, and two sampling frequencies of 12 kHz and 48 kHz were used. The generated dataset is recorded and made publicly available on the CWRU bearing data center website [26].

The CWRU dataset serves as a fundamental dataset to validate the performance of different ML and DL algorithms, and a comprehensive comparative study on previous work employing the CWRU dataset will be presented in Section V.

B. Paderborn University Dataset

The Paderborn university bearing dataset [27] includes the synchronous measurement of motor current and vibration signals, thus enabling the verification of multi-physics models and sensor fusion of different signals to increase the accuracy of bearing fault detection. Both stator current and vibration signals are measured with a high resolution and a high sampling rate, and experiments are performed on 26 damaged bearings and 6 undamaged (healthy) ones. Among the 26 damaged bearings, 12 are artificially damaged, and the other 14 have more realistic damages caused by accelerated life tests. This enables a more confident evaluation of ML algorithms in practical applications, where the real defects are generated through aging and the gradual loss of lubrication. The modular test rig used to acquire the Paderborn bearing dataset is illustrated in Fig. 3.

C. PRONOSTIA Dataset

Another popular dataset for predicting a bearing's remaining useful life (RUL) is known as the "PRONOSTIA bearings accelerated life test dataset", which serves for researchers to investigate new algorithms for bearing RUL prediction. During the International Conference on Prognostics and Health Management (PHM) in 2012, an "IEEE PHM 2012 Prognostic Challenge" was organized, where the PRONOSTIA degradation dataset [28] was provided to participants allowing them to train their prognostic methods. Every participant's method was evaluated based on the estimation accuracy of the RUL of bearings under test.

The main objective of PRONOSTIA is to provide real data related to the accelerated degradation of bearings performed at varying operating conditions [29]. The operating conditions are characterized by two sensors: a rotating speed sensor and a force sensor. In the PRONOSTIA platform as shown in Fig. 4, the bearing health is monitored by gathering two types of signals: temperature and vibration (with two uni-axis accelerometers installed in the horizontal and the vertical direction respectively). Furthermore, the data are recorded with a high sampling frequency which allows the interpretation of the entire frequency spectrum of interest during the bearing degradation process. Ultimately, the monitored data can be used for post-processing to extract the relevant features offline and continuously assess the bearing's RUL.

TABLE I
COMPARISON OF POPULAR BEARING FAULT DATASETS.

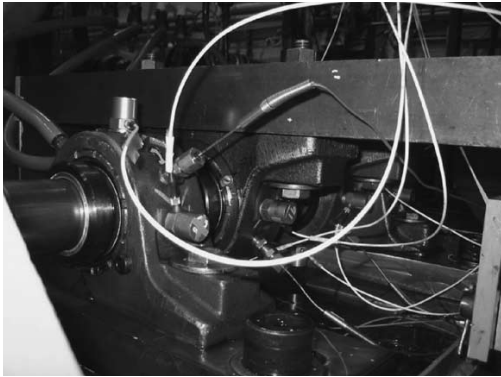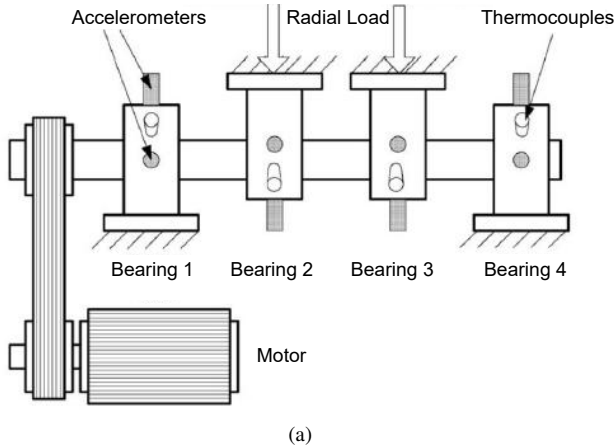| Dataset	Sensor type	Number of sensors	Sampling frequency	Fault mode
Case Western Reserve University (CWRU) Dataset	Accelerometer	2	12 & 48 kHz	Artificial
Paderborn University Dataset	Accelerometer & current sensor & thermocouple	1 & 2 & 1	64 kHz	Artificial & accelerated aging
PRONOSTIA Dataset	Accelerometer & thermocouple	2 & 1	25.6 kHz	Natural
Intelligent Maintenance Systems (IMS) Dataset	Accelerometer	2	20 kHz	Natural

Fig. 5. Illustration of the (a) bearing test rig and (b) vibration sensor placement of the IMS dataset [31].

D. Intelligent Maintenance Systems (IMS) Dataset

The IMS bearing dataset [30] is generated by the NSF I/UCR Center for Intelligent Maintenance Systems (IMS) with support from Rexnord Corp. Different from the other datasets, where the bearing faults are either artificially induced by scratching or drilling the bearing surface, or created by exerting a shaft current for accelerated life testing, the IMS dataset contains a complete record of natural bearing defect evolution. Specifically, the bearing is kept running for 30 days consecutively with a constant speed of 2,000 rpm, totaling around 86.4 million cycles before a defect is confirmed [31]. The test rig consists of four Rexnord ZA-2115 double row bearings installed on a shaft, which is coupled to an AC motor via a rubber belt as shown in Fig. 5(a). A radial load of 6,000 lbs is applied onto the shaft and bearing by a spring mechanism. Two accelerometers are installed on each bearing housing, and four thermocouples are attached to the outer race of each bearing to record bearing temperature for monitoring the lubrication purposes, as shown in Fig. 5(b).

The same experiment is repeated three times. Test 1 ends up with an inner race defect in bearing 3 and a rolling element defect in bearing 4. Test 2 and 3 end up with an outer race defect in bearing 1 and 3, respectively. The vibration data is collected every 5 or 10 minutes for a duration of 1 second with the sampling rate set at 20 kHz by National Instruments DAQCard 6062E. Since this dataset contains a complete collection of vibration signals of bearing experiments from start to failure with explicit time stamps, it is particularly suitable for predicting the RUL of rolling-element bearings.

E. Summary

A summary of the comparing the differences between different datasets is illustrated in TABLE I. So far a majority of literature on bearing fault identification with ML or DL algorithms employ the CWRU dataset due to its simplicity and popularity. The authors anticipate a growing interest on the Paderborn dataset, as it contains both the stator current signal and the vibration signal. In addition, this dataset also enables the validation of deep transfer learning and domain adaptation algorithms [32], [33] to predict a more realistic bearing fault from accelerated life testing with classifiers trained on artificially induced scratches or drills. Besides, many researchers working on RUL prediction also rely on the PRONOSTIA dataset and the IMS dataset. Since the main scope of this paper is bearing fault detection, research contributions on RUL prediction is not included in this literature survey.

III. CLASSICAL MACHINE LEARNING BASED APPROACHES

Before the recent DL boom, a variety of classical "shallow" machine learning and data mining algorithms have been around for many years, i.e., the artificial neural network (ANN). Applying these algorithms requires a lot of domain expertise and complex feature engineering. A deep exploratory data analysis is usually performed on the dataset first, followed by dimension reduction techniques such as the principal component analysis (PCA), etc., for feature extraction. Finally, the most representative features are passed along to the ML algorithm. The knowledge base of different domains and applications can be quite different and often requires extensive specialized expertise within each field, making it difficult to

perform appropriate feature extraction, or maintain a good level of transferability of ML models trained in one domain to be generalized or transferred to other contexts or settings.

Some of the earliest reviews investigating the use of artificial intelligence (AI) techniques on motor fault diagnostics can be found in [18], [19], where the characteristic fault frequencies for different motor fault types are systematically summarized, and relevant papers employing ANN and fuzzy systems are discussed. In this section, a brief summary of each classical ML method will be presented, with a comprehensive list of publications for readers' reference.

A. Artificial Neural Networks (ANN)

ANN is one of the oldest AI paradigms that has been applied to bearing fault diagnostics for almost 30 years [34]. In [34], the bearing wear of the motor is reflected in the damping coefficient B that can be inferred from a nonlinear mapping of the stator current I and the rotor speed ω. The complexity of obtaining an analytical expression for this nonlinear mapping is avoided by training a supervised neural network with stator current and motor speed measurements as input and predicted bearing condition as output. 35 training and 70 testing data patterns are collected on a laboratory test stand with the Dayton 6K624B-type bearing at different operating conditions. Highest bearing fault detection accuracy of 94.7% is achieved with the conventional neural network using two input nodes $\{I, \omega\}$. The accuracy can be further improved by utilizing five input dimensions $\{I, \omega, I^2, \omega^2, I^*\omega\}$ that are manually selected. However, besides the commonly used current sensor for bearing fault diagnostics, this method requires an addition speed encoder to collect the motor speed signal as an extra input, which is not commonly available in many low-cost induction motor drives. Similarly, the rest of the papers based on ANN [35]–[38] all require some degree of human expertise to guide its feature selection process in order to train the ANN model in a more effective manner.

B. Principle Component Analysis (PCA)

PCA is an algorithm that reveals the internal structure of the data in a way that best explains the variance in the data. If a multivariate dataset is visualized as a set of coordinates in a high-dimensional data space (one axis per variable), PCA can supply the user with a lower-dimensional projection of this object viewed from its most informative viewpoint. Since the sensitivity of various features that are characteristics of a bearing defect may vary considerably at different operating conditions, PCA has proven itself as an effective and systematic feature selection scheme that provides guidance on manually choosing the most representative features for classification purposes.

One of the earliest adoption of PCA on bearing fault diagnostics can be found in [39]. Experimental results revealed that the advantage in using only PCA identified features instead of the 13 original features is significant, as the fault diagnosis accuracy is increased from 88% to 98%. The study demonstrated that the proposed PCA technique is effective in classifying bearing faults with a higher accuracy and a lower number of input features when compared to using all of the original feature. Similarly, the rest of the papers based on PCA [40]–[43] take advantage of its data mining capability to facilitate the manual feature selection process and generate more representative features.

C. K-Nearest Neighbors (k-NN)

The k-NN algorithm is a non-parametric method used for either classification or regression. In k-NN classification, the output is the class of an object, which is identified by a majority vote of its k nearest neighbors. One early implementation of the k-NN classifier on bearing fault diagnostics can be found in [44], where k-NN serves as the core algorithm for a data mining based ceramic bearing fault classifier based on acoustic signals. Similarly, other k-NN based papers [45]–[47] employ k-NN to perform a distance analysis on each new data sample and determine whether it belongs to a specific fault class.

D. Support Vector Machines (SVM)

SVMs are supervised learning models that analyze data used for non-probabilistic classification or regression analysis. One classical work on the use of SVM towards identifying bearing faults can be found in [48], where classification results obtained by the SVM are optimal in all of the cases, with an overall improvement over the performance of ANN. Other similar SVM based papers [49]–[61] also illustrated the effectiveness and efficiency of employing SVM to serve as the fault classifier.

E. Others

Besides the commonly used ML methods listed above, many other algorithms have been applied to the identification of bearing faults, bringing in different characteristics and benefits, including neural fuzzy network [62]–[64], Bayesian networks [65]–[67], self-organizing maps [68], [69], extreme learning machines (ELM) [70], [71], transfer learning [72]–[74], linear discriminant analysis [75], [76], quadratic discriminant analysis [77], random forest [78], independent component analysis [79], softmax classifiers [80], manifold learning [81], [82], canonical variate analysis [83], particle filter [84], nonlinear preserving projection [85], artificial Hydrocarbon Networks [86], expectation maximization [87], ensemble learning [88], multi-scale permutation entropy [89], empirical mode decomposition [90]–[94], topic correlation analysis [95], affinity propagation [96], and dictionary learning [97], [98].

F. Challenges with the Classical ML Algorithms

As presented in the earlier sections, to detect the presence of a bearing fault using a classical ML algorithm, the characteristic fault frequencies are calculated based on the rotor mechanical speed and the specific bearing geometry, and these frequencies will serve as fault features. This feature determination process is known as "feature engineering". The amplitude of signals at these frequencies can be monitored to train various ML algorithms and identify any anomalies.

However, such a technique may encounter many challenges that ultimately affect the classification accuracy.

1) *Sliding:* The fault frequency is based on the assumption that no sliding occurs between the rolling element and the bearing raceway, i.e., these rolling elements will only roll on the raceway. Nevertheless, this is seldom the case in reality, as the rolling element often undergoes a combination of rolling and sliding movement. As a consequence, the calculated frequency may deviate from the real fault frequency and make this manually determined feature less informative of a bearing defect.

2) *Frequency interplay:* If multiple types of bearing faults occur simultaneously, these faults will interact and the resultant characteristic frequencies can add or subtract due to a complicated electro-mechanical process, thereby obfuscating the informative frequencies.

3) *External vibration:* There is also the possibility of interference induced from additional sources of vibration, i.e. bearing looseness and environment vibration, which can obscure the useful features.

4) *Observability:* Some faults, such as the bearing lubrication and general roughness related faults, do not even manifest themselves as a characteristic cyclic frequency, which makes them very hard to detect with the traditional model-based spectral analysis or classical data-driven ML methods.

5) *Sensitivity:* The sensitivity of various features that are characteristic of bearing defect may vary considerably at different operating conditions. A very thorough and systematic "learning stage" is typically required to test the sensitivity of these frequencies on any desirable operating condition before it can be actually put into use with the traditional approach.

Because of the aforementioned challenges, manually engineered features based on the bearing characteristic fault frequency can be difficult to interpret, and sometimes may even lead to inaccurate classification results, especially when applying the "shallow" classical ML methods that rely on human-engineered features in the training process. Therefore, many DL algorithms with automated feature extraction capabilities and better classification performance have been applied to bearing fault diagnostics, which will be discussed in detail in the next section.

IV. DEEP LEARNING BASED APPROACHES

Deep learning is a subset of machine learning that achieves great power and flexibility by learning to represent the world as nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed from less abstract ones. The trend of transitioning from classical "shallow" machine learning algorithms to deep learning can be attributed to the following reasons.

1) *Data explosion*: With the availability of exploding amount of data, and the application of crowdsourced labeling mechanisms such as Amazon mTurk [99], we are seeing a surging appearance of large scale dataset in many domains, such as ImageNet in image recognition, COCO for
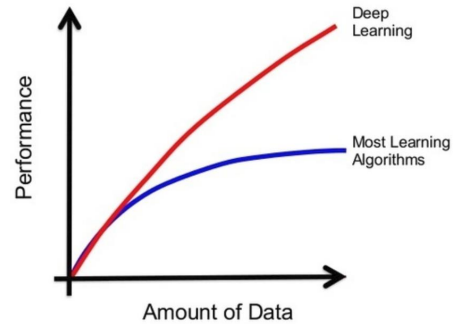

Fig. 6. Performance comparison of deep learning and most classical learning algorithms [100].

object segmentation and recognition, VoxCeleb in speaker identification, *et al*. DL generally requires a large amount of labeled data. Some DL models in computer vision were trained using more than one million images. For many applications, including the diagnostics of bearing faults, such large datasets are not readily available and will be expensive and time consuming to acquire. On smaller datasets, classical ML algorithms can compete with or even outperform deep learning networks. With the increase of the amount of data, the performance of DL can significantly outperform most classical ML algorithms, as illustrated in Fig. 6 [100] by Andrew Ng.

2) *Algorithm evolution*: More techniques are being invented and getting matured in terms of controlling the training process of deeper models to achieve faster speed, better convergence, and improved generalization. For example, algorithms such as ReLU help accelerate convergence speed; techniques such as dropout and pooling help prevent overfitting; numerical optimization methods such as mini-batch gradient descent, RMSprop, and L-BFGS optimizer help leverage more data and train deeper models.

3) *Hardware evolution*: Training deep networks is extremely computationally intensive, but running on a high performance GPU can significantly accelerate this training process. Specifically, GPU offers parallel computing capability and computational compatibility with deep neural networks, which makes them indispensable for training DL based algorithms. More powerful GPUs allows data scientists to quickly get the DL training up and running. For example, the NVIDIA Tesla V100 Tensor Core GPUs can now parse petabytes of data orders of magnitude faster than traditional CPUs [101], and leverage mixed precision to accelerate DL training throughputs across every type of neural network. In the most recent years, the emergence of the accelerators for parallel computing such as GPUs, FPGAs, ASICs and TPUs have promoted the fast evolution of DL algorithms.

All of the factors above contribute to the new era of applying DL algorithms to a variety of data-related applications. Specifically, advantages of applying DL algorithms compared to classical ML algorithms include:

1) *Best-in-class performance*: The complexity of the computed function grows exponentially with model depth [102]. DL has the best-in-class performance that significantly outperforms other solutions to problems in multiple domains, including speech, language, vision, game playing, etc.

2) *Automatic feature extraction*: DL removes the need for feature engineering. Classical ML algorithms usually demand sophisticated manual feature engineering, which unavoidably requires expert domain knowledge and numerous human effort. However, when using deep neural network, there's no need for this manual process. One can simply pass the data directly to the network, and the network can automatically learn the features from raw data by auto-tuning the weights in the network. The DL network eliminates completely the challenging stage of feature engineering.

3) *Transferability*: The strong expressive power and high performance of a deep neural network trained in one domain can be easily generalized or transferred to other contexts, settings or domains. Deep learning is an architecture that can be adapted to new problems relatively easily. For instance, problems in different domains such as vision, time series, and language are being solved using the same techniques like convolutional neural networks, recurrent neural networks, and long short-term memory, etc.

Thanks to the aforementioned reasons for the transition from traditional methods to DL methods, as well as the benefits of DL algorithms discussed above, we have witnessed an exponential increase in DL applications, such as machine health monitoring and fault diagnostics, among which the bearing fault detection is a very representative case.

A. Convolutional Neural Network (CNN)

Inspired by animal visual cortices [103], the convolution operation is first introduced to detect image patterns in a hierarchical way from simple features such as edge and corner to complex features. Specifically, lower layers in the network detect fundamental lower level visual features; and layers afterward detect higher level features, which are built upon these simple lower level features.

The first paper employing CNN to identify bearing fault was published in 2016 [104], and in the next three years many papers applying the same technique [105]–[119] have emerged and contributed to advancing bearing fault detection in various aspects. The basic architecture of a CNN-based bearing fault classifier is illustrated in Fig. 7. Specifically, the 1-D temporal raw data obtained from different accelerometers are firstly stacked to 2-D vector form similar to the representation of images, which is then passed over to a convolutional layer for feature extraction, followed by a pooling layer for down-sampling. The combination of this convolution-pooling pattern is repeated many times to further deepen the network. Finally, the output from the hidden layers will be handed over to one or several fully-connected layers, the result of which is transferred to a top classifier based on Softmax or Sigmoid functions to determine if a bearing fault is present.

In [104], the vibration data are collected using two uni-axis accelerometers installed on x- and y- direction respectively. A CNN is able to autonomously learn useful features for bearing fault detection from the raw data pre-processed by scaled discrete Fourier transform. The classification result demonstrates that the feature learning based approach significantly outperforms the feature engineering based approach of conventional ML. Moreover, another contribution of this work is to show that feature learning based approaches such as CNN can also perform bearing health prognostics, and identify some early-stage faulty conditions that have no explicit characteristic frequencies, such as lubrication degradation, which cannot be achieved using classical ML methods.

To obtain a better trade-off between the training speed and accuracy, an adaptive CNN (ADCNN) is applied on the CWRU dataset to dynamically change the learning rate in [105]. The entire fault diagnosis model employs a fault pattern determination component using 1 ADCNN and a fault size evaluation component using 3 ADCNNs, and 3-layer CNNs with max pooling. Classification results demonstrate that AD-CNN has a better accuracy compared to conventional shallow CNN and SVM methods, especially in terms of identifying the rolling element defect. In addition, this proposed ADCNN is also able to predict the fault size (defect width) with a satisfactory accuracy. On top of the conventional structure of CNN, a dislocate layer is added in [106] that can better extract the relationship between signals with different intervals in periodic forms, especially during the change of operating conditions. It is reported in [106] that the best accuracy of 96.32% is achieved with a disclose step factor $k = 3$, while the accuracy of conventional CNN without this disclose layer is only 83.39%.

Similar to earlier work [104]–[106], [107] implements a 4-layer CNN structure with 2 convolutional and 2 pooling layers employing both the CWRU dataset & dataset generated by Qian Peng Company in China, and the accuracy outperforms the conventional SVM and the shallow Softmax regression classifier, especially when the vibration signal is mixed with ambient noise. The improvement can be as large as 25%, showcasing the excellent built-in denoising capabilities of the CNN algorithm. A sensor fusion approach is applied in [108], in which both the temporal and spatial information of the CWRU raw data from two accelerometers at the drive end and the fan end are stacked by transforming 1-D time-series data into a 2-D matrix form. The average accuracy using the fusion of two sensors is increased to 99.41% from the previous 98.35% with only one sensor.

Many variations of CNN are also employed to tackle the bearing fault diagnosis challenge [109]–[116] using the CWRU dataset to obtain more desirable characteristics than conventional CNN. For example, a CNN based on LeNet-5 is applied in [109], which contains 2 alternating convolutional-pooling layers and a 2 fully-connected layers. Padding is used to control the size of learned features, and zero-padding is applied to prevent dimension loss. This improved CNN architecture is able to provide a better feature extraction capability with an astonishing accuracy (99.79%) on test set, which is higher than other deep learning based methods such
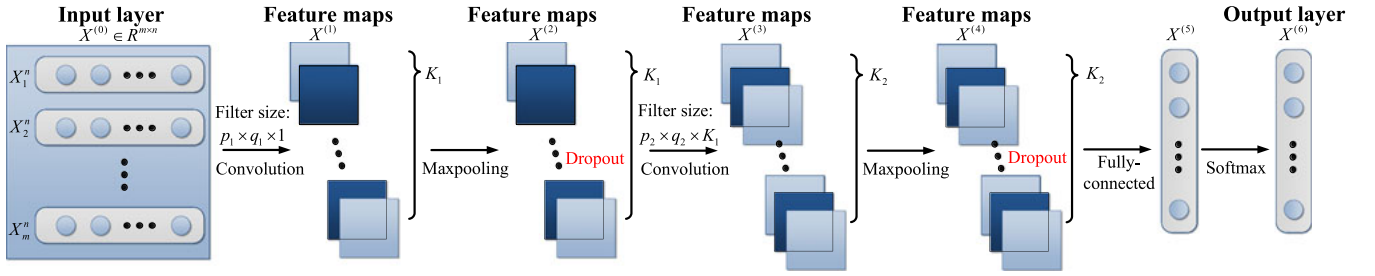
Fig. 7. Architecture of the CNN-based fault diagnosis model [109].

as the adaptive CNN (98.1%) and the deep belief network (87.45%). The proposed CNN based on LeNet-5 also dominates the classical ML methods such as SVM (87.45%) and ANN (67.70%). In addition, a deep fully convolutional neural network (DFCNN) incorporating 4 convolution-pooling layer pairs is employed in [110], while the raw data are also transformed into spectrograms for easier processing. An accuracy of 99.22% is accomplished, outperforming 94.28% of the linear SVM with particle swarm optimization (PSO), and 91.43% of the conventional SVM. These results are obtained using the same training set to train different networks and the same test set to evaluate and compare their performances.

To save the extensive training time required for most CNN based algorithms, a multi-scale CNN (MS-DCNN) is adopted in [111], where convolution kernels of different sizes are used to extract features of different scales in parallel. The mean accuracy of a 9-layer 1-D CNN, a 2-D CNN and the proposed MS-DCNN are 98.57%, 98.25% and 99.27%, respectively. In addition to the subtle increase in accuracy compared to conventional CNNs, the number of parameters to be determined during training is only 52,172, which is significantly lower than those of 1-D CNN (171,606) and 2-D CNN (213,206). Moreover, a very deep CNN of 14 layers with training interference is used in [112], which is able to maintain a high accuracy in noisy environments or during load shifts. However, the training time and the amount of parameters to be trained would increase dramatically, posing a potential threat of overfitting the data. Similarly, to overcome the impact of load variations, a novel bearing fault diagnosis algorithm based on improved Dempster-Shafer theory CNN (IDS-CNN) is employed in [113]. This improved D-S evidence theory is implemented via a distance matrix from the modified Gini Index. Extensive evaluations revealed that, by fusing complementary or conflicting evidences from different models and sensors, the proposed IDS-CNN algorithm is able to accommodate different load conditions and achieve a better fault diagnosis performance than conventional DNN models and ML approaches such as SVM.

To better suppress the impact of speed variations on bearing fault diagnosis, a novel architecture based on CNN referred to as "LiftingNet" is implemented in [114], which consists of split layers, predict layers, update layers, pooling layers, and fully-connected layers, with the main learning process performed in a split-predict-update loop. A 4-class classification is carried out with the CWRU dataset randomly and evenly split into training set and test set. The final classification accuracy is 99.63%. However, since all of the signals recorded by CWRU are measured in a small speed range (from 1,720 to 1,797 rpm), another experiment is established to record vibration signals with four distinct rotor frequencies (approximately 10, 20, 30, and 40 Hz), and the average accuracy still reaches 93.19%, which is 14.38% higher than conventional SVM algorithm. Similarly, a fault diagnosis method based on the Pythagorean spatial pyramid pooling (PSPP) CNN is proposed in [115] to enhance the classification accuracy during motor speed variations. Compared to a spatial pyramid pooling layer that has been used in an CNN, a PSPP layer is allocated in this work as a front layer of CNN, and features obtained by the PSPP layer can be delivered to convolutional layers for further feature extraction. According to the experiment result, this method has a higher diagnosis accuracy at various rotating speeds compared to other methods. In addition, the PSPP-CNN model trained by data at certain rotating speeds can be transferred and used to diagnose bearing fault at full working speed.

Since CNN excels at processing 2-D matrix data, such as images in the field of computer vision, it generally requires the transformation of 1-D time-domain vibration signal into 2-D signal to take full advantage of the strength that CNN can offer. Aiming at simplifying this conversion process and reducing the percentage of training data required due to the expensiveness of acquiring a large amount of data through experiments, an adaptive overlapping CNN (AOCNN) is proposed in [116] to directly process the 1-D raw vibration signal, and eliminate the shift variant problem of time-series signal. Compared to the conventional CNN, its novelty lies in the overlapping layer, which is used to sample the raw vibration signal. After the adaptive convolutional layer separates these samples into segments, sparse filtering is employed in the local layer to obtain local features. Classification results reveal that AOCNN with SF can identify ten health conditions of the bearing with a 99.19% test accuracy when only 5% samples are used as the training set, which is a significant improvement considering most of the DL based methods demand a minimum of 25% data allocated in the training set. In addition, the test accuracy can further rise to 99.61% when the test set data percentage increases from 5% to 20%.

Besides the CWRU dataset which contains only vibra-

tion signal, the Paderborn University bearing dataset [27], as stated in Section II, includes synchronized stator current and vibration signals. In addition, the Paderborn dataset also incorporates both artificially induced bearing fault and realistic damages caused by accelerated lifetime tests. In [117], the Paderborn dataset is used to train a deep inception net with atrous convolution, which improves the average accuracy from 75% (best result of conventional data-driven methods) to 95% for diagnosing the real bearing faults when trained only with the data generated from artificial bearing damages. The "PRONOSTIA bearings accelerated lifetime test dataset" [28], as introduced in Section II, is applied in [118] with a deep convolution structure consisting of 8 layers: 2 convolutional, 2 pooling, 1 flat, and 3 nonlinear transformation layers. Health indicators (HI) are later defined based on the CNN output, and the classification result shows the accuracy of HI predicted using CNN is superior than that of self-organizing maps (SOM).

In addition to identifying damages on rolling element bearings, the adoption of CNN on spindle bearings is also discussed in [119], in which the wavelet packet energy of the vibration signal is taken as input.

B. Auto-encoders

Auto-encoder is proposed in the 1980s as an unsupervised pre-training method for ANNs [120], [121]. After decades of evolution, the auto-encoder has become widely adopted as an unsupervised feature learning method and a greedy layer-wise neural network pre-training method. The training process of an auto-encoder with 1 hidden layer is illustrated in Fig. 8 [122]. Specifically, an auto-encoder is trained from an ANN, which consists of two parts: the encoder and the decoder. The output of the encoder is fed into the decoder as input. The ANN takes the mean squared error between the original input and output as the loss function, which essentially aims at generating the final output by imitating the input. After this ANN is trained, the decoder part is dropped while only the encoder part is kept. Therefore, the output of the encoder is the feature representation that can be employed in the next-stage classifier.

Among a large number of studies of applying auto-encoders to bearing fault diagnosis [122]–[134], an early attempt can be found in [123], where a 5-layer auto-encoder based DNN is utilized to adaptively extract fault features from the frequency spectrum and effectively classify the bearing health condition. The classification accuracy reaches 99.6%, which is significantly higher than the 70% of back-propagation based neural networks (BPNN). In [124], an auto-encoder based extreme learning machine (ELM) is employed, seeking to integrate the automatic feature extraction capability of auto-encoders and the high training speed of ELMs. The average accuracy of 99.83% compares favorably against other traditional ML methods, including wavelet package decomposition based SVM (WPD-SVM) (94.17%), EMD-SVM (82.83%), WPD-ELM (86.75%) and EMD-ELM (81.55%). More importantly, the required training time drops by around 60% to 70% using the same training and test data, thanks to the adoption of ELM.
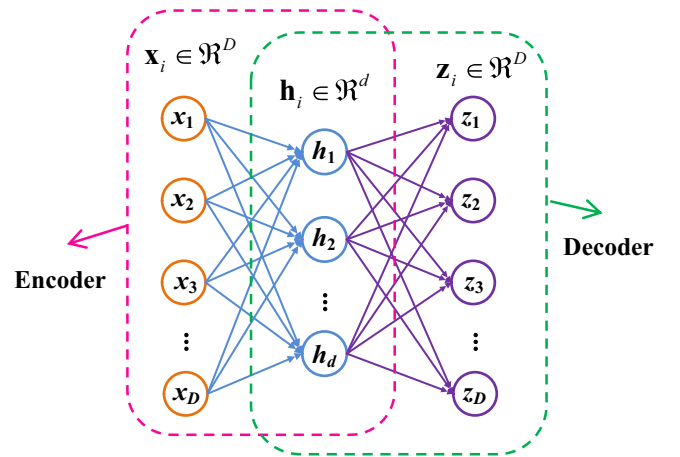

Fig. 8. Process of training a one hidden layer auto-encoder [122].

Compared to CNN, the denoising capability of conventional auto-encoders is not prominent. Thus in [125], a stacked denoising auto-encoder (SDA) is implemented, which is suitable for deep architecture based robust feature extraction on signals containing ambient noise under varying working conditions. This specific SDA consists of three auto-encoders stacked together. To strike a balance between classification performance and training speed, three hidden layers with 100, 50, and 25 units respectively are employed. The original CWRU bearing data are perturbed by a 15 dB random noise to manually create a noisy background, and data from multiple operating conditions are used as the test set to evaluate its denoising capability at different speeds and loads. The average classification result reveals the proposed SDA is able to achieve a worst case accuracy of 91.79%, which is 3% to 10% higher when compared to the conventional SAE without the denoising capability, and classical ML algorithms such as SVM and random forest (RF). Similar to [125], another form of SDA is utilized in [126] with three hidden layers of (500, 500, 500) units. Signals from the CWRU dataset are mixed with different levels of artificially induced noise in the time domain, and later transformed to the frequency domain. The proposed method has a better diagnosis accuracy than deep belief networks, particularly with the added noises, where an average improvement of 7% is achieved.

In [122], a locomotive bearing dataset developed at the Northwestern Polytechnical University is used to validate the performance of auto-encoders. Based on this dataset, the authors adopted the maximum correntropy as the loss function instead of the traditional mean squared error, and an artificial fish-swarm algorithm (AFSA) is used to optimize the key parameters in the loss function of the auto-encoder. Results show that the customized 5-layer auto-encoder composed of this maximum correntropy loss function and AFSA algorithm outperforms standard auto-encoder by an accuracy of 10% to 40% in a 5-class classification problem. Similarly, a new deep AE constructed with DAE and contractive auto-encoder (CAE) is applied to the locomotive bearing dataset for en-

hancing the feature learning capability. A single DAE is firstly used to learn low-layer features from the raw vibration data, then multiple CAEs are used to learn deeper features. In addition, locality preserving projection (LPP) is also adopted to fuse these deep features to further improve the quality of the learned features. The classification accuracy of this mixed DAE-CAE-LPP approach is 91.90%, showcasing the advantage over the standard DAE (84.60%), the standard CAE (85.10%), and the classical ML algorithms of BPNN (49.70%) and SVM (57.60%). However, all of the auto-encoder based methods are also 6 to 10 times more time-consuming when compared to classical ML methods.

In addition. an aircraft-engine inter-shaft bearing vibration dataset with the inner race, the outer race, and the rolling element defect is adopted as the input data in [127], where a new AE based on Gaussian radial basis kernel function is employed to enhance the feature learning capability. Later, a stacked AE is developed using this new AE and multiple conventional AEs. An average accuracy of 86.75% is achieved, which is much better compared to the standard SAE (44.90%) and the standard DBN (19.65%). Moreover, the importance of the proposed Gaussian radial basis kernel function is showcased in a comparative study. When the Gaussian kernel function is changed to a polynomial kernel function (PK) and a power exponent kernel function (PEK), the accuracy would drop to 24.25% and 65.55%, respectively.

Similar to the case of CNN, many variations of SAE are also employed in the last two years to tackle the bearing fault diagnosis problem [128]–[134] using the popular CWRU dataset, and all of which have achieved some form of performance elevation when compared to the traditional SAE. In [128], an ensemble deep auto-encoder consisting of a series of auto-encoders (AE) based on different activation functions is proposed for unsupervised feature learning from the measured vibration signal. Later, a decision ensemble strategy is designed to merge the classification result from each individual AE and ensure an accurate and stable diagnosis result. An average classification accuracy of 99.15% is achieved, which performs better than many classical ML methods including BPNN (88.22%), SVM (90.81%), and RF (92.07%) based on a manually selected feature of 24 dimensions. Similarly, by altering the activation function, a deep wavelet auto-encoder (DWAE) with extreme learning machine (ELM) is implemented in [129], where the wavelet function is employed as the nonlinear activation function, enabling wavelet auto-encoders (WAE) to effectively capture signal characteristics. Then a DWAE with multiple WAEs is constructed to enhance the unsupervised feature learning ability, and ELM is adopted as the output classifier. Based on the final result, this method (95.20%) not only outperforms the classical ML methods such as BPNN (85.43%) and SVM (87.97%), but also some standard DL algorithms, including the standard DAE with Softmax (89.70%) and the standard DAE with ELM (89.93%).

Considering the relatively large data size required to train deep neural nets, a 4-layer DNN with stacked sparse auto-encoder is established in [130] with a compression ratio of 70%, indicating only 30% of the original data are needed to train the proposed model. The DNN has 720 input nodes, 200 and 60 nodes in the first and the second hidden layer, and 7 nodes in the output layer, the number of which depends on the number of fault conditions. A nonlinear projection is performed to compress the vibration data and perform adaptive feature extraction in the transformed space, and the accuracy of the proposed method reaches 97.47%, which is 8% higher than the SVM, 60% higher than a three-layer ANN, and 46% higher than a multi-layer ANN. [131] summarizes two limitations of the conventional SAE. Firstly, an SAE tends to extract similar or redundant features that increase the complexity rather than the accuracy of the model. Secondly, the learned features may have shift variant properties. To overcome these issues, a new SAE-LCN (local connection network) is proposed, which consists of the input layer, the local layer, the feature layer, and the output layer. Specifically, this method learns features from the input signal locally in the local layer, then obtains shift-invariant features in the feature layer, and finally recognizes the bearing health condition in the output layer for a 10-class classification problem. The average accuracy is reported to reach 99.92%, which is 1% to 5% higher than EMD, ensemble NN, and DL based methods. Similarly, a diagnosis model using SAE and incremental support vector machines is implemented in [132], which is tested for online diagnosis purposes.

Besides the most commonly used Softmax classifiers in the output layer, the Gath-Geva (GG) clustering algorithm is implemented in [133], which induces a fuzzy maximum likelihood estimation (FMLE) of the distance norm to determine the likelihood of a sample belonging to each cluster. While an 8-layer SDAE is still used to extract the useful features from the vibration signal, GG is deployed to identify the different fault types. The worst case classification accuracy is 93.3%, outperforming the classical EMD based feature extraction schemes by almost 10%.

To further reduce the DL based model complexity, another bearing fault diagnosis method based on a fully-connected winner-take-all auto-encoder is proposed in [134], in which the model explicitly imposes lifetime sparsity on the encoded features by keeping only k% largest activations of each neuron for all of the samples in a mini-batch. A soft voting method is implemented to increase the classification accuracy and stability by aggregating the prediction result of each signal segment sliced by a sliding window. A customized dataset is generated to test the diagnosis performance under a noisy environment by adding white Gaussian noise to the original CWRU dataset. The experimental result demonstrates that with a simple *two-layer* network, the proposed method not only handles the bearing fault detection with a higher precision at normal conditions, but also demonstrates a better anti-noise capability when compared to some deeper and more complex models, such as a deep CNN.

C. Deep Belief Network (DBN)

In DL, a deep belief network (DBN) can be viewed as a composition of simple unsupervised networks such as restricted Boltzmann machines or auto-encoders, where each sub-network's hidden layer serves as the visible layer for the
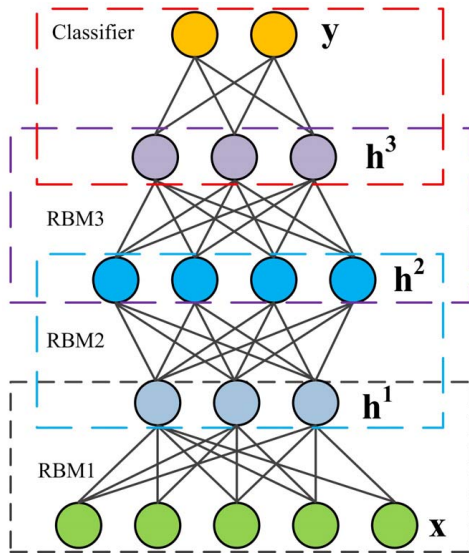
Fig. 9. Architecture of DBN [136].

next, as illustrated by different colored boxes in Fig. 9. An RBM is an undirected generative energy-based model with a "visible" input layer, a hidden layer, and connections in between, but not within layers. This composition leads to a fast layer-by-layer unsupervised training procedure, where contrastive divergence is applied to each sub-network in turn, starting from the "lowest" pair of layers in the architecture.

This greedy layer-by-layer training process has led to one of the first effective DL algorithms [135]. There are many attractive implementations of DBNs in real-life applications such as natural language understanding and drug discovery; and its first application on bearing fault diagnosis was published in 2017 [136].

In [136], a multi-sensor vibration data fusion technique is implemented to fuse the time-domain and frequency-domain features extracted using multiple 2-layer SAEs. Then a 3-layer DBN is used for classification purposes. Validation is performed on the vibration data collected at different speeds, and the 97.82% accuracy demonstrates that the proposed method can effectively identify a bearing fault at multiple operating conditions. The feature visualization using t-SNE reveals that this multi-SAE based feature fusion outperforms other methods with only one SAE or without fusion. In [137], a stochastic convolutional DBN is implemented by means of stochastic kernels and averaging, and an unsupervised CNN is built to extract 47 features. Later a 2-layer DBN is implemented with (28, 14) nodes, 5 kernels in each layer, and 1 pooling layer without overlapping. Finally, a Softmax layer is used for classification, and the average accuracy exceeds 95%.

Many DBN papers also take the CWRU bearing dataset as the input data [138]–[140] due to its popularity. For example, an adaptive DBN and dual-tree complex wavelet packet (DTCWPT) is proposed in [138]. The DTCWPT first prepossesses the vibration signal to generate a feature set with 9×8 feature parameters. Then a 3-level wavelet decomposition

of the signal is performed using the order 5 Daubechies wavelet as the basis function. Then a 5-layer adaptive DBN of the (72, 400, 250, 100, 16) structure is used for bearing fault classification. The average accuracy is 94.38%, which is much higher compared to the classical ML methods such as ANN (63.13%), GRNN (69.38%), and SVM (66.88%) using the same training and test data. In [140], data from two accelerometers mounted on the load end and fan end respectively are processed by multiple DBNs for feature extraction; then faulty conditions based on the extracted features are determined by Softmax; and the final health condition is fused by D-S evidence theory. An accuracy of 98.8% is accomplished considering the load variation from 1 to 3 hp, a significant improvement when compared to the conventional SAE and CNN. Similar to this D-S theory based output fusion method [139], a 4-layer DBN of the (400, 200, 100, 10) structure with different hyper-parameters coupled with ensemble learning is implemented in [140]. Specifically, an improved ensemble method is used to acquire the weight matrix for each DBN, and the final diagnosis result is formulated from each DBN based on their weights. The average accuracy of 96.95% is better than that of a single DBN of different weights (mostly around 80%), as well as a simple voting ensemble scheme based DBN (91.21%).

Besides the CWRU bearing dataset, many other datasets have been used to evaluate the performance of DBN on bearing fault diagnostics. In [141], a convolutional DBN constructed with convolutional RBMs is applied on the locomotive bearing vibration, where an auto-encoder is firstly used to compress the data and reduce its dimension. Without any feature extraction process, the compressed data are divided into training samples and test samples to be fed into the convolutional DBN. The convolutional DBN based on Gaussian visible units is able to learn the representative features, overcoming the problem of conventional RBMs that all visible units must be related to all hidden units by different weights. Lastly, a Softmax layer is used for classification and obtains an accuracy of 97.44%, which compares favorably against other DL methods, such as the denoising auto-encoder (90.76%), the standard DBN (88.10%), and the standard CNN (91.24%), using the same classifier and raw data. In [142], a bearing dataset directly obtained from power plants is used to evaluate the performance of a 5-hidden-layer DBN with (512, 2048, 1024, 2048, 512) nodes in each layer. The dataset contains vibration signals collected from various scaled applications, such as small testbeds and real field deployments. The unsupervised feature extraction is performed by DBN, and the fault classifier is designed using SOM which achieves a 97.13% accuracy.

DBN has also been applied to bearing RUL prediction. In [143], a DBN-feed-forward neural network (FNN) is applied to perform automatic feature learning with DBN and RUL prediction with FNN. Two accelerometers are mounted on the bearing housing, in directions perpendicular to the shaft, and the data is collected with a 102.4 kHz sampling frequency for a duration of 2 seconds. Experimental results demonstrate the proposed DBN based approach can accurately predict the true RUL as the bearing approaches the point of failure, and the accuracy of the predictions tends to increase and converge
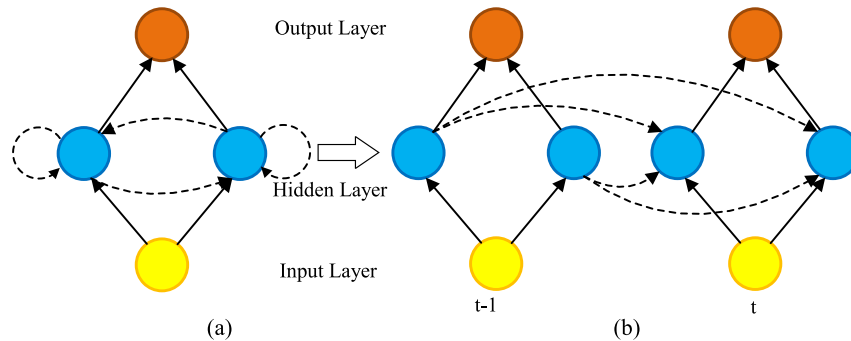
Fig. 10. Architecture of (a) RNN, and (b) RNN over a time step [144].

over time.

D. Recurrent Neural Network (RNN)

Different from a feed-forward neural network, a recurrent neural network (RNN) processes the input data in a recurrent behavior, and its architecture is shown in Fig. 10. With a flow path going from the hidden layer to itself, when unrolled in sequence, it can be viewed as a feed-forward neural network in the input sequence. As a sequential model, it can capture and model sequential relationships in sequential data or time-series data. However, often trained with back-propagation through time, RNN has the notorious gradient vanishing/exploding issue stemmed from its nature. Although the RNN is proposed as early as the 1980s, it has limited applications due to this reason, until the birth of long short-term memory (LSTM) in 1997. Specifically, LSTM is augmented by adding recurrent gates called "forget" gates. Designed for overcoming the gradient vanishing/exploding issue, LSTM has shown an astonishing capability in memorizing and modeling the long-term dependency in data, and therefore taken a dominant role in time-series and textual data analysis. So far, it has received great successes in the field of speech recognition, handwriting recognition, natural language processing, video analysis, etc.

One of the earliest applications of RNN on bearing fault diagnostics is reported in 2015 [144], where fault features are firstly extracted using the discrete wavelet transform and later selected based on the orthogonal fuzzy neighbourhood discriminative analysis. These features are then fed into an RNN to perform bearing fault detection. The experimental result has shown that the proposed scheme based on RNN is capable of accurately detecting and classifying the bearing fault. Another RNN based health indicator (RNN-HI) is proposed in [145] to predict the RUL of bearings with LSTM cells used in RNN layers. Along with time-frequency features, the related-similarity (RS) feature calculates the similarity between the currently monitored data and the data at an initial operation point. After performing a correlation and monotonicity-metrics-based feature selection process, the selected features are transferred to an RNN network to predict the bearing HI, from which the RUL can be estimated. With the input dataset collected from generator bearings of wind turbines, the proposed RNN-HI is demonstrated to offer better performance than an SOM based method.

In addition, a methodology of a combined 1-D CNN and LSTM to classify bearing fault types is presented in [146], where the entire architecture is composed of a 1-D CNN layer, a max pooling layer, a LSTM layer, and a Softmax layer as the top classifier. The system input is the raw signal without any pre-processing, and the best test accuracy of different configurations reaches 99.6%. A more recent work employing a deep recurrent neural network (DRNN) is proposed in [147] with stacked recurrent hidden layers and LSTM units. A loss function with mean squared errors is introduced and the stochastic gradient descent (SGD) method is used as the optimizer. Besides, an adaptive learning rate is also adopted to improve the training performance. The average accuracy on the test set using the proposed method is 94.75% and 96.53% at 1,750 and 1,797 rpm respectively.

E. Generative Adversarial Network (GAN)

Generative Adversarial Network (GAN) was proposed by Goodfellow *et al.* [148] in 2014 and rapidly became one of the most exciting breakthroughs in the field of deep learning. A GAN is composed of two parts: the generator F_G and the discriminator F_D, as illustrated in Fig. 11 [149]. The two parts are competing with each other in a way that the generator F_G is trying to confuse the discriminator F_D, and F_D is trying to distinguish samples generated by F_G from samples in the original dataset. Established as a zero-sum game framework, both F_G and F_D are competing to obtain an increasingly stronger capability of imitating the original data samples and discriminating in an iterative manner.

A GAN is mainly designed for generative purposes to generate samples or functions as generation modules. Despite its relatively short history, GAN has been rapidly applied to the field of bearing fault diagnostics. One of the earliest publications appears in 2017 [150], which aims at addressing the class imbalance issue using GAN. In addition, GAN is also combined with the adaptive synthetic sampling (ADASYN) approach to achieve meaningful oversampling when the original data samples are sparse. Comparison against standard oversampling techniques shows the superiority of adopting GAN. In [151], A novel approach for fault diagnosis based on deep convolution GAN (DCGAN) with imbalanced dataset is proposed. A new DCGAN model [152] with 4 convolutional layers serving as the discriminator and the generator is designed and applied on
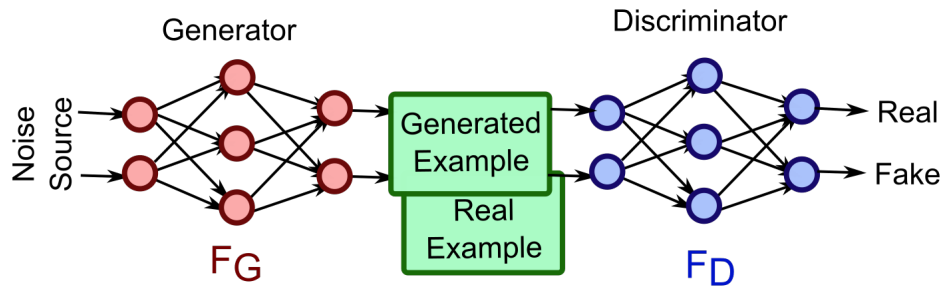
Fig. 11. Architecture of GAN [149].

raw and imbalanced vibration signals. After performing data balancing using the DCGAN model, statistical features based on time-domain and frequency-domain data are extracted to train a SVM classifier for bearing fault classification. Both the training and the test accuracy of the proposed DCGAN method demonstrate better performance than other class balancing methods, including random over-sampling, random under-sampling, and synthetic minority over-sampling technique.

We can find a number of research works in the field of bearing fault diagnostics employing GAN and its variants for data augmentation purposes due to their excellent generative capability. Besides that, there are also some works using GAN as the main framework to realize classification tasks, which heavily rely on the assumption that the data structure in latent space, although without labels, contains information that can be used to infer the labels. When a GAN is learning from un-labeled samples in a unsupervised manner, it can additionally learn the data distribution in latent space that distinguishes the data's unknown classes. In this way, the discriminator of GAN can be refined as a classifier assisted by some other modules in the framework. This class of GAN-centered frameworks has shown superiority in semi-supervised areas, especially in applications where labeled data are expensive and scarce.

In [153], for example, the authors proposed a novel GAN framework referred to as the categorical adversarial auto-encoder (CatAAE), which automatically trains an auto-encoder through an adversarial training process, and imposes a prior distribution on the latent coding space. In the next step, a classifier tries to cluster the input examples by balancing the mutual information between examples and their predicted categorical class distributions. The latent coding space and the training process are presented to investigate the advantage of the proposed model. Experiments at different signal-to-noise ratios (SNRs) and different motor load levels have indicated the preponderance of the proposed CatAAE in learning useful characteristics when compared to the categorical generative adversarial networks (CatGAN) and the K-means algorithm.

Since many real-world applications do not comply with the common assumption that the training set and the test set have the same distribution, due to the fact that the operating condition may vary frequently. Similar to [153] and inspired by GAN, a new adversarial adaptive 1-D CNN model (A2CNN) is proposed in [154] to address this problem. Experiments show that the A2CNN has a strong fault-discriminative and domain

invariant capacity, and therefore its prediction can achieve a high accuracy even at different operating conditions. Other works employing GAN to tackle the data imbalance issue can be found in [155] and [156].

F. Deep Learning based Transfer Learning

The success of ML and DL based bearing fault diagnostics relies on a massive amount of heavily annotated data. However, this is generally not feasible in most real-world applications due to 1) the dangerous and serious consequences when machines are running at faulty conditions; 2) the potential time-consuming degradation process before the desired failure appears; and 3) the possibility of a large number of operating conditions with different speeds and loads. With DL methods trained with either the publicly available datasets or self-collected datasets sampled in a laboratory environment, the classification accuracy will naturally deteriorate when determining the presence of bearing fault in a real-world application. Even if the data collected from the same machines and bearings are used, certain level of distribution discrepancy inevitably exists between features of the training and test sets if they were at different loads or speeds. As a result, the performance still suffers.

Designed to tackle this practical and widely existing issue in numerous applications, transfer learning has aroused extensive attention in the machine learning community, and various transfer learning frameworks are proposed based on classical ML algorithms [32], [33], [157], [158]. A popular method among all types of transfer learning approaches is domain adaptation. By exploring domain-invariant features, domain adaptation establishes the knowledge transfer from the source domain to the target domain [159]. Therefore, with labeled data from the source domain and unlabeled data from the target domain, the distribution discrepancy between the two domains can be mitigated by domain adaptation algorithms. Over the last few years, an integration of deep learning and transfer learning approaches has been prevalent. Specially designed domain adaptation modules are combined with deep learning architectures to endow the domain transfer ability while maintaining the extraordinary automatic feature learning ability [159]–[163].

Specifically, a domain adaptation module is proposed in [161] to facilitate a 1-D CNN to learn domain-invariant features by maximizing the domain recognition error and
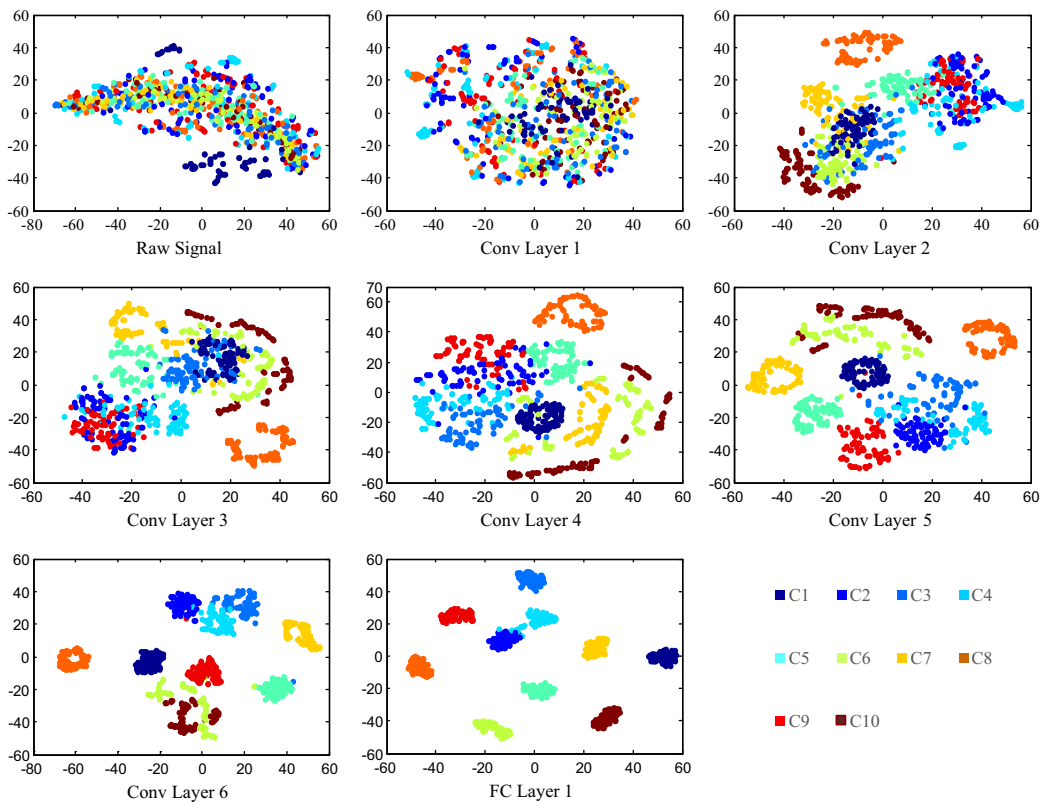
Fig. 12. Feature visualization via t-SNE: feature representations for all test signals extracted from raw signal, six convolutional layers and the last fully connected layer respectively [112].

minimizing the probability distribution distance. To validate the efficacy of domain adaptation, 3 datasets including CWRU dataset, IMS dataset, and railway locomotive bearing dataset are employed. By training on one of the three datasets and testing on another one, an average accuracy of 86.3% is achieved, which has surpassed the conventional CNN of 53.1%, and two existing domain adaptation frameworks of 75.6% [157] and 78.8% [158].

A novel framework WDCNN (deep CNN with wide first-layer kernels) combined with adaptive batch normalization (AdaBN) was proposed in [162]. Taking the raw vibration signal as input, the new framework is based on a CNN architecture with wide kernels (64) in the first convolutional layer to better suppress the high frequency noise. Then domain adaptation is implemented by extracting the mean and variance of the target domain signals and passing them to AdaBN. The CWRU dataset is used to conduct cross-domain experiments by training the proposed WDCNN in one working condition and testing in another one. An average accuracy of 90.0% is achieved, which is further improved to 95.9% by mixing with AdaBN, outperforming the conventional FFT-DNN method of 78.1%. When tested in a noisy environment (with additive white Gaussian noise), WDCNN with AdaBN achieves a 92.65% accuracy under a -4 dB SNR, in comparison to 66.95% without AdaBN.

Deep generative network can also be combined with domain adaptation to produce a novel framework. In [163], a 2-stage structure i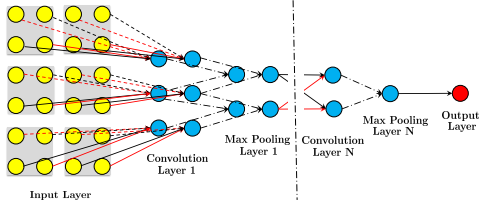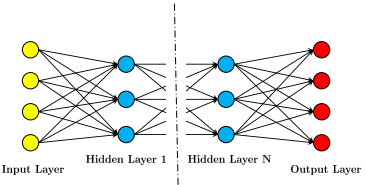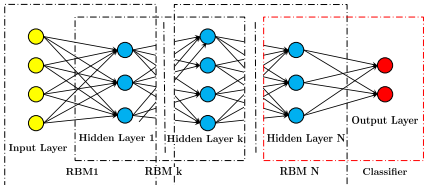s proposed. In the first stage, an 8-layer CNN component including 3 convolutional layers with basic classifiers is trained as the feature extractor to optimize the classification error under source supervision. Then $N_c - 1$ (N_c is the number of classes) CNN components, each of which consists of 3 convolutional layers and 3 dropout layers, are trained to minimize the maximum mean discrepancy. In the second stage, with the feature extractor trained in the first stage, a cross-domain classifier is trained to generate the final diagnosis result.

G. Other Variants

There are also many other DL variants implemented to cope with some of the open issues in the field of bearing fault diagnostics. Some of the selected variants are summarized as follows.

1) Variational Autoencoders: Proposed by Kingma *et al.* [164], the variational auto-encoder (VAE) is different from other autoencoder variants in that it uses the variational inference to generate a latent representation of the data, and impose a distribution over the latent variables and the data itself. Compared to some general class DL algorithms, the implementation of VAE in bearing fault detection is a relatively new domain. A representative work of such is presented in [165], where a fully unsupervised deep VAE-based approach is proposed to tackle the high dimensionality of data used for failure diagnosis. Specifically, the VAE is able to extract discriminative features from the high-dimensional input data to form their corresponding low-dimensional latent space

TABLE II
A SUMMARY OF DIFFERENT DEEP LEARNING ARCHITECTURE.

Architecture	Description	Characteristics
	Convolutional Neural Network • Well-suited for 2-D data, i.e., images, thus 1-D temporal data need to be pre-processed to form 2-D vectors • ReLU after convolutional layers helps accelerate the convergence speed. • Many variants have been proposed: ADCNN [105], LiftingNet [114], and inception net [117], etc.	**Pros:** • Few neuron connections required with respect to a typical ANN. • The classical CNN exhibits a good denosing capability [107]. **Cons:** • May require many layers to find an entire hierarchy. • May require a large labeled dataset.
	Deep Autoencoder • Mainly designed for feature extraction or dimension reduction. • Unsupervised learning method aiming at reconstructing the input vector. • Many variants have been proposed: stacked denoising AE [125], deep ensemble AE [128], and stacked sparse AE [130].	**Pros:** • Does not require labeled data. • Many AE variants can make the algorithm more noise-resilient and robust. **Cons:** • Requires a pre-training stage. • Training may suffer from the vanishing of errors.
	Deep Belief Network • Composed of RBMs where each sub-network's hidden layer serves as the visible layer for the next. • Has undirected connections just at the top two layers. • Allows unsupervised and supervised training of the network.	**Pros:** • Proposes a layer-by-layer greedy learning strategy to initialize the network. • Tractable inferences maximize the likelihood directly. **Cons:** • Training may be computationally expensive due to the initialization process and sampling stage.
	Recurrent Neural Network • An ANN capable of analyzing 1-D sequential or temporal data streams. • LSTM re-vibrated the application of RNNs. • Suitable for applications where the output depends on the previous computations.	**Pros:** • Memorizes sequential events. • Capable of modeling time dependencies. • Capable of receiving inputs of variable lengths. **Cons:** • Frequent learning issues due to gradient vanishing/exploding.
	Generative Adversarial Network • Composes of a generator and a discriminator. Originally designed to generate images that imitate real photos. • Applied for data augmentation in labeled data scarce applications • Also used in classification tasks, usually in with semi-supervised manner.	**Pros:** • Requires almost no modifications when transferring to new applications. • Requires no Monte Carlo approximations to train. • Does not introduce deterministic bias. **Cons:** • GAN training is unstable as it requires finding the Nash equilibrium of a game. • Hard to learn to generate discrete data, such as text.

representations. The experimental results show that the VAE is a more competent and promising tool for dimensionality reduction than PCA. Besides discriminative feature learning, it is also worthwhile exploring its generative capabilities for fault diagnosis in the context of semi-supervised learning.

2) Capsule Neural Networks: The capsule network is a new deep learning architecture proposed by Hinton *et al.* in [166], which has a strong capability to identify the position and orientation relationship of features through the capsule module. Meanwhile, its relatively simple structure with a limited number of parameters significantly promotes the model generalization.

For instance, in [167], a capsule network with inception block is proposed aiming at improving the model generalization capability. This goal is achieved by employing an inception-module-augmented capsule network to adopt different working conditions, upon taking the two-dimensional short-time Fourier transform graph of the raw data as input. Besides, a regression branch is added to predict the size of the bearing defect. The experiments showcase better domain adaptation performance than state-of-the-art algorithms when training and testing on bearing fault data collected at different work loads. The authors in [168] also proposes a deep capsule network with stochastic delta rule (DCN-SDR) for bearing fault diagnosis under varying working conditions and noisy environment. The network first receives raw temporal signal as input, and then extracts noise-immune representative features via incorporating a noise injection module, which is a regularization method based on SDR. The superiority of the proposed architecture is verified through extensive experiments and the subsequent feature visualization via t-SNE. Similarly, a capsule network combined with the Xception module (XCN), an extreme version of inception module, is developed in [169], aiming at improving the classification accuracy of the proposed variant of the capsule network. Trained on ideal laboratory conditions and tested on an actual system setup, the proposed diagnostic model delivers improved classification accuracy, robustness, and training speed.

3) Siamese Neural Networks: Originally proposed by Bromley and LeCun [170] in the early 1990s, the siamese neural network is designed to solve signature verification as an image matching problem. A siamese neural network consists of twin networks, which compares distinct inputs and rank similarities between them. With a growing interest in few-shot learning over the recent years due to insufficient data, Koch et al. [171] implemented the siamese neural network for one-shot image recognition, which inspired the later work of applying this similar siamese network structure to bearing fault diagnostics [172]. Specifically, in the specific siamese network model in [172], two identical networks are set up to take in sample pairs of the same or different categories, which can measure the distance of the two feature vector outputs to determine their similarity. Compared to the WDCNN benchmark with a limited number of training samples below 200, the experimental result reveals an approximately 5% increase of accuracy for the siamese net based one-shot learning, and

a 10% increase for five-shot learning.

4) Others: There are also a number of other variants for bearing fault diagnostics, either based on novel DL frameworks or mixture of multiple DL methods listed above. For example, in [173], a new large memory storage retrieval (LAMSTAR) neural network is proposed with 1 input layer, 40 input SOM modules as hidden layers, and 1 decision SOM module as the output layer. More accurate classification results compared to the conventional CNN are reported at various operating conditions, especially at low speeds. In [174], the DBN and SAE are applied simultaneously to identify the presence of a bearing fault. Other examples include a mixture of CNN and DBN [175], a deep residual network (DRN) [176], [177], a deep stack network [178], a RNN based auto-encoders [179], sparse filtering [180], etc.

V. Discussions on Deep Learning Algorithms for Bearing Fault Diagnosis

A. Automated Feature Extraction and Selection

As opposed to feature engineering of ML algorithms, which manually selects features that preserve the discriminative characteristics of the data, the DL based algorithms can learn the discriminative feature representation directly from input data in an *end-to-end* manner. The DL based approach does not require human expertise or prior knowledge of the problem, and is therefore advantageous in bearing fault diagnosis, where it is sometimes challenging to determine the fault characteristic features accurately. Specifically, DL methods perform feature learning from raw data and classification in a simultaneous and intertwined manner, as illustrated in the cluster visualization results of multiple convolutional layers in Fig. 12. A glimpse of the clustering effect can be observed in convolutional layer C2; and it becomes increasingly apparent in later convolutional layers. For comparison reasons, many DL based papers also present results using classical ML methods with human engineered features for bearing fault detection. The majority of DL based methods are reported to outperform traditional ML methods, especially in the presence of external noise and frequent change of operating conditions.

B. Comparison of Different DL Algorithms for Bearing Fault Diagnostics

Thus far, several types DNN architectures and their applications to bearing fault diagnostics have been extensively discussed, and TABLE II briefly describes the pros and cons of the commonly used deep learning approaches in the field of bearing fault diagnostics. The decision to choose which specific DL algorithm or which specific variant can be customized based on the specific setup environment, the data size, and the number and type of sensors installed. Details on algorithm customization and recommendation will be provided in Section VI.

C. Comparison of DL Algorithm Performance using the CWRU dataset

A systematic comparison of the classification accuracy of different DL algorithms employing the CWRU bearing dataset

TABLE III
COMPARISON OF CLASSIFICATION ACCURACY ON CASE WESTERN RESERVE UNIVERSITY BEARING DATASET WITH DIFFERENT DL ALGORITHMS.

Reference	Feature extraction algorithms	No. hidden layers	Classifier	Characteristics	Training sample percentage	Average accuracy
[105]	Adaptive CNN	3	Softmax	Predict fault size	50%	97.90%
[107]	CNN	4	Softmax	Noise-resilient	90%	92.60%
[108]	CNN	4	Softmax	Sensor fusion	70%	99.40%
[109]	CNN based on LeNet-5	8	FC layer	Better feature extraction	83%	99.79%
[110]	Deep fully connected CNN	8	Connectionist temporal classification	Validation with Actual filed test data	78%	99.22%
[111]	Multi-scale deep CNN	9	Softmax	Reduce training time	90%	98.57%
[112]	CNN with training interface	13	Softmax	Adapt to load change	96%	95.50%
[113]	IDS-CNN	3	Softmax	Adapt to load change	80%	98.92%
[114]	CNN-based LiftingNet	6	FC layer	Adapt to speed change	50%	99.63%
[115]	PSPP-CNN	9	Softmax	Adapt to speed change	67%	99.19%
[116]	AOCNN with SF	4	Softmax	Reduce training set %	5%	99.19%
[123]	SAE	3	ELM	Adapt to load change	50%	99.61%
[124]	SAE	3	ELM	Reduce training time	50%	99.83%
[125]	Stacked denoising AE	3	N/A	Noise-resilient	50%	91.79%
[126]	SDAE	3	Softmax	Noise-resilient	80%	99.83%
[128]	Ensemble deep AE	3	Softmax	Better feature extraction	67%	99.15%
[129]	Deep wavelet AE	3	ELM	Reduce training time	67%	95.20%
[130]	Stack sparse AE	2	N/A	Data compression	N/A	97.47%
[131]	SAE-local connection network	2	Softmax	Shift-invariant features	25%	99.92%
[132]	SAE	3	SVM	Online diagnosis	N/A	95.10%
[133]	SDAE	8	Gath-Geva (GG)	Noise-resilient	N/A	93.30%
[134]	Winner-take-all AE	2	Gath-Geva (GG)	Noise-resilient	N/A	97.27%
[138]	dual-tree complex wavelet	5	N/A	Adaptive DBN	67%	94.38%
[139]	DBN	2	Softmax	Adapt to load change	N/A	98.80%
[140]	DBN with ensemble learning	4	Sigmoid	Accurate & robust	N/A	96.95%
[146]	CNN-LSTM	3	Softmax	Accurate	83%	99.60%
[147]	Deep RNN	3	N/A	Accurate	60%	94.75%
[151]	DCGAN	8	SVM	Data augmentation	96%	86.33%
[153]	CatAAE	11	Softmax	Adapt to load changes	91%	90.68%
[154]	A2CNN	27	Softmax	Domain adaptation	N/A%	99.21%
[155]	GAN+SDAE	8	Softmax et al.	Data augmentation	50-78%	99.20%

is presented in TABLE III. As can be readily observed, the minimum number of hidden layers for all of the networks is 2, indicating the complete network has at least 4 layers incorporating the input and output layers. The maximum hidden layer size can be as large as 13 in [114], representing a very deep network that requires more time in the training process. The selection criterion for the number of hidden layers is to count the layers that are part of the model's architecture, while excluding the input and the output layer. Based on this criterion, in a CNN we count each convolutional layer and each pooling layer as an effective hidden layer, and disregard any of the dropout layer, since it is a regularization technique that only affects the training process (during evaluation, it is not active, otherwise the weights of the network will be larger than normal). For a GAN, we count all of the hidden layers in both the generator and the discriminator.

The test accuracy of all of the DL algorithms are above 95%, which validates the feasibility and effectiveness of applying deep learning to bearing fault diagnostics. However, it is worthwhile to mention that these specific values of test accuracy cannot be used as the sole indicator to compare the effectiveness of different algorithms for the following reasons:

1) *Generalization:* Some of the DL methods with an astonishing accuracy over 99% are generally applied on a very specific dataset at a fixed operating condition,

i.e., when the motor speed is 1,797 rpm and the load is 2 hp. However, this accuracy may suffer significantly under the influence of noise and variation of the motor's speed and load, which unfortunately can be a common issue in practical applications. This is in spite of the relatively strong robustness to noise disturbances of the original DL algorithms (CNN, SAE, DBN, etc.), and their capabilities to learn fault features through a general-purpose learning architecture. It is also reported in [107] that the conventional CNN has a better built-in denoising mechanism compared to other classical DL algorithms such as AE. Due to this limitation, some papers applied the stacked denoising AE (SDAE) [125], [126], [133] to increase AE's noise resilience under a small SNR, i.e., SNR = 5 or 10.

2) *Unbalanced Sampling:* Regarding the selection of training samples from the CWRU dataset, many papers did not guarantee a balanced sampling, which means the ratio of data samples selected from the healthy condition and the faulty condition is not close to 1:1. In case of a significant unbalance, accuracy should not be used as the only metric to evaluate an algorithm [154]. Compared with accuracy, other metrics, such as *precision*, *recall* and *F1-Score*, should be introduced to provide more details for evaluating the reliability of a fault identification net-

work. In addition, if the majority of training set are data from the healthy condition, many of the learnt features cannot fully indicate various fault conditions. Therefore, provided that the training data is highly unbalanced, it would be very challenging to apply the DL classifier trained with laboratory data to identify a bearing fault in practical applications, even if the DL framework adopts transfer learning with domain adaptations.

3) *Randomness:* Even when these DL methods are using the same dataset to perform classification, the percentage of training data and test data can be different, which unavoidably affects the trustability of the comparison between different approaches. What's more, even if this data distribution is identical, the training and test data might be randomly selected from the CWRU bearing dataset. Therefore this comparison is not performed on the common ground, since the classification accuracy is subject to change even with the same algorithm due to the randomness in selecting the training and the test set.

4) *Accuracy saturation:* Most of the existing DL algorithms can achieve an excellent classification accuracy of over 95% using the CWRU dataset, even with the classical CNN without any add-on architectures, which indicates that this dataset contains relatively simple features that can be easily extracted by a variety of DL methods. In fact, all of the bearing defects in the CWRU dataset are manually drilled or engraved, which are much easier to detect than the realistic bearing spalls or general roughness due to aging. Therefore, various perturbations adding on the original dataset needs to be performed to evaluate more advanced functionalities of DL algorithms, i.e., the CWRU data combined with random noise to test an algorithm's denoising capability.

All of the factors above would make the classification accuracy of different DL algorithms less convincing.

VI. Suggestions, Challenges, and Future Work Directions

A. Recommendations and Suggestions

The successful implementation of machine learning and deep learning algorithms on bearing fault diagnostics can be attributed to the strong correlations among features that follow the law of physics. For engineers and researchers considering applying ML or DL methods to solve their bearing detection problems at hand, the authors suggest the following sequences to make the best algorithm selection.

1) *Setup environment:* The first thing we recommend is to thoroughly examine the working environment and all of the possible operating conditions, for example, indoor or outdoor, operating at a fixed operating point or multiple speeds and loads. For the simplest case with an indoor and a single operating point setup, some classical ML methods or even the frequency based analytical model should suffice. For applications that are more prone to external disturbances or having multiple operating points, such as motors fed by VFD converters in electric vehicles,

more advanced deep learning approaches should be employed. Specifically, when the workbench is exposed to a noisy environment, which induces a relatively small SNR, certain denoising blocks and extra hidden layers should be added to increase the noise-resiliency and robustness of the deep neural net.

2) *Sensors:* Then we would need to check the number and type of sensors to be mounted close to the bearing. For the traditional frequency based and classical ML methods, one or two vibrations sensors mounted close to the bearing should be sufficient. For deep learning based approaches, due to the fact that many algorithms such as CNN are mainly developed for computer vision to handle 2-D image data, multiple 1-D time-series data obtained by multiple sensors in the bearing setup need to be stacked together to form this 2-D data. Alternatively, some prepossessing functions, such as the wavelet packet decomposition (WPD), need to be applied before the data is transferred to the deep neural net. Therefore, it would be better to have more than two vibration sensors installed at the same time. In addition, other types of sensors such as acoustic emission and stator current can be installed to form a multi-physics dataset to further improve the accuracy and robustness of the proposed classifier, especially in the midst of frequent and abrupt shifts of operating conditions.

3) *Data size:* If the size of the collected dataset is not sufficient to train a deep learning algorithm with a good level of generalization, specific algorithms should be selected that can make the most out of the data and computing resources available. For example, dataset augmentation techniques such as GAN, and data random sampling techniques with replacement such as Boostrapping, can be readily implemented. Before actually collecting data from the bearing setup, it is advised to interpret the required sample size beforehand [181] by considering how accurate the classification result needs to be. With a small labeled dataset, another promising routine is to leverage the unlabeled dataset, if possible, and apply the semi-supervised learning paradigm by combining the supervised and unsupervised learning approaches.

B. Current Challenges

Despite the extensive effort and the large number of academic papers devoted to this field, there are still some major challenges that need to be tackled to successfully apply ML and DL algorithms to real-world applications:

1) *Knowledge transfer from laboratories to the real world:* The majority of work included in this review is using publicly available dataset collected from laboratories setups to train their customized ML or DL algorithms. However, it would be ideal to be able to transfer the learned network structure and parameters to detect bearing faults from previously unseen setups, and a very promising example would be learning to predict naturally occuring bearing faults in the real-world by only using data collected from artificial faults in the lab. However, there are still many

technique details that need to be tackled to accomplish this ambitious goal.

2) *Limited labels:* For bearing fault detection, it is often times much easier to collect a large amount of data than to accurately obtain their corresponding labels, and this is especially the case for those faults that evolved naturally over time. Specifically, it is not easy to determine precisely when the first trace of a fault shows up and how long it lasts at the incipient stage.

3) *Data imbalance:* In certain occasions it can be challenging or expensive to collect a sufficient amount of data at various bearing faulty conditions to effectively train DL algorithms, while the vast majority of data collected would be at the healthy condition, which in fact does not significantly contribute to training an effective and robust bearing fault classifier.

4) *Noisy data:* Most of the existing work employing DL techniques for bearing fault diagnosis relies on vibration data collected from accelerometers in a laboratory environment. However, in real industrial scenarios such as wind turbines, a large amount of environmental vibration, resonance, or noise may take place. Therefore, it is still an open question if these DL algorithms, while being trained using the vibration data alone, can still deliver satisfactory fault detection performances if the collected data is contaminated by noise.

C. Future Work Directions

Regarding future research directions, the authors suggest the following methodologies and algorithms that might be helpful to address the aforementioned challenges:

1) *Transfer learning: Transfer learning is a promising technique to transfer the knowledge and experience learned from existing datasets to help identify unforeseen bearing fault conditions at different setups in real-world applications. Typical transfer learning techniques include domain randomization and domain adaptation, which can effectively increase the diversity of the source domain (existing datasets), and help facilitate faster learning and better performance in the target domain (real-world cases).*

2) *Semi-supervised learning:* To alleviate the problem of "limited labels", semi-supervised learning can be utilized to make full use of the limited labeled data and the massive unlabeled data. One potential routine is to employ the variational encoder based deep generative model perform variational inference on data with limited labels.

3) *Data augmentation:* Data augmentation techniques such as GAN can be introduced solve the "data imbalance and scarcity" issue by generating more "fake" faulty data to facilitate the training process of DL algorithms. Despite this promising feature, it has been reported in [155] that the accuracy actually declined for some classifiers after incorporating the generated data into the training process. The authors in [155] thus concludes that "the quality of generated spectrum samples" generated by GAN "isn't good enough to provide auxiliary information". Therefore, it would be interesting to explore more powerful generative models, such as BigGAN, to address this open issue.

4) *Few-Shot learning:* Another way to address the "data imbalance and scarcity" problem is to adapt few-shot learning algorithms to achieve a reasonable classification accuracy using a substantially smaller amount of data. This can be combined with transfer learning and domain adaptation to facilitate the use of deep learning for bearing fault diagnostics at the industry level.

5) *Explainability:* Rigorous interpretations of DL in general is not well developed as compared with classical ML methods. Several references, such as [162], and [161], attempted to visualize the learnt CNN kernel to interpret its physical meanings. These studies have provided intuitions on the explainability of DL, but more in-depth investigations and their adaptability to bearing fault diagnostics are necessary.

6) *Sensor fusion:* To solve the potential problem of "noisy data", it might be worthwhile to deploy other types of sensors, such as the load cell, the current sensor, and the acoustic emission sensor, etc., and apply sensor fusion techniques to synthesize these data and improve the robustness of bearing fault diagnosis. Specifically, the use of acoustic sensors should be advocates since it is reported in [173] that in comparison with vibration signals, acoustic emission signals "have certain advantages in detecting incipient faults, capturing and representing". Some existing work on applying acoustic signals to train ML and DL algorithms can be found in [58] and [173], respectively.

VII. CONCLUSIONS

In this paper, a systematic review is presented on the existing literature employing deep learning algorithms to bearing fault diagnostics. Special emphasis is placed on deep learning based approaches that has spurred the interest of the research community over the past five years. It is demonstrated that, despite the fact that deep learning algorithms require a large dataset to train, they can automatically perform adaptive feature extractions on the bearing data without any prior expertise on fault characteristic frequencies or operating conditions, making them promising candidates to perform real-time bearing fault diagnostics. A comparative study is also conducted comparing the performance of many DL algorithm variants using the common CWRU bearing dataset. Finally, detailed recommendations and suggestions are provided in regards to choosing the most appropriate type of DL algorithm for specific application scenarios. Future research directions are also discussed to better facilitate the transition of DL algorithms from laboratory tests to real-world applications.

REFERENCES

[1] M. E. H. Benbouzid, M. Vieira and C. Theys, "Induction motors' faults detection and localization using stator current advanced signal processing techniques," *IEEE Trans. Power Electron.*, vol. 14, no. 1, pp. 14–22, Jan. 1999.

[2] M. El Hachemi Benbouzid, "A review of induction motors signature analysis as a medium for faults detection," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 984–993, Oct. 2000.

[3] P. Zhang, Y. Du, T. G. Habetler and B. Lu,"A survey of condition monitoring and protection methods for medium-voltage induction motors," *IEEE Trans. Ind. Appl.*, vol. 47, no. 1, pp. 34–46, Jan./Feb. 2011.

[4] "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, Jul./Aug. 1985.

[5] "Report of large motor reliability survey of industrial and commercial installations, Part II," *IEEE Trans. Ind. Appl.*, vol. IA–22, no. 4, pp. 865–872, Jul./Aug. 1985.

[6] "Report of large motor reliability survey of industrial and commercial installations: Part 3," *IEEE Trans. Ind. Appl.*, vol. IA–23, no. 1, pp. 153158, Jan./Feb. 1987.

[7] JEMA, "On recommended interval of updating induction motors", 2000 (in Japanese).

[8] T. Harris, *Rolling Bearing Analysis*, 3rd ed. Hoboken, NJ, USA: Wiley, 1991.

[9] J. Harmouche, C. Delpha and D. Diallo, "Improved fault diagnosis of ball bearings based on the global spectrum of vibration signals," *IEEE Trans. Energy Convers.*, vol. 30, no. 1, pp. 376-383, March 2015.

[10] F. Immovilli, A. Bellini, R. Rubini and C. Tassoni, "Diagnosis of bearing faults in induction machines by vibration or current signals: A critical comparison," *IEEE Trans. Ind. Appl.*, vol. 46, no. 4, pp. 1350–1359, Jul./Aug. 2010.

[11] M. Kang, J. Kim and J. Kim, "An FPGA-based multicore system for real-time bearing fault diagnosis using ultrasampling rate AE signals," *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2319–2329, Apr. 2015.

[12] A. Ming, W. Zhang, Z. Qin and F. Chu, "Dual-impulse response model for the acoustic emission produced by a spall and the size evaluation in rolling element bearings," *IEEE Trans. Ind. Electron*, vol. 62, no. 10, pp. 6606–6615, Oct. 2015.

[13] R. R. Schoen, T. G. Habetler, F. Kamran and R. G. Bartfield, "Motor bearing damage detection using stator current monitoring," *IEEE Trans. Ind. Appl.*, vol. 31, no. 6, pp. 1274–1279, Nov./Dec. 1995.

[14] M. Blodt, P. Granjon, B. Raison and G. Rostaing,"Models for bearing damage detection in induction motors using stator current monitoring," *IEEE Trans. Ind. Electron.*, vol. 55, no. 4, pp. 1813–1822, Apr. 2008.

[15] D. López-Pérez and J. Antonino-Daviu, "Application of infrared thermography to failure detection in industrial induction motors: case stories," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 1901–1908, May/June 2017.

[16] E. T. Esfahani, S. Wang and V. Sundararajan, "Multisensor wireless system for eccentricity and bearing fault detection in induction motors," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 3, pp. 818-826, June 2014.

[17] B. Yazici and G. B. Kliman, "Adaptive, on line, statistical method and apparatus for motor bearing fault detection by passive motor current monitoring," U.S. Patent 5 726 905, Sep. 27, 1995.

[18] M. A. Awadallah and M. M. Morcos, "Application of AI tools in fault diagnosis of electrical machines and drives – an overview," *IEEE Trans. Energy Convers.*, vol. 18, no. 2, pp. 245–251, June 2003.

[19] F. Filippetti, G. Franceschini, C. Tassoni and P. Vas, "Recent developments of induction motor drives fault diagnosis using AI techniques," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 994–1004, Oct. 2000.

[20] L. Batista, B. Badri, R. Sabourin, and M. Thomas, "A classifier fusion system for bearing fault diagnosis," *Expert Syst. Appl.*, vol. 40, no. 7, pp. 6788–6797, Dec. 2013.

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

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

[23] M. Cerrada *et al.*, "A review on data-driven fault severity assessment in rolling bearings," *Mech. Syst. Signal Process.*, vol. 99, pp. 169–196, Jan. 2018.

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

[25] S. Zhang, S. Zhang, B. Wang and T. G. Habetler, "Deep Learning Algorithms for Bearing Fault Diagnostics - A Review," in *Proc. IEEE Int. Symp. Diagnostics Electr. Mach., Power Electron. Drives*, Toulouse, France, 2019, pp. 257-263.

[26] Case Western Reserve University (CWRU) Bearing Data Center. Accessed: Dec. 2018, [Online]. Available: https://csegroups.case.edu/bearingdatacenter/pages/welcome-case-western-reserve-university-bearing-data-center-website.

[27] Bearing DataCenter, Paderborn University. Accessed: Dec. 2018, [Online]. Available: https://mb.uni-paderborn.de/kat/forschung/datacenter/bearing-datacenter/.

[28] IEEE PHM 2012 Data Challenge bearing dataset. Accessed: Dec. 2018, [Online]. Available: http://data-acoustics.com/measurements/bearing-faults/bearing-6/.

[29] P. Nectoux *et al.*, "Pronostia: An experimental platform for bearings accelerated life test, in *Proc. IEEE Int. Conf. Prognostics Health Manag.*, Denver, CO, USA, 2012, pp. 18.

[30] J. Lee, H. Qiu, G. Yu, J. Lin, and Rexnord Technical Services, IMS, University of Cincinnati. "Bearing Data Set" NASA Ames Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, Accessed: Jul. 2019, [Online]. Available: http://ti.arc.nasa.gov/project/prognostic-data-repository.

[31] H. Qiu, J. Lee, J. Lin, and G. Yu, "Robust performance degradation assessment methods for enhanced rolling element bearing prognostics, *Adv. ENg. Informat*, vol. 17, no. 3–4, pp. 127–140, 2003.

[32] S. Pan and Q. Yang. "A survey on transfer learning," in *IEEE Transactions on Knowledge and Data Engineering*, October 2009

[33] C. Tan, *et al*. "A survey on deep transfer learning." in *Proc. Int. Conf. Artificial Neural Netw.*. Springer, Cham, 2018.

[34] 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, Dec. 1991.

[35] R. R. Schoen, B. K. Lin, T. G. Habetler, J. H. Schlag and S. Farag, "An unsupervised, on-line system for induction motor fault detection using stator current monitoring," *IEEE Trans. Ind. Appl.*, vol. 31, no. 6, pp. 1280–1286, Nov.-Dec. 1995.

[36] B. Li, M.-Y. Chow, Y. Tipsuwan and J. C. Hung, "Neural-network-based motor rolling bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 47, no. 5, pp. 1060–1069, Oct. 2000.

[37] M. Cococcioni, B. Lazzerini and S. L. Volpi, "Robust diagnosis of rolling element bearings based on classification techniques," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2256–2263, Nov. 2013.

[38] S. Qian, X. Yang, J. Huang and H. Zhang, "Application of new training method combined with feedforward artificial neural network for rolling bearing fault diagnosis," in *Proc. Int. Conf. Mechatronics Mach. Vision in Practice (M2VIP)*, Nanjing, China, 2016, pp. 1–6.

[39] S. Lu and X. Wang, "PCA-based feature selection scheme for machine defect classification," *IEEE Trans. Instrum. Mea.*, vol. 53, no. 6, pp. 1517–1525, Dec. 2004.

[40] X. Zhang, R. Xu, C. Kwan, S.-Y. Liang, Q. Xie and L. Haynes, "An integrated approach to bearing fault diagnostics and prognostics," in *Proc. American Control Conf.*, Portland, OR, USA, 2005, pp. 2750–2755 vol. 4.

[41] M. Xia, F. Kong and F. Hu, "An approach for bearing fault diagnosis based on PCA and multiple classifier fusion," in *Proc. IEEE Joint Int. Inf. Technol. AI Conf.*, Chongqing, China, 2011, pp. 321–325.

[42] J. Harmouche, C. Delpha and D. Diallo, "A global approach for the classification of bearing faults conditions using spectral features," in *Proc. 39th Annu. Conf. IEEE Ind. Electron. Soc. (IECON'13)*, Vienna, Austria, 2013, pp. 7352–7357.

[43] X. Xue and J. Zhou, "A hybrid fault diagnosis approach based on mixed-domain state features for rotating machinery," *ISA Trans.*, vol. 66, pp. 284–295, 2017.

[44] D. He, R. Li, J. Zhu and M. Zade, "Data mining based full ceramic bearing fault diagnostic system using AE sensors," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2022–2031, Dec. 2011.

[45] M. M. Ettefagh, M. Ghaemi and M. Yazdanian Asr, "Bearing fault diagnosis using hybrid genetic algorithm K-means clustering," in *Proc. Int. Symp. Innov. Intell. Syst. Appli. (INISTA)*, Alberobello, Italy, 2014, pp. 84–89.

[46] J. Tian, C. Morillo, M. H. Azarian and M. Pecht, "Motor bearing fault detection using spectral Kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis," *IEEE Trans. Ind. Electron.*, vol. 63, no. 3, pp. 1793–1803, Mar. 2016.

[47] T. W. Rauber, F. de Assis Boldt and F. M. Varejão, "Heterogeneous feature models and feature selection applied to bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, pp. 637–646, Jan. 2015.

[48] A. Rojas and A. K. Nandi, "Detection and classification of rolling-element bearing faults using support vector machines," in *Proc. IEEE Workshop on Mach. Learning Signal Process.*, Mystic, CT, USA, 2005, pp. 153–158.

[49] K. Teotrakool, M. J. Devaney and L. Eren, "Bearing fault detection in adjustable speed drives via a support vector machine with feature selection using a genetic algorithm," in *Proc. IEEE Instrum. Mea. Technol. Conf.*, Victoria, BC, Canada, 2008, pp. 1129–1133.

[50] P.Konar, and P. Chattopadhyay, "Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs)," *Appl. Soft Comput.*, vol. 11, no. 6, pp. 4203–4211, Sep. 2011.

[51] T. Benkedjouh, K. Medjaher, N. Zerhouni and S. Rechak, "Fault prognostic of bearings by using support vector data description," in *Proc. IEEE Conf. Progno. Health Manag.*, Denver, CO, USA, 2012, pp. 1–7.

[52] M. Kang, J. Kim, J. Kim, A. C. C. Tan, E. Y. Kim and B. Choi, "Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis," *IEEE Trans. Power Electron.*, vol. 30, no. 5, pp. 2786–2797, May 2015.

[53] S. E. Pandarakone, Y. Mizuno and H. Nakamura, "Distinct fault analysis of induction motor bearing using frequency spectrum determination and support vector machine," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3049–3056, May-June 2017.

[54] F. B. Abid, S. Zgarni and A. Braham, "Bearing fault detection of induction motor using SWPT and DAG support vector machines," in *Proc. 42nd Annu. Conf. IEEE Ind. Electron. Soc. (IECON'16)*, Florence, Italy, 2016, pp. 1476–1481.

[55] H. Hassani, J. Zarei, M. M. Arefi and R. Razavi-Far, "zSlices-based general type-2 fuzzy fusion of support vector machines with application to bearing fault detection," *IEEE Trans. Ind. Electron.*, vol. 64, no. 9, pp. 7210–7217, Sept. 2017.

[56] S. Kang, L. Cui, Y. Wang, F. Li and V. I. Mikulovich, "Method of assessing the multi-state of a rolling bearing based on CFOA-HSVM two measures combination," in *Proc. Prog. Syst. Health Manag. Conf. (PHM)*, Beijing, China, 2017, pp. 1–5.

[57] B. R. Nayana and P. Geethanjali, "Analysis of statistical time-domain features effectiveness in identification of bearing faults from vibration signal," *IEEE Sensors J.*, vol. 17, no. 17, pp. 5618–5625, 1 Sept.1, 2017.

[58] M. Elforjani and S. Shanbr, "Prognosis of bearing acoustic emission signals using supervised machine learning," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5864–5871, July 2018.

[59] W. Song and J. Xiang, "A method using numerical simulation and support vector machine to detect faults in bearings," in *Proc. 2017 Int. Conf. Sens. Diagn., Prog. Control (SDPC)*, Shanghai, China, 2017, pp. 603–607.

[60] S. E. Pandarakone, Y. Mizuno and H. Nakamura, "Evaluating the progression and orientation of scratches on outer-raceway bearing using a pattern recognition method," *IEEE Trans. Ind. Electron.*, vol. 66, no. 2, pp. 1307–1314, Feb. 2019.

[61] F. Ben Abid, S. Zgarni and A. Braham, "Distinct bearing faults detection in induction motor by a hybrid optimized SWPT and aiNet-DAG SVM," *IEEE Trans. Energy Convers.*, vol. PP, no. PP, pp. PP–PP, 2018.

[62] C. Chen, B. Zhang and G. Vachtsevanos, "Prediction of machine health condition using neuro-fuzzy and Bayesian algorithms," *IEEE Trans. Instrum. Mea.*, vol. 61, no. 2, pp. 297–306, Feb. 2012.

[63] A. Soualhi and S. Taleb, "Data fusion for fault severity estimation of ball bearings," in *Proc. IEEE Int. Conf. Ind. Technol. (ICIT)*, Lyon, France, 2018, pp. 2105–2110.

[64] A. S. Raj and N. Murali, "Early classification of bearing faults using morphological operators and fuzzy inference," *IEEE Trans. Ind. Electron.*, vol. 60, no. 2, pp. 567–574, Feb. 2013.

[65] J. Wu, C. Wu, S. Cao, S. W. Or, C. Deng and X. Shao, "Degradation data-driven time-to-failure prognostics approach for rolling element bearings in electrical machines," *IEEE Trans. Ind. Electron.*, vol. 66, no. 1, pp. 529–539, Jan. 2019.

[66] X. Zhang, J. Kang and T. Jin, "Degradation modeling and maintenance decisions based on Bayesian belief networks," *IEEE Trans. Rel.*, vol. 63, no. 2, pp. 620–633, June 2014.

[67] J. Qu, Z. Zhang, and T. Gong, "A novel intelligent method for mechanical fault diagnosis based on dual-tree complex wavelet packet transform and multiple classifier fusion," *NeuroComput.*, vol. 171, pp. 831–853, 2017.

[68] W. Li, S. Zhang and G. He, "Semisupervised distance-preserving self-organizing map for machine-defect detection and classification," *IEEE Trans. Instrum. Mea.*, vol. 62, no. 5, pp. 869–879, May 2013.

[69] Z. Zhou, J. Chen, Y. Zi and X. Chen, "Multi-domain description method for bearing fault recognition in varying speed condition,"

[70] Q. Tong *et al.*, "A fault diagnosis approach for rolling element bearings based on RSGWPT-LCD bi-layer screening and extreme learning machine," in *IEEE Access*, vol. 5, pp. 5515–5530, 2017.

[71] Q. Hu, A. Qin, Q. Zhang, J. He and G. Sun, "Fault diagnosis based on weighted extreme learning machine with wavelet packet decomposition and KPCA," *IEEE Sensors J.*, vol. 18, no. 20, pp. 8472–8483, 15 Oct. 15, 2018.

[72] F. Shen, C. Chen, R. Yan and R. X. Gao, "Bearing fault diagnosis based on SVD feature extraction and transfer learning classification," in *Proc. Prog. Syst. Health Manag. Conf. (PHM)*, Beijing, China, 2015, pp. 1–6.

[73] R. Zhang, H. Tao, L. Wu and Y. Guan, "Transfer learning with neural networks for bearing fault diagnosis in changing working conditions," in *IEEE Access*, vol. 5, pp. 14347–14357, 2017.

[74] A New Transfer Learning Method and its Application on Rotating Machine Fault Diagnosis Under Variant Working Conditions

[75] X. Zhao and M. Jia, "Fault diagnosis of rolling bearing based on feature reduction with global-local margin Fisher analysis," *Neurocomput.*, vol. 315, pp. 447–464, 2018.

[76] X. Jin, M. Zhao, T. W. S. Chow and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2441–2451, May 2014.

[77] M. Cococcioni, B. Lazzerini and S. L. Volpi, "Robust diagnosis of rolling element bearings based on classification techniques," *IEEE Trans. Ind. Informat.*, vol. 9, no. 4, pp. 2256–2263, Nov. 2013.

[78] Z. Wang, Q. Zhang, J. Xiong, M. Xiao, G. Sun and J. He, "Fault diagnosis of a rolling bearing using wavelet packet denoising and random forests," *IEEE Sensors J.*, vol. 17, no. 17, pp. 5581–5588, 1 Sept.1, 2017.

[79] T. Yang, H. Pen, Z. Wang and C. S. Chang, "Feature knowledge based fault detection of induction motors through the analysis of stator current data," *IEEE Trans. Instrum. Mea.*, vol. 65, no. 3, pp. 549-558, Mar. 2016.

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

[81] M. D. Prieto, G. Cirrincione, A. G. Espinosa, J. A. Ortega and H. Henao, "Bearing fault detection by a novel condition-monitoring scheme based on statistical-time features and neural networks," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3398–3407, Aug. 2013.

[82] B. Yao, P. Zhen, L. Wu and Y. Guan, "Rolling element bearing fault diagnosis using improved manifold learning," in *IEEE Access*, vol. 5, pp. 6027–6035, 2017.

[83] B. Wang, H. Pan, and W. Yang, "A complementary approach for fault diagnosis of rolling bearing using canonical variate analysis based short-time energy feature," *J. Vib. Control*, vol. 24, no. 18, pp. 4195–4210, Sep. 2018.

[84] B. Zhang, C. Sconyers, C. Byington, R. Patrick, M. E. Orchard and G. Vachtsevanos, "A probabilistic fault detection approach: Application to bearing fault detection," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 2011–2018, May 2011.

[85] J. Yu, "Local and nonlocal preserving projection for bearing defect classification and performance assessment," *IEEE Trans. Ind. Electron.*, vol. 59, no. 5, pp. 2363–2376, May 2012.

[86] H. Xue, M. Wang, Z. Li and P. Chen, "Fault feature extraction based on artificial hydrocarbon network for sealed deep groove ball bearings of in-wheel motor," in *Proc. 2017 Progn. Syst. Health Manag. Conf. (PHM-Harbin)*, Harbin, China, 2017, pp. 1–5.

[87] R. Razavi-Far, M. Farajzadeh-Zanjani and M. Saif, "An integrated class-imbalanced learning scheme for diagnosing bearing defects in induction motors," *IEEE Trans. Ind. Informat.*, vol. 13, no. 6, pp. 2758–2769, Dec. 2017.

[88] J. Tian, M. H. Azarian, M. Pecht, G. Niu and C. Li, "An ensemble learning-based fault diagnosis method for rotating machinery," in *Proc. 2017 Progn. Syst. Health Manag. Conf. (PHM-Harbin)*, China, 2017, pp. 1–6.

[89] V. Vakharia, V. K. Gupta and P. K. Kankar, "Ball bearing fault diagnosis using supervised and unsupervised machine learning methods," *Int. J. Acoust. Vib.*, vol. 20, no. 4, pp. 244–250, 2015.

[90] D. He, R. Li and J. Zhu, "Plastic bearing fault diagnosis based on a two-step data mining approach," *IEEE Trans. Ind. Electron.*, vol. 60, no. 8, pp. 3429–3440, Aug. 2013.

in *Proc. 42nd Annu. Conf. IEEE Ind. Electron. Soc. (IECON'16)*, Florence, Italy, 2016, pp. 423–428.

[91] M. Van and H. Kang, "Bearing-fault diagnosis using non-local means algorithm and empirical mode decomposition-based feature extraction and two-stage feature selection," *IET Sci., Mea. & Technol.*, vol. 9, no. 6, pp. 671–680, Sep. 2015.

[92] J. B. Ali, N. Fnaiech, L. Saidi, B. Chebel-Morello, and F. Fnaiech, "Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals," *Appl. Acoust.*, vol. 89, pp. 16–27, Mar. 2015.

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

[94] Q. Fu, B. Jing, P. He, S. Si and Y. Wang, "Fault feature selection and diagnosis of rolling bearings based on EEMD and optimized Elman_AdaBoost algorithm," *IEEE Sensors J.*, vol. 18, no. 12, pp. 5024–5034, 15 June 15, 2018.

[95] C. Chen, F. Shen, and R. Yan, "Topic correlation analysis for bearing fault diagnosis under variable operating conditions," *J. Phys.: Conf. Ser.*, vol. 842, art no. 012045, pp. 1–9, 2017.

[96] X. Zhao and M. Jia, "A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering and adaptive feature selection," *Knowl.-Based Syst.*, vol. 116, pp. 1–12, 2017.

[97] Y. Lu, R. Xie, and S.-Y. Liang, "Adaptive online dictionary learning for bearing fault diagnosis," *Int. J. Adv. Manuf. Technol.*, pp. 1–8, Oct. 2018.

[98] Sparse classification based on dictionary learning for planet bearing fault identification

[99] Amazon Mechanical Turk. Accessed: Dec. 2018, [Online]. Available: https://www.mturk.com/.

[100] A. Ng, "CS229 course notes: Deep learning", *Stanford University*, Fall 2018.

[101] NVIDIA TESLA: The Worlds Most Advanced Data Center GPUs. Accessed: Dec. 2018, [Online]. Available: http://wild.link/nvidia/ANrTCA.

[102] M. Raghu, B. Poole, J Kleinberg, S. Ganguli, and J. Sohl-Dickstein, "On the expressive power of deep neural networks," *arXiv preprint arXiv:1606.05336*, 2016.

[103] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biolog. Cybernetics*, vol. 36, pp. 193–202, 1980.

[104] S. Singh, C. Q. Howard, and C. Hansen, "Convolutional neural network based fault detection for rotating machinery," *J. Sound Vib.*, vol. 377, pp. 331–345, 2016.

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

[106] R. Liu, G. Meng, B. Yang, C. Sun and X. Chen, "Dislocated time series convolutional neural architecture: An intelligent fault diagnosis approach for electric machine," *IEEE Trans. Ind. Informat.*, vol. 13, no. 3, pp. 1310–1320, June 2017.

[107] C. Lu, Z. Wang, and Bo Zhou, "Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification," *Adv. Eng. Informat.*, vol. 32, pp. 139–157, 2017.

[108] 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.

[109] L. Wen, X. Li, L. Gao and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, July 2018.

[110] W. Zhang, F. Zhang, W. Chen, Y. Jiang and D. Song, "Fault state recognition of rolling bearing based fully convolutional network," *Comput. in Sci. & Eng.*, vol. PP, no. PP, pp. PP–PP, 2018.

[111] Z. Zhuang and Q. Wei, "Intelligent fault diagnosis of rolling bearing using one-dimensional multi-scale deep convolutional neural network based health state classification," in *Proc. 2018 IEEE 15th Int. Conf. Netw., Sens. & Control (ICNSC)*, Zhuhai, China, 2018, pp. 1–6.

[112] W. Zhang, C. Li, G. Peng, Y. Chen, Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–458, 2018.

[113] S. Li, G. Liu, X. Tang, J. Lu, and J. Hu, "An ensemble deep convolutional neural network model with improved D-S evidence fusion for bearing fault diagnosis," *Sensors*, vol. 17, art. no. 1729, Jul. 2017.

[114] J. Pan, Y. Zi, J. Chen, Z. Zhou and B. Wang, "LiftingNet: A novel deep learning network with layerwise feature learning from noisy mechanical data for fault classification," *IEEE Trans. Ind. Electron.*, vol. 65, no. 6, pp. 4973–4982, June 2018.

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

[116] W. Qian, S. Li, J. Wang, Z. An, ans X. Jiang, "An intelligent fault diagnosis framework for raw vibration signals: adaptive overlapping convolutional neural network," *Meas. Sci. Technol.*, vol. 29, no. 6, art. no 095009, Aug. 2018.

[117] 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," *Neurocomput.*, vol. 294, pp. 61–71, 2018.

[118] L. Guo, Y. Lei, N. Li and S. Xing, "Deep convolution feature learning for health indicator construction of bearings," in *Proc. 2017 Progn. Syst. Health Manag. Conf. (PHM-Harbin)*, Harbin, China, 2017, pp. 1–6.

[119] X. Ding and Q. He, "Energy-fluctuated multiscale feature learning with deep ConvNet for intelligent spindle bearing fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 8, pp. 1926–1935, Aug. 2017.

[120] J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Netw., vol. 61, pp. 85–117, Jan. 2015.

[121] D. H. Ballard, "Modular learning in neural networks," in *Proc. AAAI'87 National Conf. AI*, Seattle, WA, 1987, pp. 279–284.

[122] 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, 2017.

[123] 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.*, vol. 100, pp. 743–765, 2016.

[124] W. Mao, J. He, Y. Li, and Y. Yan, "Bearing fault diagnosis with auto-encoder extreme learning machine: A comparative study," *Proc. Inst. Mech. Eng. C, J. Mech. Eng. Sci.*, vol. 231, no. 8, pp. 1560–1578, 2016.

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

[126] 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.*, 2017, vol. 7. no. 41, pp. 1–17.

[127] 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)*, Houston, TX, USA, 2018, pp. 1–6.

[128] H. Shao, H. Jiang, Y. Lin, and X. Li, "A novel method for intelligent fault diagnosis of rolling bearings using ensemble deep auto-encoders," *Knowl.-Based Syst.*, vol. 119, pp. 200–220, 2018.

[129] H. Shao, H. Jiang, X. Li, S. Wu, "Intelligent fault diagnosis of rolling bearing using deep wavelet auto-encoder with extreme learning machine," *Knowl.-Based Syst.*, vol. 140, pp. 1–14, 2018.

[130] 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.

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

[132] W. Mao, S. Tian, X. Liang, and J. He, "Online bearing fault diagnosis using support vector machine and stacked auto-encoder," in *Proc. IEEE Int. Conf. Prognost. Health Manag. (ICPHM)*, Seattle, WA, 2018, pp. 1–7.

[133] F. Xu, W. P. Tse, and Y.-L. Tse, "Roller bearing fault diagnosis using stacked denoising autoencoder in deep learning and Gath-Geva clustering algorithm without principal component analysis and data label," *Appl. Soft Comput. J.*, vol. PP, pp. PP–PP, 2018.

[134] 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.

[135] Y. Bengio, "Learning deep architectures for AI, *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[136] 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, July 2017.

[137] 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. 2017 Progn. Syst. Health Manag. Conf. (PHM-Harbin)*, Harbin, China, 2017, pp. 1–7.

[138] D. T. Hoang and H. J. Kang, "Rolling bearing fault diagnosis using adaptive deep belief network with dual-tree complex wavelet packet," *ISA Trans.*, vol. 69, pp. 187–201, 2017.

[139] D.-T. Hoang and H.-J. Kang, "Deep belief network and dempster-shafer evidence theory for bearing fault diagnosis," in *Proc. IEEE 27th Int. Sym. Ind. Electron. (ISIE)*, Cairns, QLD, Australia, 2018, pp. 841–846.

[140] 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, art no. 021054, pp. 1–7, 2018.

[141] H. Shao, H. Jiang, H. Zhang and T. Liang, "Electric locomotive bearing fault diagnosis using a novel convolutional deep belief network," *IEEE Trans. Ind. Electron.*, vol. 65, no. 3, pp. 2727–2736, Mar. 2018.

[142] H. Oh, J. H. Jung, B. C. Jeon and B. D. Youn, "Scalable and unsupervised feature engineering using vibration-imaging and deep learning for rotor system diagnosis," *IEEE Trans. Ind. Electron.*, vol. 65, no. 4, pp. 3539–3549, April 2018.

[143] J. Deutsch and D. He, "Using deep learning-based approach to predict remaining useful life of rotating components," *IEEE Trans. Syst., Man, Cybern.*, vol. 48, no. 1, pp. 11–20, Jan. 2018.

[144] 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.

[145] 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," *Neurocomput.*, vol. 240, pp. 98–109, 2017.

[146] H. Pan, X. He, S. Tang, F. Meng, "An improved bearing fault diagnosis method using one-dimensional CNN and LSTM," *J. Mech. Eng.*, vol. 64, no. 7–8, pp. 443–452, 2018.

[147] 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, art. no 065107, May 2018.

[148] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, Y. Bengio, "Generative adversarial nets," in *Prof. Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, Montreal, Quebec, Canada, Dec. 2014, pp. 2672-2680.

[149] K. Fragkiadaki, "CMU 10703 course notes: Deep reinforcement learning and control lecture 14", *Carnegie Mellon University*, Spring 2017.

[150] Y. O. Lee, J. Jo and J. Hwang, "Application of deep neural network and generative adversarial network to industrial maintenance: A case study of induction motor fault detection," in *Proc. IEEE Int. Conf. Big Data*, Boston, MA, 2017, pp. 3248–3253.

[151] Y. Xie, T. Zhang, "Imbalanced learning for fault diagnosis problem of rotating machinery based on generative adversarial networks," in *Proc. 37th Chinese Control Conf. (CCC)*, Wuhan, China, 2018, pp. 3248–3253.

[152] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," arXiv preprint arXiv:1511.06434, 2015.

[153] 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," *Neurocomput.*, vol. 315, pp. 412–424, 2018.

[154] B. Zhang, W. Li, J. Hao, X.-L. Li, and M. Zhang, "Adversarial adaptive 1-D convolutional neural networks for bearing fault diagnosis under varying working condition," eprint arXiv:1805.00778, May 2018.

[155] 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.

[156] S. Suh, H. Lee, J. Jo, P. Lukowicz, and Y.-Oh Lee, "Generative Oversampling Method for Imbalanced Data on Bearing Fault Detection and Diagnosis", *Appl. Sci.*, vol. 9, no. 4, 746, 2019.

[157] E. Tzeng, J. Hoffman, N. Zhang, K. Saenko, and T. Darrell, "Deep domain confusion: Maximizing for domain invariance," arXiv preprint arXiv:1412.3474, 2014.

[158] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," Journal of Machine Learning Research, vol. 17, no. 59, pp. 1–35, 2016

[159] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang. "Deep model based domain adaptation for fault diagnosis," in *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2296–2305, 2017.

[160] W. Zhang, C. Li, G. Peng, Y. Chen and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Processing*, 100 (2018), pp. 439–453.

[161] L. Guo, Y. Lei, S. Xing, T. Yan and N. Li, "Deep Convolutional Transfer Learning Network: A New Method for Intelligent Fault Diagnosis of Machines with Unlabeled Data," *IEEE Trans. Ind. Electron.*, vol. PP, no. PP, pp. PP–PP, 2018.

[162] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors.*, vol. 17, no. 2, 425, 2017.

[163] 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.

[164] D. P. Kingma, and M. Welling, "Auto-encoding variational bayes," arXiv preprint arXiv:1312.6114, 2013.

[165] G. Martin, E. Droguett, V. Meruane, and M. Moura, "Deep variational auto-encoders: A promising tool for dimensionality reduction and ball bearing elements fault diagnosis," *Structural Health Monitoring*, pp. 1092–1128, 2018.

[166] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3859–3869.

[167] 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," *Neurocomput.*, vol. 323, pp. 62–75, 2019.

[168] T, Chen, Z. Wang, X. Yang, and K. Jiang. "A deep capsule neural network with stochastic delta rule for bearing fault diagnosis on raw vibration signals," *Measurement*, vol. 148, 106857, 2019.

[169] Z. Wang *et al.* "A Novel Method for Intelligent Fault Diagnosis of Bearing Based on Capsule Neural Network." *Complexity*, pp. 1-17, 2019.

[170] J. Bromley, I. Guyon, Y. LeCun, E. Sckinger, and R. Shah. "Signature verification using a " siamese" time delay neural network." In *Proc. Adv. neural inf. process. syst. (NIPS)*, pp. 737-744. 1994.

[171] G. Koch, R. Zemel, and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition," *ICML deep learning workshop*, vol. 2. 2015.

[172] 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*, 110895-110904, Jul. 2019.

[173] M. He and D. He, "Deep learning based approach for bearing fault diagnosis," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3057–3065, May/June 2017.

[174] 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, 2017.

[175] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, and S. Wu, "Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–458, 2018.

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

[177] M. Zhao, M. Kang, B. Tang and M. Pecht, "Multi-wavelet coefficients fusion in deep residual networks for fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. PP, no. PP, pp. PP–PP, 2018.

[178] C. Sun, M. Ma, Z. Zhao and X. Chen, "Sparse deep stacking network for fault diagnosis of motor," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 3261–3270, July 2018.

[179] 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.

[180] W. Qian, S. Li, J. Wang, and Q. Wu, "A novel supervised sparse feature extraction method and its application on rotating machine fault diagnosis," *Neurocomput.*, vol. 210, pp. 129–140, 2018.

[181] C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft, and J. Popp, "Sample size planning for classification models," *Analytica Chimica Acta*, vol. 760, pp. 25–33, 2013.