Faculty of Engineering
University of Nigeria, Nsukka
22nd- 25th June, 2021

Proceedings of the
2021 Sustainable Engineering and
Industrial Technology Conference

Review of Deep Learning Algorithms in Motor Bearing Fault Detection

O. S. Jimoh, S. A. Bashir, I. M. Kolo, O. A. Abisoye

Department of Computer Science, Federal University of technology, Minna. ***Emails:*** oladebojimoh@gmail.com (OSJ), bashirsulaimon@futminna.edu.ng (SAB), idris.kolo@futminna.edu.ng (IMK), o.abisoye@futminna.edu.ng (OAA)

Abstract

Bearing fault in any rotary machines can cause equipment to break down thus causing critical safety, environmental or economic effect. Many mechanical equipment operate under tough working environment, which makes them vulnerable to various types and degrees of faults. As a result, bearing fault detection (BFD) and consistent monitoring of the health status of bearings has become important so as to ensure efficiency, avert complete breakdown or any catastrophic event and prevent/reduce financial loss. This has attracted researchers to work on BFD during the past few years because of its great influence on the operational continuation of many industrial processes. This paper provides a survey on some deep learning (DL) methods for motor BFD. Some common existing DL methods are briefly reviewed, highlighting their contributions, drawbacks and their significance in motor BFD. Finally, we point out a set of promising future works and draw our own conclusions by recommending long short term memory (LSTM) autoencoder (AE) as the best method to use for BFD based on certain advantage that we presented in this paper.

Keywords: fault detection, deep learning, bearing fault, machine learning, deep neural network

1. INTRODUCTION

Motor bearings are regarded as one of the most important parts of rotating machines such as turbines, industrial machines and automobiles [1, 2]. The importance of this component which is visually presented in Fig. 1 cannot be over emphasized since they are responsible for the smooth running of the rotary parts of machines.

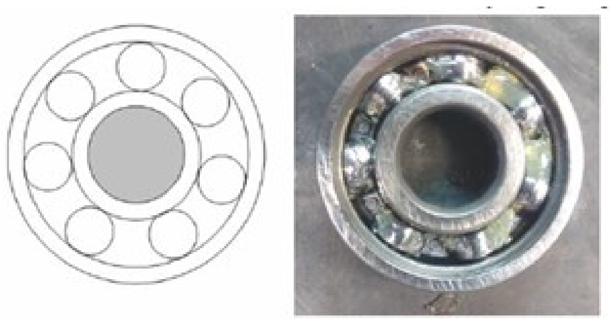

Figure 1: Diagram and plate of a healthy bearing [3].

However, it is important to note that the rate of degradation of motor bearings vary from one machine to another. This phenomenon is unavoidable but vary as a result of operational and environmental factors [3] such as the time of operation of machines where installed, ambient temperature, load factor [4] and maintenance ethics like constant oiling. For sustainability of machines, safe points of these factors have to be considered to avoid anomaly of bearings that could lead to high maintenance cost such as bearing faults, fatigue and accelerated aging [1] or even breakdown which accounts for 30% to 40% failures of machines [5]. With this, production in industries could halt as a result of prolonged downtime due to anomaly [1, 6] or breakdown. Furthermore, it is important to note that complete breakdown could be catastrophic [6] especially in automobiles and heavy duty production machines which could as a result of sudden failure during operation, jet parts from the machine can lead to accidents and sometimes death [3]. This therefore suggests that consistent monitoring of the health status of bearings is important [7] so as to ensure efficiency and avert complete breakdown. According to [1], motor bearings which are made up of the inner race way, outer race way and ball bearing as shown in Fig. 2 could suffer three different kinds of defects leading to anomaly and on the long run breakdown [1, 8]. These defects shown in Fig. 2 include inner race way defect, outer race way defect and lubrication defects [1, 8]. [3] also, pointed out defects which could be on the ball bearing as shown in Fig. 3. All these can be curbed via adequate maintenance.

2. DEEP LEARNING ALGORITHMS FOR BEARING FAULT DETECTION

Over the years, as a result of the quest to run efficient industrial process involving the use of rotary machines, anomaly detection in motor bearing have been looked into by a lot of researchers and come up with techniques to detect impending defects in motor bearing before it becomes permanent damage.

DL algorithms has been successfully employed in fault detection such as autoencoder (AE), denoised

ISBN 0000-0000

Figure 2: (a) Healthy bearing (b) inner race defect (c) outer race defect (d) lubrication defects [1, 8].

Figure 3: Diagram and plate of faulty [3].

autoencoders (DAE), deep belief networks (DBN), convolutional neural network (CNN), recurrent neural networks (RNN), and long short-term memory (LSTM), etc. Before the recent widespread adoption of DL, a variety of classical less effective machine learning (ML) and data mining algorithms have been used for many years, e.g. the artificial neural network (ANN). These algorithms requires a lot of domain expertise, complex feature engineering and also, it is hard to keep up a reasonable level of transferability of ML models trained in one domain to be generalized or transferred to other contexts or settings. Therefore, many DL algorithms with automated feature extraction capabilities and better classification performance have been applied to machine health monitoring and fault diagnostics, among which BFD is a very representative case.

DL, a subset of ML that attains great power and flexibility by learning to represent a problem as a nested hierarchy of concepts, where each concept is defined in relation to simpler concepts, and representations that are more abstract are computed from less abstract ones.

2.1. Convolutional Neural Network (CNN)

CNN is a successful algorithm in DL, based on ANN and inspired biologically by mammalian visual cortex. This operation was first introduced to find patterns in images in an orderly manner from its simple features like edge and corner down to its complex features. First study that implemented CNN to detect motor bearing fault was published in 2016 [9], subsequent years; several papers that applied same method were published and they added to the progress of BFD in different areas.

2.2. Autoencoder (AE)

AE was first recommended as an unsupervised pretraining method for ANN in the 1980s [10]. Subsequently, decades of advancement has made AE become widely adopted. AE is trained to encode the input x into a representation r(x) in a way that input can be reconstructed from r(x) [11]. The target output of the AE is thus the AE input itself. Hence, the output vectors have the same dimensionality as the input vector. There are other variants of autoencoders that have been proposed over the years.

Generally, AE are neural networks that deduce features of low level signals by comparing the differences between the input data and the output data. AEs has less denoising capacity when compared to CNNs [12], this challenge brought about the implementation of stacked denoised autoencoder (SDAE) which is suitable for feature extraction on signals that contain ambient noise under different working conditions [13].

2.2.1. Denoise Autoencoders (DAE)

Denoise Autoencoders (DAE) is the stochastic version of AE where the input is randomly corrupted, but the uncorrupted input is still used as target for the reconstruction [20]. It tries to encode the input; preserve the information about the input, and then it tries to undo the effect of a corruption process applied to the input of the autoencoder. The use of autoencoders for denoising was introduced in earlier works such as [21], but the significant contribution of [20] lies in the demonstration of the successful use of the method for unsupervised pretraining of a deep architecture and linking the denoise autoencoder to a generative model. DAE is developed from AE but is more robust, since DAE assumes that the input data contain noise and is suitable to learn features from noisy data. As a result, the generalization ability of DAE is better than AE. Furthermore, DAE can be stacked to attain high level features, which give rise to stacked denoise autoencoders (SDAE) approach.

2.2.2. Stacked Denoise Autoencoder (SDAE)

Stacked Denoise Autoencoder (SDAE) is stacking up DAE to build a deep network which has more than one hidden layer [16]. In a typical SDAE structure, that includes two encoding layers and two decoding layers, in the encoding part the output of the first encoding layer acted as the input data of the second encoding layer.

Table 1: Summary of DL algorithms for fault detection.

Deep Learning Algorithms	Features	Applications/Related Works
Convolutional Neural Network (CNN)	Advantages: • Few neuron connections are required than in a typical ANN. • Classical CNN exhibits a good capability for denoising Shortfall: • May need several layers to lo-cate a whole hierarchy and large labeled datasets.	• [14] enhanced CNN with enlarged receptive fields to capture fault information. • [14] developed a CNN to learn features from raw data, frequency spectrum and combined time-frequency data. • In [9] CNN independently learn useful features for BFD from raw data pre-processed by the scaled discrete fourier transform (DFT). Different variations of CNN has been employed to tackle the bearing fault diagnosis challenge to achieve a more desirable characteristics and accuracy.
Autoencoder (AE)	Advantages: • Labeled dataset is not required, more noise-resilient and robust. • Ability to work without any preprocessing or predetermined transformations or manual feature engineering. Shortfalls: • Pre-training stage is mandatory. • Vanishing of error issue.	Many researchers employed different variations of autoencoders in BFD, each achieved a different case accuracy. • [13] carried out a detailed study of SDA for bearing fault diagnosis and achieved a worst case accuracy of 91.79%. • [10] proposed a deep AE constructed with DAE and contractive auto-encoder (CAE) to locomotive bearing dataset for the enhancement of feature learning ability which achieved classification accuracy 91.90%.
Denoise Autoencoder (DAE)	Advantages: • It's very robust to noise. • It can be shown to correspond to a generative model. Shortfall: • It. May eliminate important information in the input data.	• [15] enhanced depth feature fusion method for fault diagnosis of rotating machinery. • In order to learn more abstract features, [16] stacked together multiple AEs to form a stacked AE which give a better reconstruction of data. • A SDAE was introduced in [13] to distinguish anomaly and health condition of rotary machinery components.
Stacked Denoise Autoencoder (SDAE)	Advantage: • Can obtain higher level features. Shortfall: •It do not correspond to a generative model	
Deep Belief Network (DBN)	Advantage: • It makes use of a layer-by-layer greedy learning approach to initialize the network. Shortfall: • Training can end up being computationally expensive due to the initialization process and the sampling stage.	• [15] implemented a 3-layer RBM based DBN which attain a 97.82% accuracy of effectively identify bearing faults even after a change of operating conditions. • [11] trained one of the first effective deep learning algorithms. • [15] proposed a sparse AE-DBN technique in order to improve fault detection and diagnosis reliability which effectively identify the condition and performed better than other methods.
Recurrent Neural Network (RNN)	Advantages: • Sequential events are memorized by this model which enable it to make predictions based on the past. • Capable of modeling time dependencies and receiving inputs of variable lengths. Shortfall: • It has gradient vanishing/exploding issue.	• The earliest application on bearing fault diagnostics is in [17] where it was shown that proposed scheme based on RNN is capable of detecting and classifying bearing faults accurately, even under non-stationary operating conditions. • [18] proposed RNN based Health Indicator that predict the bearing health indicator from which RUL was estimated which give a better performance than any other self-organizing map method.
Long Short Term Memory (LSTM)	Advantages: • Can deal with unlimited number of states. • Doesn't have problem with notion of recency • Ability to bridge very long time lags. • It's local in both space and time. Shortfall: • It increase computational complexity.	• LSTM-AE approach was proposed in [19] to detect faults and anomalies in rotating machines where 70% of the data was used to train the AE, threshold was calculated with 10% and the last 20% was used for evaluating the approach (i.e. detection ability). The approach gives a really good results of 99.6% accuracy. And can be applied to different rotating machines (RMs) of a similar kind, which enables transfer of anomaly knowledge from one RMs to another.

2.3. Deep Belief Network

Deep Belief Networks (DBN) is a generative graphical model or class of deep neural network, composed of multiple layers of hidden variables (values) with relation between the layers but not the values [22]. Each sub network's hidden layer serves as the visible layer for the next, there are many attractive implementations of DBNs in real life applications and its first application on BFD was published in 2017 [15].

2.4. Recurrent Neural Network

Recurrent Neural Networks (RNN) are relatively old like many other DL algorithms, they are initially created in the 1980's. RNN as a sequence-based model processes the input data in a recurrent pattern, capture and model sequential relationships in sequential data or time-series data. In 2015, one of the earliest applications of RNN on bearing fault diagnostics was reported [17]. RNN has gradient vanishing/exploding issue emerged from its nature which make it have limited applications until the birth of LSTM in 1997.

2.5. Long Short Term Memory

Long Short Term Memory (LSTM) are a type of RNN that can integrate the temporal information into the network and maintain a hidden state vector which acts as a memory for the past information [23]. RNN based health indicator (RNN-HI) was proposed in [18] to predict the remaining useful life (RUL) of bearings with LSTM cells used in RNN layers. There are other DL algorithms and combined approaches use in BFD such as: Restricted Boltzmann Machine (RBM), Deep CNN, LSTM-CNN, LSTM-AE etc. A brief reviewed is done due to guidelines restriction in paging.

3. CONCLUSION

A review has been presented on some of the existing DL algorithms for BFD that has captured the attention of the research community over the years in this paper. Due to the fact that DL algorithms have automated feature extraction capabilities, better classification performance and transferability which makes them promising alternatives to perform real-time BFD. However, our overall recommendation for BFD is the combination of LSTM and Autoencoder (LSTM-AE). Since AE can handle unlabeled data, more resilient to noise and more robust, combined with LSTM's ability to retain information of former time stamp, deal with unlimited number of states and large multiple data. Future work can be done to build better model with the use of LSTM-AE by transforming time domain signals into frequency domain signals with the use of Fast Fourier transform (FFT) to detect fault in bearing faster with better sensitivity and accuracy at a lower threshold than that of [19].

References

[1] A. Choudhary, D. Goyal, and S. S. Letha. Infrared thermography based fault diagnosis of induction motor bearings using machine. *IEEE Sensors Journal*, 21(2), 2020.

[2] G.D. Novaes, P. Leite, A. Maurício, P. André, C. Rosas, T. Stosic, and B. Stosic. Entropy measures for early detection of bearing faults. physical a. *Physica A: Statistical Mechanics and its Applications*, 514:458–472, 2019.

[3] A. Glowacz. Fault diagnosis of single-phase induction motor based on acoustic signals. *Mechanical Systems and Signal Processing*, 117:65–80, 2019.

[4] Z. Shen, Y. Fei, W. Bingnan, and G.H. Thomas. Semi-supervised learning of bearing anomaly detection via deep variational autoencoders. *MITSUBISHI ELECTRIC RESEARCH LABORATORIES*, TR2019-153, 2019.

[5] C. Cooper, J. Zhang, R.X. Gao, P. Wang, and I. Ragai. Anomaly detection in milling tools using acoustic signals and generative adversarial networks. *Procedia Manufacturing*, 48:372–378, 2020.

[6] X. Jin, S. Member, Z. Que, Y. Sun, Y. Guo, and W. Qiao. A data-driven approach for bearing fault prognostics. 9994(c).

[7] S. Zhang, F. Ye, B. Wang, and T. G. Habetler. Few-shot bearing anomaly detection via model-agnostic meta-learning. *2020 23rd International Conference on Electrical Machines and Systems (ICEMS)*, pages 1341–134, 2020.

[8] M. Meire and P. Karsmakers. Comparison of deep autoencoder architectures for real-time acoustic based anomaly detection in assets. *2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, pages 786–790, 2019.

[9] A. Singh, C. Howard, and C. Hansen. Convolutional neural network based fault detection for rotating machinery. *Sound Vib. J.*, 377:331–345, 2016.

[10] H. Shao, H. Jiang, H. Zhao, and F. Wang. A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mech. Sys. Signal Process*, 95:187–204, 2017.

[11] Y. Bengio. Learning deep architectures for ai, foundations and trends in machine learning. *Foundations and Trends in Machine Learning*, 2(1):1–27, 2009.

[12] S. Zhang, S. Zhang, B. Wang, and T.G. Habetler. Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification. In *Proc. IEEE Int. Symp. Diagnostics Electr. Mach., Power Electron. Drives,*, pages 257–263, Toulouse, France, 2019.

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

[14] L. Jing, M. Zhao, P. Li, and X. Xu. A convolutional neural network based feature method for the condition monitoring of gearbox. *Measurement*, 111:1–10, 2017.

[15] Z. Chen and W. Li. Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. *IEEE Trans. Instrum. Meas.*, 66(7):1693–1702, 2017.

[16] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.A. Manzagol. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11:3371–3408, 2010.

[17] 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.*, 26(3):241–254, 2015.

[18] 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.*, 240:98–109, 2017.

[19] S. Ahmad, K. Styp-Rekowski, S. Nedelkoski, and O. Kao. Autoencoder-based condition monitoring and anomaly detection method for rotating machines. 2021.

[20] P. Vincent, H. Larochelle, Y. Bengio, and P.A. Manzagol. Extracting and composing robust features with denoising autoencoders. In W.W. Cohen, A. McCallum, and S.T. Roweis, editors, *Proceedings of the Twenty-fifth Int. Conf. on ML (ICML'08)*, 2008.

[21] P. Gallinari, Y. LeCun, S. Thiria, and F. Fogelman-Soulie. Memoires associatives distribuees. In *Proceedings of COGNITIVA 87*, Paris, La Villette, 1987.

[22] G. Hinton. Deep belief networks. *Scholarpedia*, 4(5):5947, 2009.

[23] K. Greff, R.K. Srivastava, J. Koutnik, B.R. Steunebrink, and J. Schmidhuber. LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10):2222–2232, 2016.