Machine Learning and Deep Learning Algorithms for Bearing Fault Diagnostics – A Comprehensive Review

Shen Zhang*†, Shibo Zhang‡, Bingnan Wang*, and Thomas G. Habetler†
*Mitsubishi Electric Research Laboratories, 201 Broadway, Cambridge, MA 02139, USA
†School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA
‡Computer Science, Northwestern University, Evanston, IL 60201, USA

***Abstract*—In this survey paper, we systematically summarize the current literature on studies that apply machine learning (ML) and data mining techniques to bearing fault diagnostics. Conventional ML methods, including artificial neural network (ANN), principal component analysis (PCA), support vector machines (SVM), etc., have been successfully applied to detecting and categorizing bearing faults since the last decade, while the application of deep learning (DL) methods has sparked great interest in both the industry and academia in the last five years. In this paper, we will first review the conventional ML methods, before taking a deep dive into the latest developments in DL algorithms for bearing fault applications. Specifically, the superiority of the DL based methods over the conventional ML methods are analyzed in terms of metrics directly related to fault feature extraction and classifier performances; the new functionalities offered by DL techniques that cannot be accomplished before are also summarized. In addition, to obtain a more intuitive insight, a comparative study is performed on the classifier performance and accuracy for a number of papers utilizing the open source Case Western Reserve University (CWRU) bearing data set. Finally, based on the nature of the time-series 1-D data obtained from sensors monitoring the bearing conditions, recommendations and suggestions are provided to applying DL algorithms on bearing fault diagnostics based on specific applications, as well as future research directions to further improve its performance.**

***Index Terms*—Bearing fault; diagnostics; machine learning; deep learning; feature extraction**

I. Introduction

Electric machines are widely employed in a variety of industry application processes and electrified transportation systems. However, 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 concerns [1]–[3]. The malfunction of electric machines can be generally attributed to various faults of different categories, which include the drive inverter failures, stator winding insulation breakdown, as well as bearing faults and air gap eccentricity. Several surveys regarding the likelihood of induction machine failures conducted by the IEEE Industry Application Society (IEEE-IAS) [4]–[6] and the Japan Electrical Manufacturers' Association (JEMA) [7] reveal that a bearing fault is the most common fault type that accounts for 30% to 40 % percent of the total faults.

Since a bearing is the most vulnerable component in a motor and drive system, bearing fault detection has been a research frontier for engineers and scientists for the past decades. Specifically, this problem is approached by interpreting a variety of available signals, including vibration [10], [11], acoustic noise [12], [13], stator current [14], [15], thermal-imaging [16], and multiple sensor fusion [17]. 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 the components at the 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 a defect inside the bearing.

However, accurately identifying the occurrence of a bearing fault can be sometimes difficult in practice, especially when the fault is still at its incipient stage due to the small signal-to-noise ratio of the vibration and acoustic signals. In addition, unlike other motor failures (stator inter-turn, broken rotor bar, etc. [3]) that can be accurately determined by the electric signals, the uniqueness of bearing failures lies in its multi-physics nature, incorporating a closed-loop interplay between the initiative mechanical vibration and the corresponding induced harmonics in the electric signals, which further affects the output torque , motor speed, and finally the vibration pattern itself. As a result, the accuracy of the traditional mathematical model-based bearing fault diagnostics can be further influenced by background noise due to external mechanical excited motion, while its sensitivity is also subject to change based on sensor mounting positions and spatial constraints in a highly-compact design. Therefore, a popular alternative approach for bearing fault detection is accomplished by analyzing the stator current [14], [15], which is measured in most motor drives and thus would not bring extra device or installation costs. Despite its advantages such as economic savings and simple implementation, stator current signature analysis can encounter many practical issues. For example, the magnitude of stator currents at bearing fault signature frequencies can vary at different loads, different speeds, and different power ratings of the motors themselves, thus bringing challenges to identify the threshold stator current values to trigger a fault alarm at an arbitrary operating condition. Therefore, a thorough and systematic testing is usually required while the motor is still at the healthy condition, and the healthy data would be collected while the targeted motor is running at different load and speed. This process, summarized as a "Learning Stage" in patent US5726905 [18], is unfortunately tedious and expensive to perform, and needs to be repeated for any motor with a different power rating.

Most of the challenges described above can be attributed to the fact that all the conventional methods rely solely upon the values at the fault characteristic frequencies to determine the presence of a bearing fault. However, there may exist some unique patterns or relationships hidden in the data themselves that can potentially reveal a bearing fault, and these special features can be almost impossible for humans to identify at the first place. Therefore, many researchers began applying various machine learning algorithms, i.e., artificial neural networks (ANN), principal component analysis (PCA), support vector machines (SVM), etc., to better parse the data, learn from them, and then apply what they've learned to make intelligent decisions regarding the presence of bearing faults [30]–[33]. Most of the literature applying these ML algorithms report very satisfactory results with classification accuracy over 90%.

To achieve even better performance and higher classification accuracy under versatile operating conditions or noisy conditions, deep learning based methods are becoming increasingly popular to meet this need. This literature survey incorporates around 160 papers, around 60 of which employed some type of deep learning based approaches. And the number of papers grows steadily, with 2 papers published in 2015, 7 papers in 2016, 17 papers in 2017 and 37 papers in 2018, clearly indicating booming interests in employing DL methods for bearing fault diagnostics in the recent years. Deep learning generally requires extremely large datasets, and some DL networks in computer vision were trained using as many as 1.2 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. For smaller datasets, classical ML algorithms can compete with or even outperform deep learning networks.

In this context, this paper seeks to present a thorough overview on the recent research work devoted to applying machine learning techniques on bearing fault diagnostics. The rest of the paper is organized as below. In Section II, we introduce some of the most popular datasets used for bearing fault detection. Next, in Section III, we look into traditional ML methods, including ANN, PCA, *k*-nearest neighbors (*k*-NN), SVM, etc., with a brief explanation of each method. For the main part of the paper, in Section IV, we take a deep dive into the research frontier of identifying a bearing fault with DL techniques. In this part, we will provide our understanding of the research trend toward DL networks. The advantages of the DL based methods over the conventional ML methods will be analyzed in terms of metrics directly related to fault feature extraction and classifier performances; the new functionalists offered by DL techniques that cannot be accomplished before are also summarized. We will also give detailed analysis to each of major DL techniques, including convolutional neural network (CNN), auto-encoder, deep belief network (DBN), recurrent neural network (RNN), generative adversarial network (GAN), and their applications to bearing fault detection. In Section V, to obtain a more intuitive insight, a comparative study is performed on the different DL algorithms, as well as their classifier performances utilizing the common open source Case Western Reserve University (CWRU) bearing data set. Finally, based on the nature of the time-series 1-D data obtained from sensors monitoring the bearing conditions, recommendations and suggestions on applying DL algorithms for bearing fault diagnostics for specific applications are provided in section VI, as well as future research directions to further improve its performance by adopting more complex datasets and predict the actual bearing fault with classifiers trained with artificially induced faults.

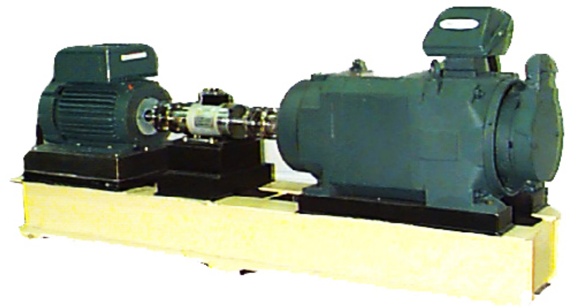


Fig. 1. Experimental setup collecting the CWRU bearing dataset [26].

II. Popular Bearing Fault Datasets

Data is the foundation for all machine learning and artificial intelligence methods. To develop effective ML and DL algorithms for bearing fault detection, a good collection of datasets is necessary. Since the bearing degradation process may take many years, most people conduct experiment and collect data either using bearings with artificially injected faults, or with accelerated life testing method. While the data collection is still time consuming, fortunately a few organizations have made the effort and provided bearing fault datasets for people to work on the ML research. These datasets also serve as standards for the evaluation and comparison of different algorithms.

Before getting into the details of various ML 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. 1, 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 were introduced to

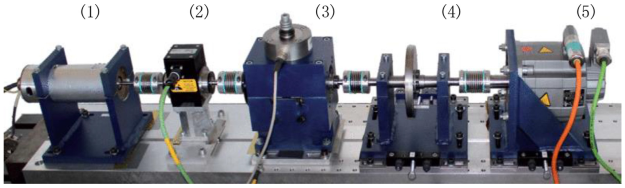


Fig. 2. Modular test rig collecting the Paderborn bearing dataset consisting of (1) an electric motor, (2) a torque-measurement shaft, (3) a rolling bearing test module, (4) a flywheel and (5) a load motor [27].

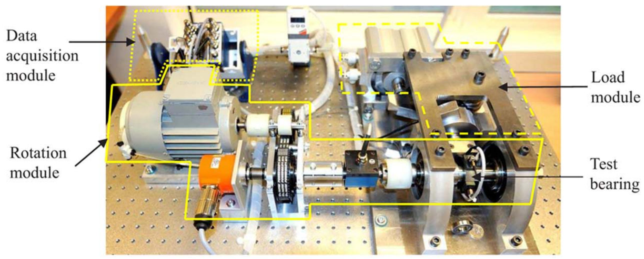


Fig. 3. PRONOSTIA testbed (Department of Automatic Control and Micro-Mechatronic Systems (AS2M), Franche-Comté Electronique Mécanique [29].

the test bearings using electro-discharge machining with fault diameters of 7 mils, 14 mils, 21 mils, 28 mils, and 40 mils, at the inner raceway, rolling element and outer raceway. Vibration data was collected for motor loads of 0 to 3 horsepower and motor speeds of 1797 to 1720 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 was 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 algorithms, and a comprehensive comparative study on all prior work employing the CWRU dataset will be presented in section V.

B. Paderborn University Dataset

The Paderborn university bearing dataset [27] includes the synchronously measured motor currents and vibration signals, thus enabling the verification of multi-physics models, as well as the cross-validation and fusion of different signals to increase bearing fault diagnostic accuracy. Both the stator current and vibration signals are measured with high resolution and sampling rate, and experiments were performed on 26 damaged bearing states and 6 undamaged (healthy) states for reference. Among the damaged bearings, 12 were artificially damaged, and the other 14 were with real damages caused by accelerated life tests. This enables more accurate testing and implementation of the ML algorithms in practical applications, where the real defects are generated through aging and the gradual lost of lubrication. The modular test rig to acquire the Paderborn bearing dataset is illustrated in Fig. 2.

C. PRONOSTIA Dataset

Another popular dataset for predicting bearing's remaining useful life (RUL) is known as "PRONOSTIA bearings accelerated life test dataset", which serves as the fundamental dataset for researchers investigating new algorithms on bearing RUL prediction. The main objective of PRONOSTIA is to provide real data related to accelerated degradation of bearings performed under constant and/or variable operating conditions, which are online controlled.

The operating conditions are characterized by two sensors: a rotating speed sensor and a force sensor. In PRONOSTIA platform as shown in Fig. 3, the bearings health monitoring is ensured by gathering two types of signals: temperature and vibration (with horizontal and vertical accelerometers). Furthermore, the data are recorded with a specific sampling frequency which allows the catching of the whole frequency spectrum of the bearing during its degradation process. Ultimately, the monitoring data provided by the sensors can be used for further processing in order to extract relevant features and continuously assess the health condition of the bearing.

During the International Conference on Prognostics and Health Management (PHM), a "IEEE PHM 2012 Prognostic Challenge" was organized. For this purpose, a web link to the degradation data [28] is provided to the competitors to allow them testing and verifying their prognostic methods. The results of each method can then be evaluated regarding its capability to accurately estimate the RUL of the tested bearings.

D. Summary

So far a majority of papers on bearing fault identification with ML algorithms employ the CWRU dataset for its simplicity and popularity. The authors anticipate a growing trend of interests will emerge on the Paderborn dataset as it contains both the stator current signal and the vibration signal. Besides, many researchers working on RUL prediction also use the PRONOSTIA dataset. Since the main scope of this paper is on bearing fault identification, research contributions on RUL prediction is not included in this literature survey.

III. Traditional Machine Learning Based Approaches

The structure of a rolling-element bearing is illustrated in Fig. 4, which contains the outer race typically mounted on the motor cap, the inner race to hold the motor shaft, the balls or rolling elements, and the cage for restraining the relative distances of the rolling elements. The four common scenarios of misalignment that are likely to cause bearing failures are demonstrated in Fig. 4(a) to (d).

Before ML actually goes deep, there are a variety of classical "shallow" machine learning and data mining algorithms that have been around for many years, i.e., the artificial neural networks (ANN) based on back-propagation. Applying these algorithms requires a lot of domain expertise and complex feature engineering, since a deep exploratory data analysis is

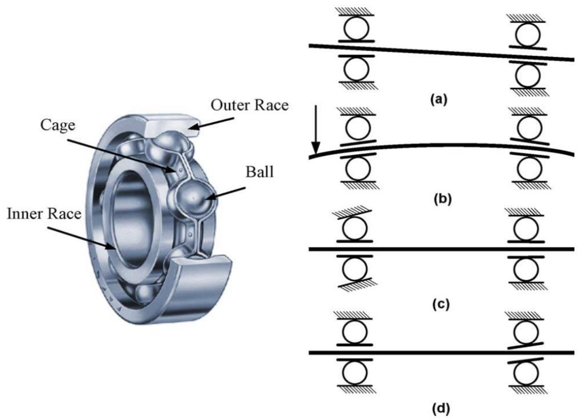


Fig. 4. Structure of a rolling-element bearing with four types of abnormal scenarios: (a) misalignment (out-of-line), (b) shaft deflection, (c) crooked or tilted outer race and (d) crooked or tilted inner race [8].

usually performed on the dataset first, then a dimension reduction might be performed with principal component analysis (PCA), etc., for easier processing. Finally, the best features must be carefully selected to pass over to the ML algorithms. The knowledge base of classical ML for different domains and applications is quite different and often requires extensive specialized expertise within each field.

One of the earliest reviews summarizing the use of artificial intelligence techniques on motor fault diagnostics can be found in [30], [31], where the characteristic fault frequencies for different motor fault types are systematically investigated, and relevant papers employing ANN and fuzzy systems are discussed. Since the main scope of this review paper lies on the deep learning based approaches, a very brief summary of each classical ML method will be presented in this section, with a complete list of references.

A. Artificial Neural Networks (ANN)

Artificial neural networks (ANNs) are one of the oldest AI paradigms that have been applied to bearing fault diagnostics for almost 30 years [37]. In [37], 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 rotor speed ω. The complexity of obtaining an accurate analytical expression for this nonlinear mapping is avoided by training a supervised neural network with accurate measurements as input and known bearing condition as output. 35 training and 70 testing data patterns have been collected over the operating curves on the laboratory test stand with Dayton 6K624B type bearing. A highest accuracy of 94.7% was recorded for the bearing fault detection with the conventional neural network using two input nodes $\{I, \omega\}$. The accuracy can be further improved by utilizing five input dimensions $\{I, \omega, I^2, \omega^2, I^*\omega\}$. Similarly, the rest of the papers based on ANN [38]–[41] all require some degree of human expertise to guide the feature selection process in order to train the ANN model more effectively.

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 (1 axis per variable), PCA can supply the user with a lower-dimensional projection of this object when viewed from its most informative viewpoint. Since the sensitivity of various features that are characteristics of a bearing defect may vary considerably under different operating conditions, PCA has proven itself as an effective and systematic feature selection scheme which provides guidance on choosing the most representative features manually for defect classification.

One of the earliest adoption of PCA to bearing fault diagnosis can be found in [42]. Experimental results revealed that the advantage in using only PCA identified features instead of 13 original features was significant, as the fault diagnosis accuracy increased from 88% to 98%. The study demonstrated that the proposed PCA technique is effective in classifying bearing faults with higher accuracy and lower number of feature inputs as compared to using all the original feature inputs. Similarly, the rest of the papers based on PCA [43]–[46] takes advantage of the its data mining capability to aid this manual feature selection process.

C. K-Nearest Neighbors (k-NN)

k-nearest neighbors algorithm (*k*-NN) is a non-parametric method used for either classification or regression. In *k*-NN classification, the output is a class membership of an object, which is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. One early implementation of the *k*-NN classifier on bearing fault diagnosis is found in [42], where *k*-NN serves as the core algorithm for a data mining based ceramic bearing fault classifier based on acoustic emission signals. Similarly, other *k*-NN based papers [48]–[50] all employ *k*-NN to perform distance analysis on each new data sample and determine whether it belongs to a specific fault class.

D. Support Vector Machines (SVM)

Support vector machines (SVM) are supervised learning models that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. One classical work on the use of SVMs towards identifying bearing faults can be found in [51], where the classification results obtained by the SVM were optimal in all the cases, and demonstrated a maximum of 20% improvement over the performance of the neural network. Other similar SVM based papers [52]–[64] mostly employ SVM to serve as the fault classifier with manually selected spectral features or statistical features.

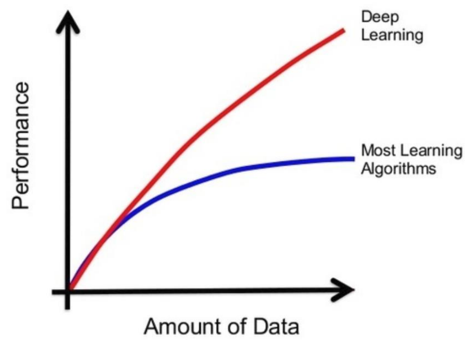


Fig. 5. Performance comparison of deep learning and most other learning algorithms [103].

E. Others

Besides the commonly used ML methods listed above, many other algorithms with different characteristics have been applied to the identification of bearing faults, including the neural fuzzy network [65]–[67], Bayesian networks [68]–[70], self-organizing maps [71], [72], extreme learning machines (ELM) [73], [74], transfer learning [76]–[78], discriminate analysis [79]–[81], random forest [82], independent component analysis [83], Softmax classifiers [84], manifold learning [85], [86], canonical variate analysis [87], particle filter [88], nonlinear preserving projection [89], artificial Hydrocarbon Networks [90], class imbalanced learning (CIL), ensemble learning (EL), multi-scale permutation entropy (MPE) [93], empirical mode decomposition [94]–[98], topic correlation analysis [99], affinity propagation [100], and dictionary learning [101], [102].

IV. Application of 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 in terms of less abstract ones. The trend of transitioning from traditional shallow ML methods to deep learning can be attributed to the following reasons.

1) *Hardware evolution*: Training deep networks is extremely computationally intensive, but running on a high performance GPU can significantly accelerate this training process. Specifically, the GPUs offer parallel computing capability and computational compatibility with deep neural network, which makes them indispensable for training DL based algorithms. More powerful GPUs allows data scientists to quickly get the deep learning training up and running. For example, the NVIDIA Tesla V100 Tensor Core GPUs [104] can now parse petabytes of data orders of magnitude faster than traditional CPUs, and leverage mixed precision to accelerate deep learning training throughputs across every type of neural network. In most recent years, the emerging of different DL computing platforms such as FPGA, ASIC, and TPU (Tensor Processing Units) as well as high performance GPUs for embedded system expedites the fast evolution of deep learning algorithms.
2) *Algorithm evolution*: More techniques and frameworks are invented and getting matured in terms of controlling the training process of deeper models to achieve faster speed, better convergence, and better generalization. For example, algorithms such as RELU that helps accelerate convergence speed; techniques that prevent overfitting such as dropout and pooling. The development of numerical optimization also provides the horsepower of leveraging more data and training deeper models, such as mini-batch Gradient Descent, RMSprop, and L-BFGS optimizer.
3) *Data explosion*: With the availability of more sensors installed that collect an increasing amount of data, and the application of crowdsourced labeling mechanism such as Amazon mTurk [105], we have seen a surging appearance of large scale dataset in many domains, such as ImageNet in image recognition, MPI Sintel Flow in image optical flow, VoxCeleb in speaker identification, et al. The performance of deep learning can significantly outperform most traditional ML algorithms, especially with the increase of dataset dimension, as illustrated in Fig. 5 [103] by Andrew Ng.

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

1) *Best-in-class performance*: The complexity of the computed function grows exponentially with depth [106]. Deep learning has best-in-class performance that significantly outperforms other solutions on problems in multiple domains, including speech, language, vision, playing games like Go, etc.
2) *Automatic feature extraction*: No need for feature engineering. Traditional machine learning algorithms usually call for sophisticated manual feature engineering which unavoidably involves expert domain knowledge and numerous human effort. However, when using a deep network, there's no need for this; as one can just pass the data directly to the network and usually achieve good performance right off the bat. This totally eliminates the challenging feature engineering stage.
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 or settings. Deep learning is an architecture that can be adapted to new problems relatively easily. For instance, problems in vision, time series, and language are using same techniques like convolutional neural networks, recurrent neural networks, long short-term memory etc.

Thanks to the above reasons for the transition from ML methods to DL methods, as well as the explicit benefits of the DL algorithms discussed above, we have witnessed an exponential increase in deep learning applications. One such example is the fault diagnostics and health prognostics, and bearing fault identification is a very representative case.

A. Convolutional Neural Network (CNN)

Inspired by animal visual cortex [107], convolution operation is first introduced to detect image patterns in a hierarchical way from simple features, such as edge and corner, to complex features. The low layers detect fundamental low level visual features such as edge and corner; and the layers afterward detect higher level features, which are built upon simple low level features.

The first paper employing CNN to identify bearing fault was published in 2016 [108] and for the next three years many papers [109]–[112], [114]–[124] emerged on this topic that contributed to its performance advancement in various aspects. The basic architecture of a CNN-based bearing fault classifier is illustrated in Fig. 6. The 1-D temporal raw data obtained from different accelerometers are firstly stacked to a 2-D matrix form, similar to the representation of images, which is then passed through a convolution layer for feature extraction, followed by a pooling layer for down-sampling. The combination of this convolution-pooling can be repeated many times to further deepen the network. Finally, the output from the hidden layers will be passed along to one or several fully-connected (FC) layers, the result of which serves as the input to a top classifier such as Softmax or Sigmoid functions.

In [108], the vibration data were acquired using two accelerometers installed on the x- and y- directions, one on top of the housing, and the other on the back of the housing. CNN is able to autonomously learn useful features for bearing fault detection from the raw data pre-processed by the scaled DFT transform. The classification results demonstrate that the feature-learning based approach significantly outperforms the feature-engineering based approach based on conventional ML methods. Moreover, the authors have shown that feature-learning based approaches such as CNN can also perform bearing health prognostics, and identify some early-stage faulty conditions without explicit characteristic frequencies, such as lubrication degradation, which cannot be identified by traditional ML methods.

An adaptive CNN (ADCNN) is applied on the CWRU dataset to dynamically change the learning rate, for a better trade-off between training speed and accuracy in [109]. The entire fault diagnosis model employs of a fault pattern determination component using 1 ADCNN and a fault size evaluation component using 3 ADCNNs, 3-layer CNNs with max pooling. Classification results on the test set demonstrate ADCNN has a better accuracy compared to conventional shallow CNN and SVM methods, especially in terms of identifying the rolling element (ball) defect. In addition, this proposed ADCNN is also able to predict the fault size (defect width) with satisfactory accuracy. On top of the conventional structure of CNN, a dislocate layer is added in [110] that can better extract the relationship between signals with different intervals in periodic forms, especially during the change of operating conditions. 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 [108]–[110], [111] 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 shallow Softmax regression classifier, especially when the vibration signal is mixed with ambient noise, this improvement can be as large as 25%, showcasing the excellent built-in denoising capabilities of the CNN algorithm itself. A sensor fusion approach is applied in [112], in which both temporal and spatial information of the CWRU raw data from the two accelerometers at both the drive end and fan end are stacked by transforming 1-D time-series data into 2-D input matrix and sent to CNN as input. 70% of the samples are used for training, 15% for validation, and 15% for testing; average accuracy with two sensors is 99.41%, while the accuracy with only one sensor is 98.35%.

Many variations of CNN are also employed to tackle the bearing fault diagnosis challenge [114]–[121] on the CWRU dataset to obtain more desirable characteristics compared to the conventional CNN. For example, a CNN based on LeNet-5 [113] is applied in [114] containing 2 alternating convolutional-pooling layers and a 2 fully-connected (FC) 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 believed to provide better feature extraction capability, as the accuracy on the test set is an astonishing 99.79%, which is better than other deep learning based methods such as the adaptive CNN (98.1%) and deep belief network (87.45%), and dominates the traditional ML methods such as SVM (87.45%) and ANN (67.70%). In addition, a deep fully CNN (DFCNN) incorporating 4 convolution-pooling layer pairs is employed in [115], 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 percentage numbers are obtained by the same authors, using the same training and test set to train the conventional ML algorithms that serve as benchmark cases. To save the extensive training time required for most CNN based algorithms, a multi-scale deep CNN (MS-DCNN) is adopted in [116], where convolution kernel of different sizes are used to extract different-scale features in parallel. The mean accuracy of 9-layer 1d-CNN, 2d-CNN and the proposed MS-DCNN are 98.57%, 98.25% and 99.27%, respectively. Despite the subtle increase in accuracy compared to conventional CNNs, the number of parameters to be determined via training is only 52,172, significantly fewer than 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

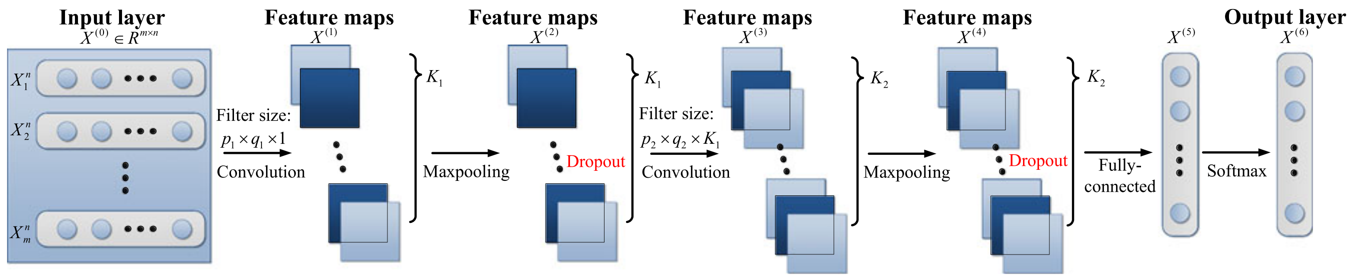


Fig. 6. Architecture of the CNN-based fault diagnosis model [114].

[117], which is able to sustain a high accuracy under noisy environment or load shifts. However, the training time and the amount of parameters would drastically increase, posting a potential threat of overfitting the data. Similarly, to overcome the impact of load variations, a novel bearing fault diagnosis algorithm based on improved DempsterShafer theory CNN (IDS-CNN) is employed in [118]. This improved D-S evidence theory is implemented via distance matrix from evidences and modified Gini Index. Extensive evaluations showed that, by fusing complementary or conflicting evidences from different models and sensors and adapting to different load conditions, the IDSCNN algorithm can achieve better fault diagnosis performance than conventional ML approaches such as SVM and DNN models.

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 [119], 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 and testing set, and the final classification accuracy is 99.63%. However, since all signals recorded by CWRU are at almost the same rotating speed (1,720 to 1,797 rpm), another experimental test is established to record vibration signals with four distinct rotating frequencies (approximately 10, 20, 30, and 40 Hz), and the average accuracy still reaches 93.19%, which is 14.38% higher than the conventional SVM algorithm. Similarly, a fault diagnosis method based on Pythagorean spatial pyramid pooling CNN (PSPP-CNN) is proposed in [120] for enhancing accuracies during motor speed variations. Compared with a spatial pyramid pooling (SPP) layer that has been used in an CNN, a PSPP layer is allocated as front layer of CNN. Thus, the features obtained by PSPP layer can be delivered to convolutional layers for further feature extraction. According to experiment results, this method has higher diagnosis accuracy for variable rotating speed bearing than other methods. In addition, the PSPP-CNN model trained by data at some rotating speeds can be used to diagnose bearing fault at full working speed.

Since CNN excels in processing 2-D matrix data such as photos in the field of computer vision, it generally requires the 1-D time domain vibration signals to be transformed into 2-D signals in the first place. Aiming at simplify this conversion process and reducing the percentage of training data required due to the complexity involved in acquiring a large amount of data, an adaptive overlapping CNN (AOCNN) employed in [121] is able to deal with the 1-D raw vibration signals directly, and eliminate the shift variant problem embedded in time domain signals. Compared to conventional CNNs, the novelty lies in the overlapping layer, which is firstly used to sample raw vibration signals. After the adaptive convolutional layer separates these samples into segments, sparse filtering (SF) is also employed to obtain local features in the local layer. Classification results reveal AOCNN with SF can diagnose ten health conditions of the bearing with 99.19% testing accuracy when only 5% samples are used in the training process, which is a significant improvement considering most DL based methods reported demand a minimum of 25% data allocated in the training set. And the testing accuracy further rises to 99.61% when the percentage increases to 20%.

Besides the CWRU dataset which only have vibration signals, the Paderborn University bearing dataset [27], as stated in section II, includes bearing dataset with both artificially induced faults and real natural faults due to aging. In [122], it is used to train a deep inception net with atrous convolution (ACDIN), which improves the accuracy from 75% (best results of conventional data-driven methods) to 95% on diagnosing the real bearing faults when trained with only the data generated from artificial bearing damages. The "PRONOSTIA bearings accelerated life test dataset" [28], as introduced in Section II, is applied in [123] 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 designed based on CNN output, and the classification results show the HI performance of CNN with automatically learned data are superior compared to that of self-organizing maps (SOM).

In addition to identifying failures on rolling element bearings, the adoption of CNN on spindle bearings is discussed in [124], in which the wavelet packet energy of the vibration signal is taken as input.

B. Auto-encoders

Auto-encoder is proposed in 1980s as a method of unsupervised pre-training method for ANN [9], [182]. Originally

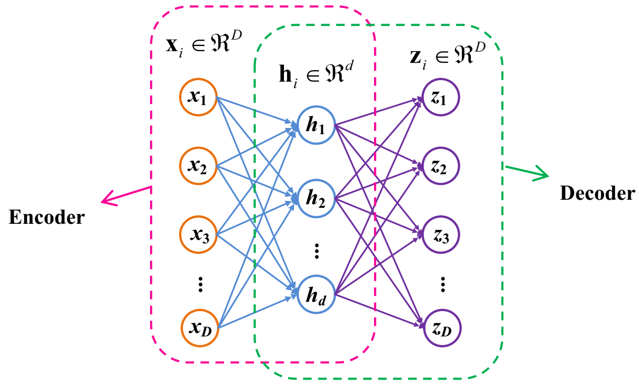


Fig. 7. Process of training a one hidden layer auto-encoder [129].

it was not called with the same terminology we are using today. After evolving for decades, 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 a one hidden layer auto-encoder is illustrated in Fig. 7. An auto-encoder is trained from an ANN, which composes of two parts: encoder and decoder. The output of encoder is fed into the decoder as input. The ANN takes the mean square error between the original input and output as loss function, which essentially aims at imitating the input as the final output. After this ANN is trained, the decoder part is dropped and only the encoder part is kept. Therefore the output of the encoder is the feature representation that can be employed in next-stage classifiers.

There are many studies employing auto-encoder in bearing fault diagnosis [125]–[138]. One of the earliest can be found in [125], where a 5-layer auto-encoder based DNN is utilized to mine fault characteristics from the frequency spectrum adaptively for various diagnosis issues, and effectively classify the health conditions of the machinery. The classification accuracy reaches 99.6%, which is significantly higher than the 70% of back-propagation based neural networks (BPNN). In [126], 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.

Compared to CNN, the denoising capabilities of original auto-encoders is not prominent. Thus in [127], a stacked denoising autoencoder (SDA) is implemented, which is suitable for deep architecture-based robust feature extraction on signals containing ambient noise and working condition fluctuations. This specific SDA consists of three auto-encoders stacked together. To balance between performance and training speeds, three hidden layers with 100, 50, and 25 units respectively are employed. The original CWRU bearing data are perturbed by a 15dB random noise to mimic the noisy condition, and multiple operating condition datasets are used as test sets to examine its fault identification capability under speed and load changes. This method achieves a worst case accuracy of 91.79%, which is 3% to 10% higher compared to conventional SAE without denoising capability, SVM and random forest (RF) algorithms. Similar to [127], another form of SDA is utilized in [128] with three hidden layers of (500, 500, 500) units, signals of the CWRU dataset are combined with different levels of artificially random noises in the time domain and later converted to frequency signals. The proposed method has better diagnosis accuracy than deep belief networks (DBN), particularly with the added noises, where an improvement of 7% in fault diagnosis accuracy is achieved.

Besides the most commonly used CWRU dataset, vibration data collected from an electrical locomotive bearing test rig, developed at the Northwestern Polytechnical University of China, is used to validate the performance of auto-encoders. Based on this dataset, the authors in [129] adopted the maximum correntropy as the loss function instead of the traditional mean square error, and one of swarm intelligence algorithms called artificial fish-swarm algorithm (AFSA) is used to optimize the key parameters in the loss function of auto-encoder. Results show that the customized 5-layer auto-encoder composed of maximum correntropy loss function and AFSA algorithm outperforms a standard auto-encoder by an accuracy of 10% to 40% on the test set 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 the enhancement of feature learning ability. A DAE is first used to learn the low-layer features from the raw vibration data, then multiple CAEs are used to learn the deep features based on the learned low-layer features. In addition, locality preserving projection (LPP) is also adopted to fuse the deep features to further improve the quality of the learned features. The classification accuracy of the proposed DAE-CAE-LPP approach is 91.90%, showcasing an advantage over standard DAE (84.60%), standard CAE (85.10%), BPNN(49.70%) and SVM (57.60%). However, all autoencoder based methods are also 6 to 10 times more time-consuming compared to conventional BPNN and SVM.

In addition. an aircraft-engine inter-shaft bearing vibration dataset with inner race, outer race and rolling element defects is adopted as the input data in [131], where a new AE based on Gaussian radial basis kernel function is employed to enhance the feature learning capability. Later, a stacked AE was developed with this new AE and multiple conventional AEs with an accuracy of 86.75%, which is much better compared to standard SAE (44.90%) and standard DBN (19.65%). Moreover, the importance of the proposed Gaussian radial basis kernel function was presented: if the kernel function is changed to either a polynomial kernel function (PK) or 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 to tackle the bearing fault diagnosis challenge [132]–[138] on the CWRU dataset to achieve better performance compared to the traditional SAE. An ensemble deep auto-encoders (EDAE) is employed in [132] with a series of auto-encoders (AE) with different activation functions for unsupervised feature learning from the measured vibration signals. Later a combination strategy is designed to ensure accurate and stable diagnosis results. The classification accuracy is 99.15%, which shows performance boosts compared to BPNN (88.22%), SVM (90.81%), and RF (92.07%), after performing a 24-dimension manual feature extraction process. Similarly, by altering the activation function, a deep wavelet auto-encoder (DWAE) with extreme learning machine (ELM) is implemented in [133], where the wavelet function is employed as the nonlinear activation function to design wavelet auto-encoders (WAE) to effectively capture the 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. The input data dimension is, and the output accuracy achieves 95.20%. According to [133], this method not only outperforms conventional ML methods such as BPNN (85.43%) and SVM (87.97%), but also some standard deep learning algorithms, including the standard DAE + soft-max (89.70%) and standard DAE + ELM (89.93%).

Considering the relatively large dataset required to train deep neural nets, a 4-layer DNN with stacked sparse autoencoder (SSAE) is established in [134] 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 second hidden layer, and 7 nodes in the output layer (dependent on the number of fault conditions in a dataset). A nonlinear projection is performed to compress the vibration data and automatic adaptive feature extraction in the transform domain; accuracy can reach 97.47%, which is 8% better than SVM, 60% better than a three-layer ANN and 46% better than multi-layer ANNs. In [135], two limitations of conventional SAE are summarized. The first one is that SAE tends to extract similar/redundant features that increases the complexity rather than 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, local layer, feature layer and output layer. Specifically, this method first uses SAE to locally learn features from input signals in the local layer, then obtains shift-invariant features in the feature layer, and finally recognizes mechanical health conditions in the output layer for a 10-classes problem. The average accuracy is 99.92%, which is 1% to 5% higher than many EMD, ensemble NN, and DL based methods. Similarly, the fault data are utilized to automatically extract the representative features by means of SAE in [136], while the diagnosis model is constructed by using ISVM. This model has been also 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 [137], which induces a fuzzy maximum likelihood estimation (FMLE) of the distance norm to determine the degree of a sample belonging to each cluster. While an 8-layer SDAE is still used to extract the useful features and reduces the dimension of the vibration signal, GathGeva is deployed to identify the different fault types. The worst case classification accuracy is 93.3%, outperforming the traditional EMD based feature extraction schemes by almost 10%.

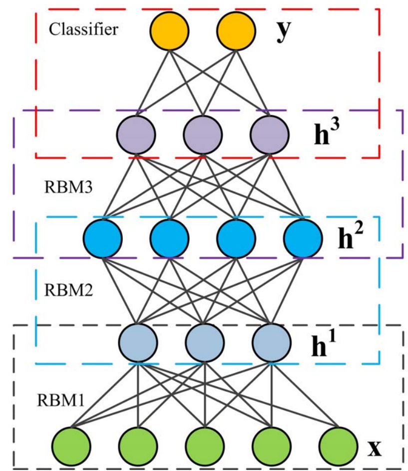


Fig. 8. Architecture of DBN [140].

To reduce DL based model complexity, another bearing fault diagnosis method based on fully-connected winner-take-all autoencoder is proposed in [138], in which the model explicitly imposes lifetime sparsity on the encoded features by keeping only k% largest activations of each neuron across all samples in a mini-batch. A soft voting method is implemented to aggregate prediction results of signal segments sliced by a sliding window to increase accuracy and stability. A simulated dataset is generated by adding white Gaussian noise to original signals of the CWRU dataset to test the diagnosis performance under noisy environment. The experimental results demonstrate that with a simple two-layer network, the proposed method is not only capable of diagnosing with high precision under normal conditions, but also has better robustness to noise than some deeper and more complex models, such as CNN.

C. Deep Belief Network (DBN)

In deep learning, a Deep Belief Network (DBN) can be viewed as a composition of simple, unsupervised networks such as Restricted Boltzmann Machines (RBMs) or autoencoders, where each sub-network's hidden layer serves as the visible layer for the next, as illustrated by the boxes of different colors in Fig. 8. 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 (the lowest visible layer in a training set).

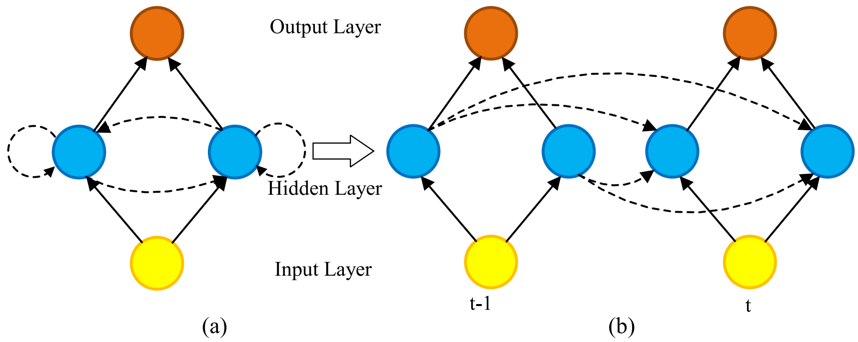


Fig. 9. (a) Architecture of RNN, and (b) architecture of RNN across a time step [148].

The observation that DBNs can be trained greedily, one layer at a time, led to one of the first effective deep learning algorithms [139]. There are many attractive implementations and uses of DBNs in real-life applications such as drug discovery; and its first application on bearing fault diagnosis was published in 2017 [140].

In [140], a multi-sensor vibration data fusion technique is implemented to fuse the time domain and frequency domain features extracted via multiple 2-layer SAEs. Then a 3-layer RBM based DBN is used for classification. Validation is performed on vibration data under different speeds, and a 97.82% accuracy demonstrated that the proposed method can effectively identify bearing faults even after a change of operating conditions. Further feature visualization using t-SNE reveals that this multi-SAE based feature fusion outperforms other cases with only one SAE or with no fusion. In [141], a stochastic convolutional DBN (SCDBN) is implemented by means of stochastic kernels and averaging processing, and unsupervised CNN is built to extract 47 features. Later a 2-layer DBN is implemented with (28, 14) nodes, 5 kernels in each layer, 1 pooling layer without overlapping. Finally, a Softmax layer is used for classification that achieves an average accuracy of over 95%.

Many DBN papers also employ the CWRU bearing dataset as the input data [142]–[144] thanks to its popularity. For example, an adaptive DBN and dual-tree complex wavelet packet (DTCWPT) is proposed in [142]. The DTCWPT first prepossesses the vibration signals, where an original feature set with 9×8 feature parameters is generated. The decomposition level is 3, and the db5 function, which defines the scaling coefficients of the Daubechies wavelet, is taken as the basis function. Then a 5-layer adaptive DBN of (72, 400, 250, 100, 16) structure is used for bearing fault classification. The average accuracy is 94.38%, which is much better compared to convention ANN (63.13%), GRNN (69.38%), and SVM (66.88%) using the same training and test data. In [144], data from two accelometers mounted on the load end and fan end are processed by multiple DBNs for feature extraction; then the faulty conditions based on each extracted feature are determined with Softmax; and the final health condition is fused by DS evidence theory. An accuracy of 98.8% is accomplished while including the load change from 1 hp to 2 and 3 hp. In contrast, the accuracy of SAE suffers the most from this load change, and the accuracy employing CNN is also lower than DBN. Similar to this D-S theory based output fusion [143], a 4-layer DBN of (400, 200, 100, 10) structure with different hyper-parameters coupled with ensemble learning is implemented in [144]. An improved ensemble method is used to acquire the weight matrix for each DBN, and the final diagnosis result is formulated by each DBN based on their weights. The average accuracy of 96.95% is better compared to the accuracies of employing 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, DBN has also been applied to many other datasets. A convolutional DBN, constructed with convolutional RBMs, is employed in [145] on locomotive bearing vibration data, where an auto-encoder is firstly used to compress data and reduce the dimension. Without any feature extractions, the compressed data are divided into training samples and testing 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 is much better compared to other DL methods using the same classifier and raw data, but with inferior feature extraction capabilities, such as the denoising autoencoder (90.76%), standard DBN (88.10%) and CNN (91.24%). In [146], a 5-hidden-layer DBN with (512, 2048, 1024, 2048, 512) nodes is employed on bearing data directly obtained from power plants, with vibration image generation incorporating data from systems with various scales, such as small testbeds and real field deployed systems. Unsupervised feature extraction is performed by DBN, and the fault classifier is designed using SOM. The resultant clustering accuracy is 97.13%, which would decrease with fewer hidden layers or fewer nodes.

DBN has also been applied to bearing RUL prediction. In [147], a DBN-feedforward neural network (FNN) is applied to perform automatic self-taught feature learning with DBN and

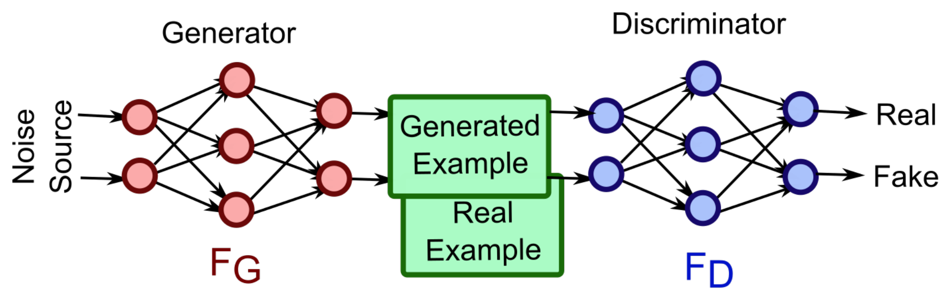


Fig. 10. Architecture of GAN [153].

RUL prediction with FNN. Two accelerometers were mounted on the bearing housing, in directions perpendicular to the shaft, and data is collected every 5 min, with a 102.4 kHz sampling frequency, and a duration of 2s. Experimental results demonstrate the proposed DBN based approach can accurately predict the true RUL 5 min and 50 min into the future.

D. Recurrent Neural Network (RNN)

Different from feed-forward neural network (FNN), RNN processes input data in a recurrent behavior. The architecture is shown in Fig. 9. With a flow path going from its hidden layer to itself, when unrolled in the sequence, it can be viewed as a feed-forward neural network across the input sequence. As a sequential model, it can capture and model sequential relationship in sequential data or time series data. However, trained with back propagation through time (BPTT), RNN has the notorious gradient vanishing/exploding issue stemmed from its nature. Although proposed as early as 1980s, RNNs have limited applications for 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 astonishing capability in memorizing and modeling the long-term dependency in data, and therefore is taking a dominant role in time series and text 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 application of RNN on bearing fault diagnostics can be found in in 2015 [148], where the fault features were first extracted using discrete wavelet transforms and later selected based on orthogonal fuzzy neighbourhood discriminative analysis (OFNDA). These features were then fed into an RNN for fault classification, enabling the fault classifier to incorporate a dynamic component. The experiment has shown the proposed scheme based on RNN is capable of detecting and classifying bearing faults accurately, even under non-stationary operating conditions. Another RNN based Health Indicator (RNN-HI) is proposed in [149] to predict the RUL of bearings with LSTM cells used in RNN layers. Combined with time-frequency features, the related-similarity (RS) features, which calculates the similarity between the currently inspected data and data at an initial operation point, are extracted. Then after correlation and monotonicity-metrics-based feature selection, the selected features are transferred to an RNN network to predict the bearing HI, from which RUL is estimated. With Experiment on generator bearings from wind turbines, the proposed RNN-HI is demonstrated to give better performance than an SOM based method.

In addition, a methodology of a combined 1-D CNN and LSTM to classify the bearing fault types is presented in [150], 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 sampling signal without any pre-processing, and the best testing accuracy of different configurations reaches as high as 99.6%. A more recent work employing a deep recurrent neural network (DRNN) is proposed in [151] with stacked recurrent hidden layers and LSTM units. Mean Square Error (MSE) function is used as the loss function and stochatic gradient descent (SGD) is used as the optimizer. Besides, an adaptive learning rate is also adopted to improve the training performance. The average testing accuracy of the proposed method is 94.75% and 96.53% at 1750 and 1797 rpm, respectively.

E. Generative Adversarial Network (GAN)

Generative Adversarial Network (GAN) was proposed by Goodfellow et al. [152] in 2014 and rapidly became one of the most exciting breakthroughs in the field of machine learning. A GAN composes two parts: generator (F_G) and discriminator (F_D), as illustrated in Fig. 10. GAN artfully takes the concept of competition. Generator (F_G) and Discriminator (F_D) are competing with each other, in the manner that F_G is trying to confuse the discriminator and F_D is trying to distinguish the samples generated by F_G and the samples from the original dataset. GAN is set up as a zero-sum game framework, where both F_G and F_D are competing to individually gain stronger and stronger capability of imitating original data samples and discrimination capability iteratively.

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 [154], which aims at addressing the class imbalance issues using GAN. The authors compared its performance with that of standard oversampling

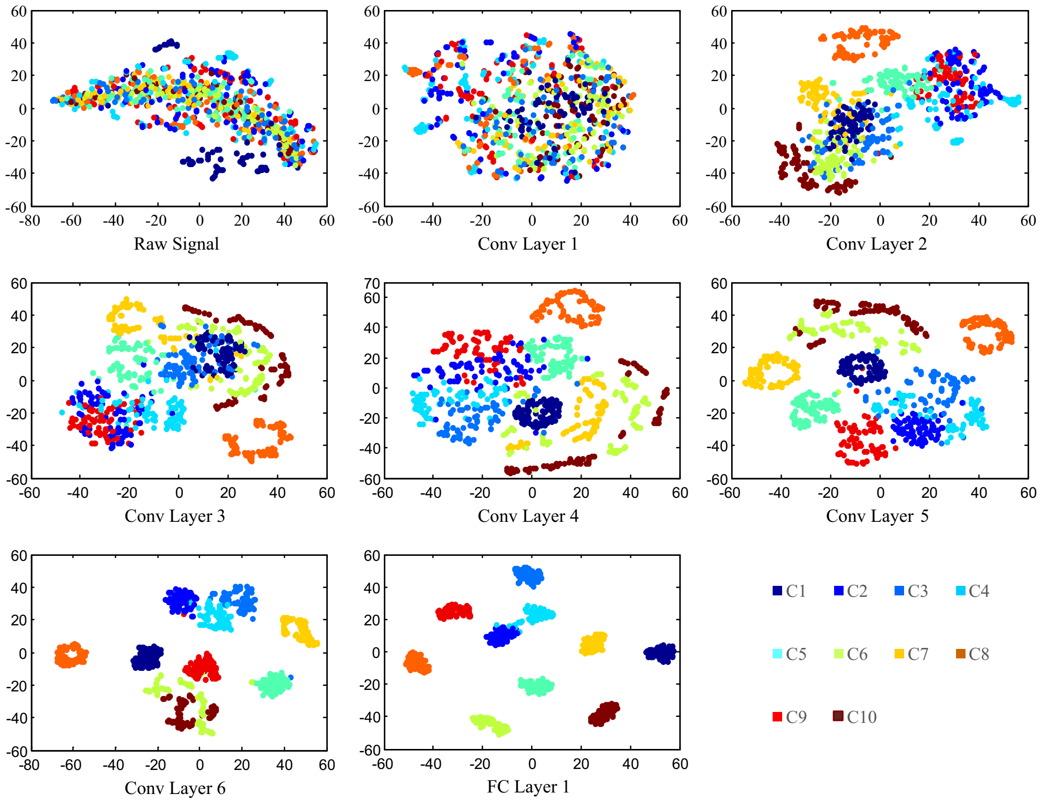


Fig. 11. Feature visualization via t-SNE: feature representations for all test signals extracted from raw signal, six convolutional layers and the last fully connected layer respectively [119].

techniques, and showed the advantage of adopting GAN. In [154], GAN is combined with Adaptive Synthetic sampling (ADASYN) approach to resolve the inability of the GAN generator to create meaningful data oversampling when the original sample data are scarce.

In [155], the authors proposed a new GAN framework called categorical adversarial autoencoder (CatAAE), which automatically trains an autoencoder through an adversarial training process, and imposes a prior distribution on the latent coding space. Then a classifier tries to cluster the input examples by balancing mutual information between examples and their predicted categorical class distribution. The latent coding space and training process are presented to investigate the advantage of proposed model. The experiments under different signal-to-noise ratios (SNRs) and fluctuating motor loads have indicated the preponderance in learning useful characteristics of the proposed CatAAE.

For real-world applications, due to the fact that the operating condition varies, they usually cannot comply with the common assumption that training set and test set have the same distribution. Similar to [155] and inspired by GAN, a new adversarial adaptive 1-D CNN model (A2CNN) is proposed in [156] to address this problem. Experiments show that A2CNN has strong fault-discriminative and domain invariant capacity, and therefore can achieve high accuracy under different operating conditions. In [157], A novel approach for fault diagnosis based on deep convolution GAN (DCGAN) with imbalanced dataset is proposed. A new DCGAN model [159] with 4 convolutional layers for the discriminator and generator individually is designed and applied on raw and imbalanced vibration signals. Following data balancing using DCGAN model, statistical features based on time-domain and frequency-domain data are extracted to train a SVM classifier for bearing fault classification. The training accuracy and test accuracy of the proposed DCGAN method show better performance than other class balancing methods, including random over-sample, random under-sample, and synthetic minority over-sampling technique (SMOTE).

F. Deep Learning based Transfer Learning

The success of machine learning and deep learning based bearing fault diagnostics relies on massive heavily annotated data. However, this is generally not feasible in most real world applications due to 1) the unafforable and serious consequences when machines running under fault conditions, and 2) the potential time-consuming degradation process before desired failure happens. With publicly available datasets or self-collected datasets sampled from non-target machines, the result will naturally deteriorate on target machines. Even on the identical machine under different loads or settings, since there inevitably exists distribution discrepancy between the features from the training set and the test set, the performance still suffers. Designed to tackle this practical and widely existent issue in numerous areas, transfer learning has aroused extensive attention in machine learning community, and various transfer learning frameworks are proposed based on classical machine learning algorithms [75]. A popular method among all types of transfer learning approaches is domain adaptation. By exploring domain-invariant features, domain adaptation establishes knowledge transfer from source domain to target domain [173]. Therefore, with labeled data from source domain and unlabeled data from target domain, the distribution discrepancy between the two domains can be mitigated by domain adaptation algorithms. In recent years, an integration of deep learning and transfer learning approaches has been witnessed. Specially designed domain adaptation modules are combined with deep learning architectures to endow domain transfer ability while maintaining the extraordinary automatic feature learning ability [173]–[177].

[176] proposed a novel framework WDCNN (Deep CNN with wide first-layer kernels) combined with Adaptive Batch Normalization (AdaBN). It takes raw vibration signal as input. It's based on a CNN architecture with wide kernels (64) in the first convolutional layer for better suppressing high frequency noise. Then domain adaptation is implemented by extracting the mean and variance of target domain signals and passing to AdaBN. It used CWRU dataset and conducted cross-domain experiments when training in one working condition and testing in another one. WDCNN has achieved the average accuracy of 90.0% outperforming FFT-DNN method 78.1% and reveived further improvement to 95.9% with AdaBN. When tested under noise environment (with additive white Gaussian noise), WDCNN with AdaBN achieves 92.65% accuracy under -4(dB) SNR, in comparison with 66.95% without AdaBN.

In [174], a domain adaptation module facilitates a one-dimensional CNN to learn domain-invariant features by maximizing domain recognition errors and minimizing probability distribution distance. To validate the efficacy of the domain adaptation ability, 3 dataset including CWRU dataset, IMS dataset and railway locomotive (RL) bearing dataset are employed. By training on one of the three datasets and testing on another one, the achieved average accuracy 86.3% has surpassed the traditional CNN 53.1% and two another existing domain adaptation frameworks 75.6% [171] and 78.8% [172].

[177] proposed a novel framework combining domain adaptation with deep generative network. It employs a 2 stage structure. In the first stage, A 8-layer CNN component including 3 convolutional layers with basic classifier is trained as feature extractor while optimizing the classification error under source supervision; Then $N_c - 1$ (N_c is the number of classes) CNN components, each of which composes 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 and outputs the final diagnosis results.

G. Others

There are also numerous other DL methods employed for bearing fault diagnostics, some are based on new algorithms, some are mixture of the DL methods listed above. For example, in [160], a new large memory storage retrieval (LAMSTAR) neural network is proposed with one input layer, 40 input SOM modules as hidden layers, and one decision SOM module as output layer. More accurate classification results compared to conventional CNN are reported on various operating conditions, especially at low speeds. In [161], both DBN and SAE are applied simultaneously to identify the presence of a bearing fault. Other examples include a mixture of CNN and DBN [162], deep residual network (DRN) [163], [164], deep stack network (DSN) [165], deep generative network [78], RNN based autoencoder [166], sparse filtering [170], a new deep learning architecture capsule network [168] proposed by Hinton [167].

V. Discussions on Different Machine Learning Algorithms for Bearing Fault Diagnosis

A. Feature Extraction and Selection

To detect if a bearing fault is present, for a traditional ML algorithm, the characteristic fault frequencies are calculated based on the rotor mechanical speed and the specific bearing geometry, which serve as the fault features. This feature calculation process is known as "feature engineering". The amplitude at these frequencies can then be monitored to train various ML algorithms and identify any anomalies. However, such a technique may encounter many challenges that can affect the classification accuracy.

1) *Sliding:* The frequency calculations have the assumption that there is no sliding, i.e., the rolling elements only roll on the raceways. Nevertheless, this is seldom the case. Often, a bearing undergoes a combination of rolling and sliding. As a consequence, the calculated frequencies may differ slightly, i.e. 1-2 percent, compared to the actual frequencies.
2) *Frequency interplay:* If multiple types of bearing faults occur simultaneously, these faults will interact and the characteristic frequencies generated can add and subtract due to superposition effects, obfuscating important frequencies.
3) *External vibration:* There is also the possibility of interference induced due to additional sources of vibration, i.e. bearing looseness, which obscures useful features.
4) *Observability:* Some faults, such as lubrication related faults, do not even manifest themselves as a characteristic cyclic frequency, which makes them very hard to detect via traditional vibration spectral analysis techniques.
5) *Sensitivity:* The sensitivity of various features that are characteristic of a bearing defect may vary considerably under different operating conditions. A very thorough and systematic "learning stage" is required to test the sensitivity of these frequencies on any desirable operating condition before it can be actually put into use.

TABLE I
A SUMMARY OF DIFFERENT DEEP LEARNING ARCHITECTURE.

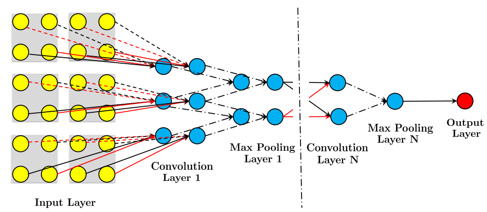
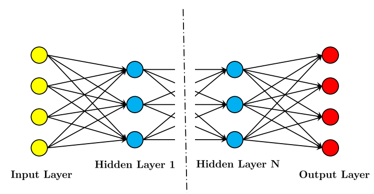
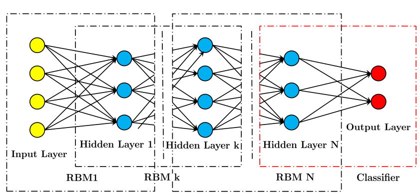
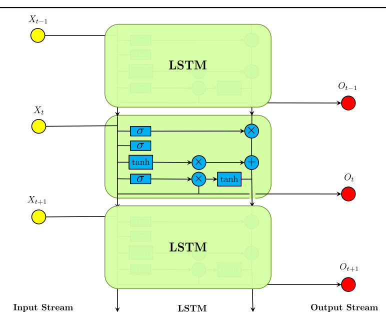
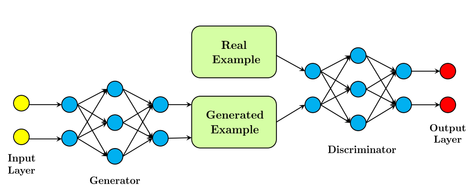
Architecture	Description	Characteristics
	**Convolutional Neural Network** • Well-suited for 2D data, i.e., images. • Inspired by the neural-biological model of the visual cortex [107]. • RELU after convolutional layers helps accelerate convergence speed.	**Pros:** • Few neuron connections required with respect to a typical ANN. • Many variants have been proposed: LeNet-5 [112], LiftingNet [119], and inception net [122], etc. **Cons:** • May require many layers to find an entire hierarchy. • May require a large dataset of labeled 2D data.
	**Deep Autoencoder** • Mainly designed for feature extraction or dimension reduction. • Has the same number of input and output nodes. • Unsupervised learning method aiming to recreate the input vector.	**Pros:** • Does not require labeled data. • Many variations have been proposed to make the representation more noise-resilient and robust: stacked denoising AE [127], ensemble deep AE [132], and stacked sparse AE [134]. **Cons:** • Requires a pre-training stage. • Training may suffer from the vanishing of errors.
	**Deep Belief Network** • Composed of RBMs where each subnetwork'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 minimizes the likelihood directly. **Cons:** • Training may be computationally expensive due to the initialization process and sampling.
	**Recurrent Neural Network** • An ANN capable of analyzing data streams. • LSTM re-vibrated the application of RNNs. • Suitable for applications where the output depends on the previous computations.	**Pros:** • Memorizes sequential events. • Models time dependencies. • Has shown great success in many applications: speech recognition, natural language processing, video analysis. **Cons:** • Learning issues are frequent due to gradient vanishing/exploding.
	**Generative Adversarial Network** • Mainly designed for generate photographs that look superficially authentic to human observers. • Implemented by a system of generative and discriminative networks. • Semi-Supervised learning method aiming to recreate the input vector.	**Pros:** • Requires almost no modifications 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 a Nash equilibrium of a game. • Hard to learn to generate discrete data, like text.

TABLE II
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	Testing accuracy
[109]	Adaptive CNN (ADCNN)	3	Softmax	Predict fault size	50%	97.90%
[111]	CNN	4	Softmax	Noise-resilient	90%	92.60%
[112]	CNN	4	Softmax	Sensor fusion	70%	99.40%
[114]	CNN based on LeNet-5	8	FC layer	Better feature extraction	83%	99.79%
[115]	Deep fully CNN (DFCNN)	8	Connectionist temporal classification	Validation with actual service data	78%	99.22%
[116]	Multi-scale deep CNN (MS-DCNN)	9	Softmax	Reduce training time	90%	98.57%
[117]	CNN with training interface	12	Softmax	Adapt to load change	96%	95.50%
[118]	IDS-CNN	3	Softmax	Adapt to load change	80%	98.92%
[119]	CNN-based LiftingNet	6	FC layer	Adapt to speed change	50%	99.63%
[120]	PSPP-CNN	9	Softmax	Adapt to speed change	5%	99.19%
[121]	AOCNN with SF	4	Softmax	Reduce training set %	5%	99.19%
[125]	SAE	3	ELM	Adapt to load change	50%	99.61%
[126]	SAE	3	ELM	Reduce training time	50%	99.83%
[127]	Stacked denoising AE (SDAE)	3	N/A	Noise-resilient	50%	91.79%
[128]	SDAE	3	Softmax	Noise-resilient	80%	99.83%
[132]	Ensemble deep AE (EDAE)	3	Softmax	Better feature extraction	67%	99.15%
[133]	Deep wavelet AE (DWAE)	3	ELM	Reduce training time	67%	95.20%
[134]	Stack sparse AE	2	N/A	Data compression	N/A	97.47%
[135]	SAE-LCN (local connection network)	2	Softmax	Shift-invariant features	25%	99.92%
[136]	SAE	3	SVM	Online diagnosis	N/A	95.10%
[137]	SDAE	8	Gath-Geva (GG)	Noise-resilient	N/A	93.30%
[138]	Winner-take-all AE	2	Gath-Geva (GG)	Noise-resilient	N/A	97.27%
[142]	dual-tree complex wavelet	5	N/A	Adaptive DBN	67%	94.38%
[143]	DBN	2	Softmax	Adapt to load change	N/A	98.80%
[144]	DBN with ensemble learning	4	Sigmoid	Accurate & robust	N/A	96.95%
[150]	CNN-LSTM	3	Softmax	Accurate	83%	99.6%

Because of these various challenges, manually engineered features based on vibration signals can be difficult to interpret for people other than by an experienced vibration analysts, and sometimes it may even lead to inaccurate classification results, especially under the influence of external disturbances,

As opposed to feature engineering, which manually selects features that preserve the discriminative characteristics of the data, the deep learning based algorithms can learn the discriminative feature representation directly from input data in an end-to-end manner. The latter approach does not require human expertise or prior knowledge of the problem, and is advantageous in bearing fault diagnosis, where it is sometimes challenging to accurately determine the fault characteristic features. 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 convolution layers in Fig. 11. A glimpse of the clustering effect can be observed in convolution layer C2; and it become increasingly apparent in later convolution layers. For comparison reasons, many deep learning 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.

In summary, the features relevant to bearing faults can be hand-picked based on the conventional characteristic fault frequencies in the frequency domain, which are prone to many challenges listed above that can impact its accuracy, or via the deep learning based algorithms themselves to perform simultaneous automatic feature extraction and classification, which appears to be a better option in the future of bearing fault diagnostics.

B. Comparison of Different DL Algorithms for Bearing Fault Diagnostics

Thus far, several DNNs architectures have been introduced in literature and TABLE I briefly describes the pros and cons of the commonly used deep learning approaches in the field of bearing fault diagnostics.

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 is presented in TABLE II. As can be readily observed, the minimum number of hidden layers is 2, indicating that the complete network has at least 4 layers incorporating the input and output layers. The maximum hidden layer size can be as large as 12 in [119], representing a very deep network that requires more time for the training process.

The testing accuracies of all 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 the specific values of testing accuracy are only demonstrated for an intuitive understanding for the following reasons:

1) *Generalization:* Some of the DL methods with astonishing accuracies over 99% are generally applied on a very specific dataset at a fixed operating conedition, i.e., when the motor speed is 1797 rpm and load is 2 hp. However, this accuracy would suffer, and may drop to below 90% under the influence of noise and variation of motor speed/load, which is common in practical applications. This is in spite of the relatively strong robustness to noise disturbance of 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 [111] that CNN has a better built-in denoising mechanism compared to other common DL algorithms such as SAE. Due to this limitation, some papers applied stacked denoising AE (SDAE) [127], [128], [137] that can effectively increase accuracy for bearing fault identification with a small signal to noise ratio (SNR), i.e., SNR = 5 or 10.
2) *Evaluation metrics:* 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 [156]. Compared with accuracy, other metrics, such as *precision* and *recall*, should be introduced to provide more details for evaluating the reliability of a fault identification model.
3) *Randomness:* Even these DL methods are using the same dataset to perform classifications, the percentage of training data and test data are different. 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 randomness in the training and test dataset selection.
4) *Accuracy saturation:* Most existing DL algorithms can achieve an excellent classification accuracy of over 95% while using the CWRU dataset, which seems to indicate that this dataset contains relatively clear features that can be readily extracted by a variety of DL methods. The superiority of certain DL methods, however, is not apparent. Thus various perturbations on the original dataset needs to be performed, i.e., pollute the signals with random noise to test the algorithm's denosing capability. In addition, the level of background noise and all bearing defects are artificially induced, which makes the validation results of different DL algorithms less convincing.

VI. Recommendations, Suggestions, 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/DL methods to solve their bearing problems at hand, the authors suggest the following sequences to make the best decision.

1) *Examine the setup environment:* Thoroughly examine the working environment and all possible operating conditions of the setup. For example, indoor or outdoor, always maintaining a fixed operating point or experiencing frequent change of speed and load. For the simplest case with an indoor and a single operating point setup, the traditional ML methods or even the frequency-based analytical models should suffice. For applications that are more prone to external disturbances or constant change of operating points, such as the motors fed by VFD converters in electric vehicles, more advanced deep learning approaches should be considered, and certain denoising blocks and extra layers should be considered to increase the deep neural net robustness. Specifically, when the workbench is exposed to a noisy environment, which induces a relatively inferior signal-to-noise ratio, special denoising components should be designed to increase the robustness of the architecture to external disturbances.
2) *Sensors:* Determine the number and type of sensors to be mounted close to the bearings. For the traditional frequency based and ML based 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 the 2-D image data, multiple 1-D time-series data obtained by the sensors in the bearing setup need to be stacked to form this 2-D data, or some prepossessing functions such as wavelet packet decomposition (WPD) need to be applied before the data is passed on to the deep neural nets. 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 AE and stator current can be installed to form a multi-physics dataset to further improve the classifier performance, especially during frequent and abrupt changes of the operating conditions.
3) *Data Size:* If the size of the collected dataset is not sufficient to train a deep learning model with good generalization, the algorithms and their training processes should be selected that make the most of the data and computing resources available. For example, the dataset augmentation techniques such as GAN, and data random sampling with replacement such as Boostrapping can be implemented. Before actually collecting data from the bearing system setup, it is also useful to consider

beforehand how precise the classification results need to be, in order to allow a useful interpretation of the needed sample size. [183] With the problem of small labeled dataset, another promising routine is to leverage unlabeled dataset, if possible, and apply the so-called semi-supervised learning paradigm by combining supervised and unsupervised learning approaches.

B. Future Work Directions

As a future research trend, the authors suggest the following directions that are crucial to the advancement of the topic and the smooth transition from the research labs to real-world applications:

1) *More complicated datasets:* Apply more complicated bearing datasets such as the Paderborn University dataset, since the accuracy saturation problem of the commonly used CWRU dataset prevails.
2) *From artificial to the real-world:* Make attempts to predict those natural faults with the algorithms trained by artificial faults, which is actually a reasonable expectation for the DL-based fault indicators to be applied in real-world applications.

VII. Conclusions

In this paper, a systematic review is conducted on existing literature employing machine learning algorithms to motor bearing fault diagnostics. Special emphasis is placed on the deep learning based approaches that has spurred the interests of academia for the past five years. It is demonstrated that, despite the fact that the 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 suitable for real-time bearing fault diagnostics.

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

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] W. Xin, "A hybrid intelligent technique for induction motor condition monitoring," Ph.D. dissertation, Dept. Mech. Des. Eng., Univ. Portsmouth, Portsmouth, U.K., 2011.

[20] S. Singh, C. Q. Howard, and C. Hansen, "An extensive review of vibration modelling of rolling element bearings with localised and extended defects," *J. Sound Vib.*, vol. 357, pp. 300–330, 2015

[21] H. Cao, L. Niu, S. Xia, and X. Chen, "Mechanical model development of rolling bearing-rotor systems: A review," *Mech. Syst. Signal Process.*, vol. 102, pp. 37–58, Mar. 2018.

[22] P. D. McFadden and J. D. Smith, "Model for the vibration produced by a single point defect in rolling element bearing," *J. Sound Vib.*, vol. 96, no. 1, pp. 69–82, 1984.

[23] N. Tandon and A. Choudhury, "An analytical model for the prediction of the vibration response of rolling element bearings due to a localized defect," *J. Sound Vib.*, vol. 205, no. 3, pp. 275–292, Aug. 1997.

[24] Z. Kıral and H. Karagulle, "Vibration analysis of rolling element bearings with various defects under the action of an unbalanced force," *Mech. Syst. Signal Process.*, vol. 20, pp. 1967–1991, 2006.

[25] S. Singh *et al*, "Analyses of contact forces and vibration response for a defective rolling element bearing using an explicit dynamics finite element model," *J. Sound Vib.*, vol. 333, pp. 5356–5377, 2014.

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

[27] (2018, December). NVIDIA TESLA: The Worlds Most Advanced Data Center GPUs., [Online]. Available: https://mb.uni-paderborn.de/kat/forschung/datacenter/bearing-datacenter/.

[28] (2018, December). IEEE PHM 2012 Data Challenge bearing dataset, [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] 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.

[31] 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.

[32] 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.

[33] 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.
[34] 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.
[35] 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.
[36] 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.
[37] 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.
[38] 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.
[39] 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.
[40] 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.
[41] 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.
[42] 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.
[43] 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.
[44] 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.
[45] 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.
[46] 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.
[47] 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.
[48] 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.
[49] 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.
[50] 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.
[51] 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.
[52] 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.
[53] 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.
[54] 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.
[55] 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.
[56] 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.
[57] 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.
[58] 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.
[59] 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.
[60] 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.
[61] 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.
[62] 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.
[63] 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.
[64] 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.
[65] 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.
[66] 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.
[67] 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.
[68] 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.
[69] 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.
[70] 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.
[71] 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.
[72] Z. Zhou, J. Chen, Y. Zi and X. Chen, "Multi-domain description method for bearing fault recognition in varying speed condition," in *Proc. 42nd Annu. Conf. IEEE Ind. Electron. Soc. (IECON'16)*, Florence, Italy, 2016, pp. 423–428.
[73] 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.

[74] 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.

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

[76] 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.

[77] 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.

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

[79] 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.

[80] 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.

[81] 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.

[82] 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.

[83] 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.

[84] 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.

[85] 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.

[86] 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.

[87] 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.

[88] 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.

[89] 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.

[90] 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.

[91] 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.

[92] 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)*, Harbin, China, 2017, pp. 1–6.

[93] 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.

[94] 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.

[95] 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.

[96] 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.

[97] 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.

[98] 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.

[99] 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.

[100] 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.

[101] 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.

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

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

[104] (2018, December). NVIDIA TESLA: The Worlds Most Advanced Data Center GPUs., [Online]. Available: http://wild.link/nvidia/ANrTCA.

[105] (2018, December). Amazon Mechanical Turk, [Online]. Available: https://www.mturk.com/.

[106] 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.

[107] 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.

[108] 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.

[109] 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.

[110] 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.

[111] 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.

[112] 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.

[113] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, Gradient based learning applied to document recognition, *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[114] 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.

[115] 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.

[116] 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.

[117] 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.

[118] 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.

[119] 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.

[120] 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.

[121] 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.

[122] 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.

[123] 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.

[124] 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.

[125] 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.

[126] 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.

[127] 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.

[128] 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.

[129] 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.

[130] H. Shao, H. Jiang, F. Wang, and H. Zhao, "An enhancement deep feature fusion method for rotating machinery fault diagnosis," *Mech. Syst. Signal Process.*, vol. 102, pp. 278–297, 2017.

[131] 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.

[132] 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.

[133] 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.

[134] 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.

[135] 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.

[136] 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.

[137] 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.

[138] 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.

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

[140] 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.

[141] 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.

[142] 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.

[143] 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.

[144] 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.

[145] 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.

[146] 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.

[147] 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.

[148] 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.

[149] 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.

[150] 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.

[151] 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.

[152] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014*, December 8-13 2014, Montreal, Quebec, Canada, 2014, pp. 26722680.

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

[154] 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.

[155] 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.

[156] 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.

[157] 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.

[158] S. Khan, A. Prosvirin, and J. Kim, "Towards bearing health prognosis using generative adversarial networks: Modeling bearing degradation," in *Proc. Int. Conf. Adv. Comput. Sci. (ICACS)*, Lahore, Pakistan, 2018, pp. 1–6.

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

[160] 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.

[161] 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.

[162] 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.

[163] 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.

[164] 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.

[165] 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.

[166] 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.

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

[168] 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.

[169] X. Li, W. Zhang, and Q. Ding, "A robust intelligent fault diagnosis method for rolling element bearings based on deep distance metric learning," *Neurocomput.*, vol. 310, pp. 77–95, 2018.

[170] 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.

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

[172] 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

[173] W. Lu, B. Liang, Y. Cheng, D. Meng, J. Yang, and T. Zhang. "Deep model based domain adaptation for fault diagnosis," in *IEEE Transactions on Industrial Electronics*, vol. 64, no. 3, pp. 2296-2305, 2017.

[174] 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.

[175] 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.

[176] W. Zhang, G. Peng, C. Li, Y. Chen, Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors.*, 17 (2) (2017), pp. 425.

[177] 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. PP, no. PP, pp. PP–PP, 2018.

[178] M. R. Islam, M. M. M. Islam and J. Kim, "Feature selection techniques for increasing reliability of fault diagnosis of bearings," in *Proc. 2016 9th Int. Conf. Electr. Comput. Eng. (ICECE)*, Dhaka, 2016, pp. 396–399.

[179] H. O. A. Ahmeda M. L. D.Wongb, A. K. and Nandiac, "Intelligent condition monitoring method for bearing faults from highly compressed measurements using sparse over-complete features," *Mech. Syst. Signal Process.*, vol. 99, pp. 459–477, Jan. 2018.

[180] L. Zhao and X. Wang, "A deep feature optimization fusion method for extracting bearing degradation features," *IEEE Access*, vol. 6, pp. 19640–19653, 2018.

[181] H. Ahmed and A. K. Nandi, "Compressive sampling and feature ranking framework for bearing fault classification with vibration signals," in *IEEE Access*, vol. 6, pp. 44731–44746, 2018.

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

[183] 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.