

Deep Learning Algorithms for Bearing Fault Diagnostics — A Review

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

TR2019-084 September 05, 2019

Abstract

This paper presents a comprehensive review on applying various deep learning algorithms to bearing fault diagnostics. Over the last ten years, the emergence and revolution of deep learning (DL) methods have sparked great interests in both industry and academia. Some of the most noticeable advantages of DL based models over conventional physics based models or heuristic based methods are the automatic fault feature extraction and the improved classifier performance. In addition, a thorough and intuitive comparison study is presented summarizing the specific DL algorithm structure and its corresponding classifier accuracy for a number of papers utilizing the same Case Western Reserve University (CWRU) bearing data set. Finally, to facilitate the transition on applying various DL algorithms to bearing fault diagnostics, detailed recommendations and suggestions are provided for specific application conditions such as the setup environment, the data size, and the number of sensors and sensor types. Future research directions to further enhance the performance of DL algorithms on healthy monitoring are also presented.

Symposium on Diagnostics for Electric Machines, Power Electronics and Drives (SDEMPED)

This work may not be copied or reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for nonprofit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of Mitsubishi Electric Research Laboratories, Inc.; an acknowledgment of the authors and individual contributions to the work; and all applicable portions of the copyright notice. Copying, reproduction, or republishing for any other purpose shall require a license with payment of fee to Mitsubishi Electric Research Laboratories, Inc. All rights reserved.

Deep Learning Algorithms for Bearing Fault Diagnostics – A 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—This paper presents a comprehensive review on applying various deep learning algorithms to bearing fault diagnostics. Over the last ten years, the emergence and revolution of deep learning (DL) methods have sparked great interests in both industry and academia. Some of the most noticeable advantages of DL based models over conventional physics based models or heuristic based methods are the automatic fault feature extraction and the improved classifier performance. In addition, a thorough and intuitive comparison study is presented summarizing the specific DL algorithm structure and its corresponding classifier accuracy for a number of papers utilizing the same Case Western Reserve University (CWRU) bearing data set. Finally, to facilitate the transition on applying various DL algorithms to bearing fault diagnostics, detailed recommendations and suggestions are provided for specific application conditions such as the setup environment, the data size, and the number of sensors and sensor types. Future research directions to further enhance the performance of DL algorithms on healthy monitoring are also presented.

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], [2]. The malfunction of electric machines can be generally attributed to various faults of different categories, which include the drive inverter failure, the stator winding insulation breakdown, the bearing fault, and the air gap eccentricity. Several surveys on the likelihood of induction machine failures conducted by the IEEE Industry Application Society (IEEE-IAS) [3] and the Japan Electrical Manufacturers' Association (JEMA) [4] reveal that bearing fault is the most common fault type that accounts for 30% to 40% of the total faults.

Since 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 decade. Specifically, this problem is approached by interpreting a variety of available signals, including vibration [7], acoustic noise [8], stator current [9], thermal-imaging [10], and multiple sensor fusion [11]. The existence of a bearing fault as well as its specific fault type can be readily determined by performing

frequency spectral analysis on monitored signals and analyzing their component at the characteristic fault frequency, which can be calculated by a well-defined mechanical model [5] that depends on the motor speed, bearing geometry and the specific location of a defect.

However, the accuracy of the mainstream bearing fault diagnostics method based on vibration or acoustic signals can be affected by background noise due to external mechanical excitation 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 [9], 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 current at bearing fault signature frequencies can vary at different loads, different speeds, and different power ratings of the motor itself, thus bringing challenges to identify the threshold stator current values to trigger a fault alarm at an arbitrary operating condition. Therefore, a thorough 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 loads and speeds. This process, summarized as “Learning Stage” in [12], is unfortunately very tedious and expensive to perform, and needs to be repeated for motors with different power ratings.

Most of the challenges described above can be attributed to the fact that all of the conventional methods rely solely upon values at fault characteristic frequencies to determine the presence of a bearing fault. However, there may exist some unique patterns or relationships hidden in the data 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 [13]–[16]. Most of the literature applying these ML algorithms report satisfactory results with accuracy over 90%.

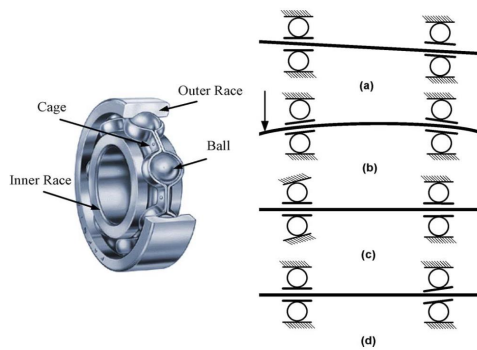


Fig. 1. Structure of a rolling-element bearing with four types of abnormal scenarios: (a) misalignment (out-of-line), (b) shaft deflection, (c) crooked or tilted outer race and (d) crooked or tilted inner race [5].

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. Although this literature survey does not include all DL papers on bearing fault diagnostics due to length limit, it is observed that the number of papers grew steadily in the last three years, clearly indicating booming interests in employing DL methods for bearing fault diagnostics. In this context, this paper seeks to present a thorough overview on recent research work devoted to applying deep learning techniques to bearing fault diagnostics.

The rest of this paper is structured as follows. In Section II, we give a brief overview to the bearing structure, and the CWRU dataset to be used for a comparative study. In Section III, we discuss the advantages of DL approaches compared with conventional machine learning methods, and introduce a variety of DL approaches that have been applied for bearing fault diagnostics. Then in Section IV, a systematic comparison is provided on the classification accuracy of DL algorithms. Based on those observations, in Section V we provide our recommendations for applying DL algorithms to bearing fault diagnostics, and future directions in this field.

II. BEARING STRUCTURE AND FAULT DATASET

The structure of a rolling-element bearing is illustrated in Fig. 1, which contains the outer race typically mounted on the motor cap, the inner race to hold the motor shaft, the ball or the rolling element, and the cage for restraining the relative distance of the rolling elements. Four common scenarios of misalignment are demonstrated in Fig. 1 (a) to (d).

Data is the foundation for all machine learning and artificial intelligence methods. To develop effective 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. A few organizations have made the effort and provided bearing fault datasets for people to work on the DL research, i.e., the CWRU dataset, which can serve as standards for the comparison of different algorithms.

III. DEEP LEARNING BASED APPROACHES

As a subset of machine learning, deep learning possesses powerful capability to learn and represent real-world applications with great flexibility as nested hierarchy from convoluted concepts to simpler concepts, and abstract representations computed in terms of perceptual ones. The trend of transitioning from conventional ML methods to deep learning can be attributed to the following reasons.

- 1) *Hardware evolution*: Training deep neural networks is extremely computationally intensive, but running on a high performance GPU can significantly accelerate this training process.
- 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.
- 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 [17], 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.*

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

- 1) *Best-in-class performance*: The complexity of the computed function grows exponentially with depth [18]. Deep learning has best-in-class performance that significantly outperforms other solutions on problems across multiple domains, including speech, language, vision, and gaming.
- 2) *Automatic feature extraction*: No need for feature engineering. Conventional machine learning algorithms usually call for sophisticated manual feature engineering which unavoidably involves expert domain knowledge and numerous human effort. With a deep network, there's no need for this.
- 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.

Due to these advantages, We are witnessing an exponential increase in DL applications. One such example is fault diagnostics and health prognostics, and bearing fault identification is a very representative case.

A. Convolutional Neural Network (CNN)

Inspired by animal visual cortex [19], convolution operation is first introduced to detect image patterns in a hierarchical way from simple to complex features. The low layer detects fundamental low level visual features such as edge and corner, and layers afterward detect higher level features.

The first paper employing CNN to identify bearing fault was published in 2016 [20]. For the next three years many papers

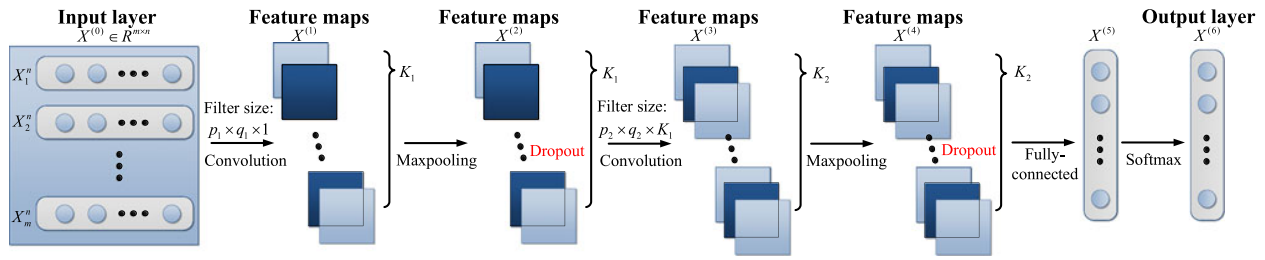


Fig. 2. Architecture of the CNN-based fault diagnosis model [25].

[21]–[23], [25], [26] 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. 2. 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 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 [20], the vibration data were acquired using two accelerometers, one installed on top of the housing, the other on the back. CNN is able to autonomously learn useful features for bearing fault detection from the raw data pre-processed by the scaled Fourier transform. The classification results demonstrate that feature-learning based approaches such as CNN can also identify some early-stage faulty conditions without explicit characteristic frequencies, such as lubrication degradation, which cannot be identified by traditional methods.

An adaptive CNN (ADCNN) is applied to the CWRU dataset to dynamically change the learning rate, for a better trade-off between training speed and accuracy in [21]. The entire fault diagnosis model employs 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 provides a better accuracy compared to conventional shallow CNN and SVM methods, especially for identifying rolling element (ball) defect. In addition, this proposed ADCNN is also able to predict the fault size (defect width) with satisfactory accuracy.

Similar to earlier work [20], [21], [22] implements a 4-layer CNN structure with 2 convolutional and 2 pooling layers on bearing fault detection, and the accuracy outperforms conventional SVM and shallow Softmax regression classifier. In particular, when the vibration signal is mixed with ambient noise, the improvement can be as large as 25%, showcasing the excellent built-in denoising capabilities of the CNN algorithm. A sensor fusion approach is applied in [23], in which both temporal and spatial information of the CWRU data from two accelerometers at both 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. The average accuracy with two sensors is 99.41%, comparing favorably against one sensor of 98.35%.

Some other variations of CNN are also employed to tackle the bearing fault diagnosis challenge [25], [26] on the CWRU dataset to obtain more desirable characteristics compared to the conventional CNN. For example, a CNN based on LeNet-5 [24] is applied in [25] containing 2 alternating convolutional-pooling layers and 2 fully-connected (FC) layers. 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 DL methods such as the adaptive CNN (98.1%) and deep belief network (87.45%), while dominating conventional 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 [26], with raw data transformed into spectrograms for easier processing. An accuracy of 99.22% is achieved, outperforming the 94.28% of linear SVM with particle swarm optimization (PSO), and 91.43% of conventional SVM.

B. Auto-encoders

Auto-encoder is proposed in 1980s as a method of unsupervised pre-training method for ANN [6], [51]. After evolving for decades, it 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. 3. An auto-encoder is trained from an ANN, which is composed of two parts: encoder and decoder. The output of the 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 aims at imitating the input as the final output. 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 various classifiers.

There are many studies employing auto-encoder in bearing fault diagnosis [27]–[30]. One of the earliest can be found in [27], where a 5-layer auto-encoder is utilized to mine fault characteristics from the frequency spectrum and effectively classify the health condition of machines. The classification accuracy reaches 99.6%, which is significantly higher than the 70% of back-propagation based neural networks (BPNN). In [28], an auto-encoder based extreme learning machine (ELM)

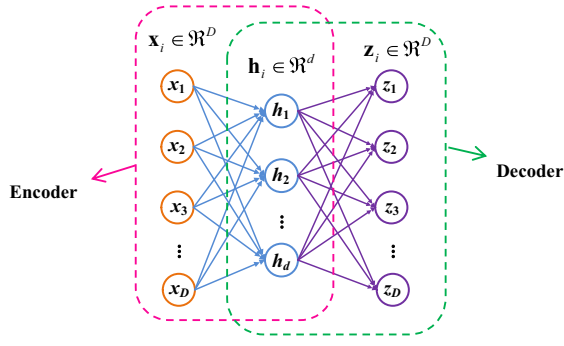


Fig. 3. Process of training a one hidden layer auto-encoder [31].

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 conventional 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 ELM.

Compared to CNN, the denoising capabilities of original auto-encoders is not prominent. Thus in [29], a stacked denoising autoencoder (SDA) is implemented, which is suitable for deep architecture-based robust feature extraction on signals containing ambient noise in volatile working conditions. To balance between performance and training speed, three hidden layers with 100, 50, and 25 units respectively are employed. The original CWRU bearing data are perturbed by a 15 dB random noise to 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 [29], another form of SDA is utilized in [30] with three hidden layers. Signals of the CWRU dataset are combined with different levels of random noises in the time domain and converted to frequency domain signals. The proposed method has better diagnosis accuracy than deep belief networks (DBN), particularly with the added noise, where an improvement of 7% in diagnosis accuracy is achieved.

C. Deep Belief Network (DBN)

In deep learning, a Deep Belief Network (DBN) is composed of unsupervised networks such as Restricted Boltzmann Machines (RBMs) or autoencoders, where each preceding hidden layer becomes the visible layer for the next, as demonstrated by boxes of different colors in Fig. 4. Often recognized as an undirected, generative energy-based model, an RBM consists of a “visible” input layer and a hidden layer with connections in between. The composition of RBNs leads to a fast and unsupervised training procedure and starts from the

lowest visible layer. The DBNs can be trained greedily, one layer at a time, and the contrastive divergence is applied to the next RBM in turn [32].

Its first application on bearing fault diagnosis was published in 2017 [33], in which a multi-sensor vibration data fusion technique is implemented to fuse the time domain and frequency domain features extracted via multiple 2-layer SAEs. A 3-layer RBM based DBN is then 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 the change of operating condition. In [34], 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 and 5 kernels in each layer. Finally, a Softmax layer is used with an average accuracy of over 95%.

Many DBN papers also employ the CWRU bearing dataset as the input data [35]–[37] thanks to its popularity. For example, an adaptive DBN and dual-tree complex wavelet packet (DTCWPT) is proposed in [35]. The DTCWPT first preprocesses 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 multiple scaling coefficients of the Daubechies wavelet, is taken as the basis function. Then a 5-layer adaptive DBN 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 [37], data from two accelerometers 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, while the accuracy employing CNN is also lower than DBN. Similar to this D-S theory based output fusion [36], a 4-layer DBN with different hyper-parameters coupled with ensemble learning is implemented in [37]. 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 those employing a single DBN of different weights (around 80%) and a simple voting ensemble DBN (91.21%).

D. Recurrent Neural Network (RNN)

Different from feed-forward neural network (FNN), RNN processes input data in a recurrent manner. The architecture is shown in Fig. 5. 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 sequence model, it can capture and model the sequential relationship in sequential data or time series data.

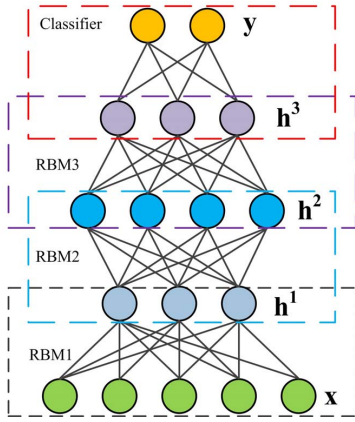


Fig. 4. Architecture of a DBN [33].

However, trained with back propagation through time (BPTT), RNN has the notorious gradient vanishing issue stemmed from its nature. Although proposed in 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 issue, LSTM has shown astonishing capability in modeling the long-term dependency in data, and therefore is taking a dominant role in time series and text analysis. It has received great success in speech recognition, natural language processing, video analysis, etc.

One of the earliest application of RNN on bearing fault diagnostics can be found in 2015 [38], where 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 showed that the proposed scheme based on RNN is capable of detecting and classifying bearing faults accurately, even under non-stationary operating conditions. In addition, a methodology using combined 1-D CNN and LSTM to classify bearing fault types is presented in [39], 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 reaches as high as 99.6%.

E. Others

There are also numerous other DL methods employed for bearing fault diagnostics, some are based on new algorithms, some are mixtures of many DL methods listed above. For example, in [40], a new large memory storage retrieval (LAM-STAR) neural network is proposed. In [41], both DBN and SAE are applied simultaneously to identify the presence of a bearing fault. Other examples include a mixture of CNN and DBN [42], deep residual network (DRN) [43], [44], deep stack network (DSN) [45], RNN based autoencoder [46], sparse

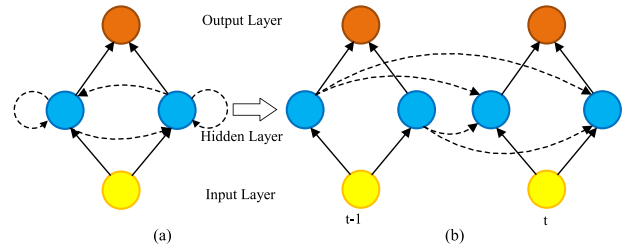


Fig. 5. (a) Architecture of the RNN, and (b) RNN across a time step [38].

filtering [50], and a new deep learning architecture capsule network [48] proposed by Hinton *et al.* [47].

IV. COMPARISON OF DL ALGORITHM PERFORMANCE USING THE CWRU DATASET

A systematic comparison of the accuracy of different DL algorithms employing the CWRU dataset is presented in TABLE. I. The testing accuracy of all of the DL algorithms are above 95%, which validates the effectiveness of applying deep learning to bearing fault diagnostics. However, we would like to stress that the specific values of testing accuracy are only demonstrated for an intuitive understanding for the following reasons:

- 1) *Generalization*: Some DL methods with accuracy over 99% are generally applied on a very specific dataset at a fixed operating condition. 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.
- 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 this scenario, accuracy should not be used as the only metric to evaluate an algorithm, and other metrics including the *precision* and *recall* should be introduced to further evaluation.
- 3) *Randomness*: Although these DL methods use the same dataset to perform classifications, the percentages 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.

V. RECOMMENDATIONS AND FUTURE DIRECTIONS

A. Recommendations and Suggestions

The successful implementation of ML and DL algorithms on bearing fault diagnostics can be attributed to the strong correlations among features that follow the law of physics. For researchers considering applying ML/DL methods to solve bearing problems, the author suggests the following:

TABLE I
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
[21]	Adaptive CNN (ADCNN)	3	Softmax	Predict fault size	50%	97.90%
[22]	CNN	4	Softmax	Noise-resilient	90%	92.60%
[23]	CNN	4	Softmax	Sensor fusion	70%	99.40%
[25]	CNN based on LeNet-5	8	FC layer	Better feature extraction	83%	99.79%
[26]	Deep fully CNN (DFCNN)	8	Connectionist temporal classification	Validation with actual service data	78%	99.22%
[27]	SAE	3	ELM	Adapt to load change	50%	99.61%
[28]	SAE	3	ELM	Reduce training time	50%	99.83%
[29]	Stacked denoising AE (SDAE)	3	N/A	Noise-resilient	50%	91.79%
[30]	SDAE	3	Softmax	Noise-resilient	80%	99.83%
[35]	dual-tree complex wavelet	5	N/A	Adaptive DBN	67%	94.38%
[36]	DBN	2	Softmax	Adapt to load change	N/A	98.80%
[37]	DBN with ensemble learning	4	Sigmoid	Accurate & robust	N/A	96.95%
[39]	CNN-LSTM	3	Softmax	Accurate	83%	99.6%

- 1) *Examine the setup environment*: Thoroughly examine the working environment and all possible operating conditions of the setup (i.e., indoor/outdoor, fixed/volatile operating conditions, etc.). For the simplest case with an indoor and a single operating point setup, conventional ML methods or even frequency-based analytical models should suffice. Otherwise, more advanced DL approaches with certain denoising blocks and extra layers should be considered to improve the diagnostic robustness.
- 2) *Sensors*: Determine the number and type of sensors to be mounted close to bearings. For traditional frequency based and ML based methods, one or two vibrations sensors mounted close to the bearing should be sufficient. For DL 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. 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 sensors can be installed to form a multi-physics dataset to enhance the performance.
- 3) *Data Size*: If the size of the collected dataset is not sufficient to train a DL model with good generalization, the algorithm and its training process should be selected to make the most out of the data and computation resources. For example, the dataset augmentation techniques such as generative adversarial network (GAN), and data random sampling with replacement such as bootstrapping can be used, as well as some cross validation methods such as leave-one-out, Monte Carlo, etc. With the problem of a small labeled dataset, a promising routine is to apply a semi-supervised learning paradigm.

B. Future Research Directions

The following research directions may facilitate the transition from research labs to real-world applications:

- 1) *More complicated datasets*: Apply more complicated bearing datasets to prevent the accuracy saturation problems.
- 2) *From artificial to real-world*: Make attempts to predict natural faults with algorithms trained by artificial faults, which is a reasonable expectation for DL-based fault indicators in real-world applications.
- 3) *Explainability*: Explain how DL methods work in fault detection applications. The interpretation of DL in general is not so developed as traditional ML methods. Several works, such as [53], [54], visualized the learnt CNN kernel to interpret its physical meaning. These works have provided intuitions on the explainability of DL, but more in-depth investigations are necessary.

VI. CONCLUSIONS

In this paper, a systematic review is conducted on existing literature employing deep learning algorithms to motor bearing fault diagnostics, which has spurred the interest of academia for the last five years. It is demonstrated that, despite the fact that deep learning algorithms require a large dataset to train, they can automatically perform adaptive feature extractions on the bearing data without any prior expertise on fault characteristic frequencies or operating conditions, making them very suitable for real-time bearing fault diagnostics. A comparative study is presented and discussed regarding the effectiveness of different algorithms. Insights and future research directions toward performance improvement and real-world implementation are provided.

REFERENCES

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

- [4] JEMA, "On recommended interval of updating IMs", 2000.
- [5] T. Harris, *Rolling Bearing Analysis*, Hoboken, NJ, USA: Wiley, 1991.
- [6] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Netw.*, vol. 61, pp. 85–117, Jan. 2015.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] D. López-Pérez and J. Antonino-Daviu, "Application of infrared thermography to failure detection in industrial IMs: case stories," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 1901–1908, May 2017.
- [11] E. T. Esfahani, S. Wang and V. Sundararajan, "Multisensor wireless system for eccentricity and bearing fault detection in IMs," *IEEE/ASME Trans. Mechatronics*, vol. 19, no. 3, pp. 818–826, 2014.
- [12] 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.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] (2018, December). Amazon Mechanical Turk, [Online]. Available: <https://www.mturk.com/>.
- [18] M. Raghu, B. Poole, J. Kleinberg, S. Ganguli, and J.-S. Dickstein, "On the expressive power of DNNs," *arXiv:1606.05336*, 2016.
- [19] 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.
- [20] 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.
- [21] X. Guo, L. Chen, and C. Shen, "Hierarchical adaptive deep convolutional neural network and its application to bearing fault diagnosis," *Meas.*, vol. 93, pp. 490–502, 2016.
- [22] 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.
- [23] M. Xia, T. Li, L. Xu, L. Liu and C. W. de Silva, "Fault diagnosis for rotating machinery using multiple sensors and CNNs," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 1, pp. 101–110, Feb. 2018.
- [24] 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.
- [25] 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.
- [26] 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.
- [27] 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.
- [28] 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*, vol. 231, no. 8, pp. 1560–1578, 2016.
- [29] C. Lu, Z.-Y. Wang, W.-L. Qin, J. Ma, "Fault diagnosis of rotary machinery components using a stacked denoising AE-based health state identification," *Signal Process.*, vol. 130, pp. 377–388, 2017.
- [30] 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.
- [31] 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.
- [32] Y. Bengio, Learning deep architectures for AI, *Foundations and Trends in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.
- [33] 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.
- [34] 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. Progn. Syst. Health Manag. Conf.*, 2017, pp. 1–7.
- [35] 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.
- [36] D.-T. Hoang and H.-J. Kang, "Deep belief network and dempster-shafer evidence theory for bearing fault diagnosis," in *Proc. IEEE Int. Sym. Ind. Electron. (ISIE)*, 2018, pp. 841–846.
- [37] 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.
- [38] 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.
- [39] 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.
- [40] M. He and D. He, "DL based approach for bearing fault diagnosis," *IEEE Trans. Ind. Appl.*, vol. 53, no. 3, pp. 3057–3065, June 2017.
- [41] 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.
- [42] 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.
- [43] 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.
- [44] 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.
- [45] 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.
- [46] 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.
- [47] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. NIPS*, 2017, pp. 3859–3869.
- [48] 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.
- [49] 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.
- [50] 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.
- [51] D. H. Ballard, "Modular learning in neural networks," in *Proc. AAAI'87 National Conf. AI*, Seattle, WA, 1987, pp. 279–284.
- [52] 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.
- [53] 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), p. 425.
- [54] 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. and Sig. Proc.*, vol. 100, 2018, p. 439.