2021 IEEE International Conference on Prognostics and Health Management (ICPHM) | 978-1-6654-1970-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICPHM51084.2021.9486591

Application of Deep CNN-LSTM Network to Gear Fault Diagnostics

T. Haj Mohamad*, A. Abbasi†, E. Kim‡, and C. Nataraj†
*Palo Alto Research Center (PARC, a Xerox Co.)
3333 Coyote Hill Road, Palo Alto, CA 94304, USA
tmohamad@parc.com
†Villanova Center for Analytics of Dynamic Systems
Villanova University, Villanova, PA 19085, USA
http://vcads.org
a.abbasi@villanova.edu
nataraj@villanova.edu
‡College of Computing and Informatics
Drexel University, PA 19104, USA
ek826@drexel.edu

***Abstract*—Condition-based maintenance (CBM) is an optimum predictive maintenance framework that proposes maintenance actions based on monitoring the state data of an asset. Diagnostics is a principle concept in this framework and deals with fault detection, identification and isolation. Improving performance of diagnostics methods is of importance since it can result in reducing downtimes, improving operation reliability, reducing operations and maintenance costs. On the other hand, development of computational resources and sensory facilities could contribute highly to data based diagnostics approaches. The current paper studies one of these approaches that is categorized under deep learning (DL) concepts for a fault classification problem. A convolutional neural network (CNN) is used along with a long short term memory (LSTM) network for fault classification of vibration data of a helicopter gearbox mock-up system. In experimental tests, multiple gears at different conditions e.g. healthy gear and defective gears with root crack on one tooth, multiple cracks on five teeth and missing tooth, are taken into account. A deep learning model is built and its performance is evaluated using post processing techniques.**

I. Introduction

Rapid development of mechanical assets in the past decades has led to high demands for operational reliability, availability, maintainability and safety (RAMS). Historically, these demands have been addressed by different strategies like post-failure repairs, preventive maintenance and condition based maintenance (CBM). In contrast to the first two strategies, CBM can make maintenance decisions efficiently only when it is needed. CBM includes two principal concepts: diagnostics and prognostics. Diagnostics is responsible for fault detection, identification and isolation; prognostics tries to predict fault progression and degradation evolution [1], [2]. Fault detection and identification are critically important in the CBM framework since they detect abnormal behavior of a system and its source. Diagnostics methods can be categorized into physics-based, data-driven and hybrid ones. Performance of physics-based approaches decreases as the mechanical assets becomes more complex. That can be related to many simplifications and assumptions required in such approaches. Hence, mathematical abstractions of complex and nonlinear phenomena are not always achievable. As a case in point, fault diagnostics of industrial components such as gears which are complex and nonlinear has always been a challenging problem [1], [3]–[5].

On the other hand data-driven based methods do not need any prior knowledge of the underlying physics of a complex system. However, their performance is highly dependent on availability and quality of data. Such dependency on data has resulted in emergence of many data processing techniques like scrubbing, fusion, transformation, feature extraction, feature ranking, and feature selection. Specifically in feature extraction, the primary objective is characterizing the input signal to attain general information about the system. These features should be informative and non-redundant. Pattern recognition is considered as one of the data-driven approaches that widely uses this technique. It is mostly concerned with extracting a feature set that is used as a condition indicator or condition signature to detect and identify the system's health condition [6], [7]. In the area of gear diagnostics, that is the focus of the current paper, a wide range of feature extraction techniques have been studied in literature [8]–[17]

Recent advances in sensing technologies could address high demands of these methods to data. In addition, improvement of computation resources made complicated data-driven approaches applicable. Hence, deep learning (DL) as an advanced data-driven approach has evinced a lot of interest. DL methods like conventional ones try to build the best mapping from input to output domains; however, this mapping is highly complex that makes the mapping more general and robust. Many deep learning methods have been used in diagnostics, prognostics and health management like deep belief networks [18], auto-encoders [19], [20] and sparse filtering [21].

In deep learning studies, convolutional neural networks (CNNs) have been at the center of interest in recent years. They were fundamentally inspired by the mammalian visual cortex and have been widely implemented in image recognition, natural language processing and voice recognition, video classification, and object detection [22]. A notable point about CNNs is their ability in self-extraction of features as a part of the learning process. This specification results in deeper exploration of data domain and highly robust and informative features. This capability makes these networks a powerful tool in machinery health management and CBM. Chong [23] showed that CNNs can efficiently convert one dimensional vibration signals to two dimensional gray images with capability of extracting robust features. The performance of the proposed approach was demonstrated in fault detection of an induction motor. Ince et al. [24] proved the robustness of a one dimensional CNN in real time motor fault diagnostics. Other studies also showed that considering the frequency spectrum of mechanical systems as input of CNNs can provide an accurate mapping to fault space for diagnostics application [25].

In another application of DL, a notable part of the literature has focused on modeling sequential data, and specifically time series prediction. The primary idea is on giving the model a notion of time. This goal was achieved in Artifical Neural Networks (ANNs) by considering the dependency between successive inputs and was implemented in recurrent neural networks (RNNs). RNNs have been modified to many new networks with higher capabilities for many applications like machine translation, face recognition, speech recognition and response prediction [26]. Specifically, in machinery health management area, time awareness of these networks has been used in many applications like time response classification [27]–[29] and fault prognostics [30]–[32].

The current paper studies application of DL in fault classification of gears. Specifically, in early detection of faults in a complex rotating system with multi-fault modes which is a very challenging problem as established in prior studies [14], [15], [33]. In the developed framework, CNN's ability in self extraction of features is coupled with a special class of RNNs called long short-term memory (LSTM) network. Also, a novel approach for fusing the sensory data is proposed. The integrated network is trained using the obtained experimental data. Upon training the network, its performance is evaluated using post-processing techniques. Section II presents methodology including deep learning concepts, architecture of networks, data acquisition considerations and model training. In Section III the performance of the proposed framework is evaluated and finally conclusion is presented in Section IV.

II. Methodology

A. Convolutional Neural Networks

CNNs are a class of artificial neural networks with capability of self-extraction of features as a part of learning from the training data. According to Fig. 1, the architecture of CNNs consist of two main parts. The first part extracts features and feeds them to the second part for mapping to outputs. The first part includes an input layer and stack of pairs of convolutional and pooling layers. The convolutional layers are not fully

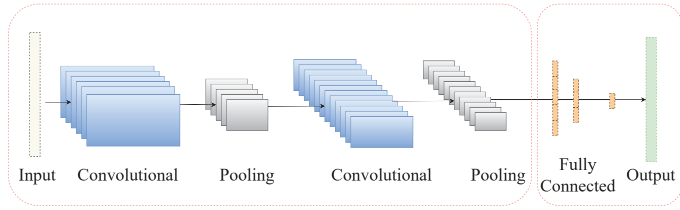

Fig. 1: Typical CNNs architecture

connected. This means neurons in the first convolutional layer are connected partially to the input layer neurons. This results in a gradual feature extraction in these networks. In detail, it initially concentrates on small superficial features in the primary convolutional layers and enriches them to extract larger deep features on subsequent ones. In this pass, nonlinear activation functions are used in the convolutional layers neurons. Nonlinear ones are used, since without a nonlinear transformation deep level of abstraction and generalization in features cannot be achieved. The local connections and enriching features step by step are said to be analogous to the perceptive fields and information narrowing down of simple and complex cells of the visual cortex. Although feature extraction by a network causes additional computation cost, it gives the opportunity of exploring larger parameter spaces. Unlike fully connected deep networks, e.g., deep multi-layer perceptrons (MLPs), CNNs use partial connections as it was mentioned. This characteristic improves information flow in CNNs that results in higher performance in dealing with large data-sets compared to fully connected deep networks.

In this architecture, pooling layers are concerned with subsampling of data structure in the network for computational cost reduction. The features obtained after several convolutional and pooling layers in the first part are passed over to the second part. This part maps features into the output layer with a fully-connected network. It can be even a conventional classifier that uses prior extracted features for classification.

B. Long Short-Term Memory Neural Networks

Classical feedforward ANNs do not have any notion of time due to single directional flow of information. That results in weakening their performance in learning from sequential data like time series. RNNs have emerged to address this limitation by modification in network architecture. In detail, they have recurrent connections along with forward ones as can be seen in Fig. 2a. Each recurrent connection can be interpreted as a feedback loop that feeds computations of the past time steps to the current step. This capability helps to learn dynamics and provides a hidden memory that retains what has been learned from in previous time steps. Although adding the loops highly improves networks performance in handling sequential data, it adversely affects trainability due to problems like gradient vanishing and exploding. They happen when low and high rates of the network error gradient prevents it from updating

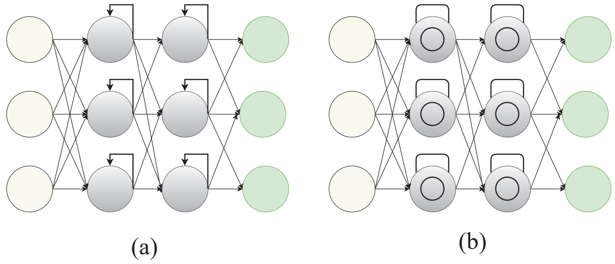

Fig. 2: Typical RNN and LSTM architecture

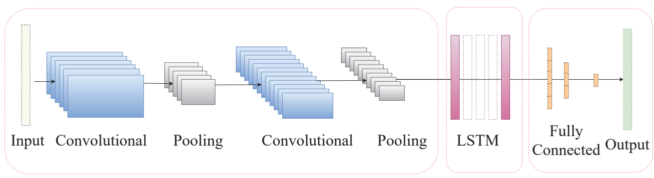

Fig. 3: Typical CNN-LSTM architecture

Fig. 4: Gear-train experimental setup [14], [15]

and converging, respectively. LSTM neural networks have emerged to address these problems by further modification in architecture. This goal was attained by adding a new element to the neurons called gate that controls flow of information. Fig. 2b shows a typical architecture of these networks in which gray circles are gated neurons. In this networks, in addition to hidden memory gates with three types of input, output and forget are implemented. The input gate sets the rate of keeping information from the past layer in the the cell. The output gate determines the rate of passing information from cell. The forget gate is concerned with disregarding the information and canceling out the dependency between the layers.

C. CNN-LSTM

The ability of CNNs in feature extraction can be leveraged by that of LSTMs in sequential data modeling in a coupled architecture. The idea of combination of these two networks led to CNN-LSTM networks with a hybrid architecture according to Fig. 3. In a CNN-LSTM network, conventional layer's ability in abstraction and generalization of features extracted from time-series data is powered by LSTM's ability in modeling short-term and long-term dependencies in these data. That results in a more accurate mapping from input to output domains. This mapping can be a fault classification using time series data.

CNN-LSTM network application in fault classification of gears is implemented in the following sections using recorded time series data.

D. Test and Data Acquisition

Deep networks, and especially CNN-LSTM, which is the focus of the current study, need larger datasets for training compared to shallow ones. That can be clearly related to a more complex architecture and higher number of parameters. This demand intensifies the importance of field operation considerations that are discussed in detail in following sections.

1) Experimental Setup: In this study a gear-train experimental set up is considered for a case study. This set up is a mock-up of a helicopter gear box system located at the United Technologies Research Center (UTRC) [14], [15]. The gear-train with length of 5 m includes motor, dynamometer and four gearboxes according to Fig. 4. In this set up each gearbox contains four spur gears. A schematic of this structure can be seen in Fig. 5. Multiple test gears with various health conditions were considered in the experiments. For this purpose, the gear located in the gearbox number 3 (see green color in Fig. 5) was changed with various test gears without any change to rest of the setup.

Three defective gears and one healthy were used in this study. Each gear has 23 teeth and the defects are for crack on one tooth, multiple cracks on five teeth and missing tooth. These defects can be seen in Fig. 6 in which a, b and c show a single root crack of 2 mm, locations of the five root cracks and corresponding teeth numbers, and sizes of the five root cracks, respectively.

2) Vibration Signal Acquisition: For data acquisition, vibrational signals were recorded at a sampling frequency of $102, 400$ Hz by a triaxial accelerometer mounted on the gearbox number 3. Encoders and a tachometer were used to measure rotational speeds of shafts A, C, and B with a 360 pulse/rev resolution. The tachometer on shaft "B" measured shaft rotational speed at a rate of 1 pulse/rev. Given the gear teeth ratio, the test gear shaft and shaft "B" had the same rotational speed. The rotational speed of the motor and test gear shaft was set at 900 rpm and 94 rpm. The time length of signal recording of the gearbox number 3 was 64 seconds at healthy, single crack tooth and multiple crack teeth conditions and for 3.2 seconds for the missing tooth condition. Samples

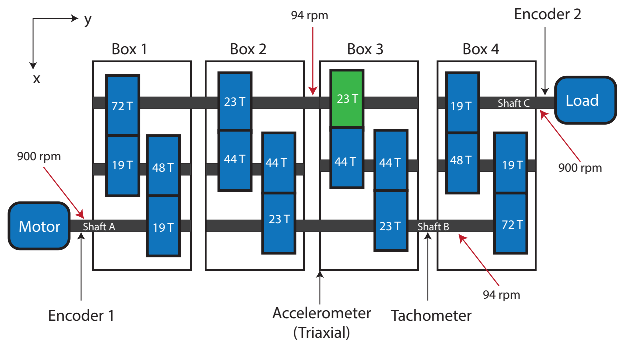

Fig. 5: Gear-train schematic [14], [15]

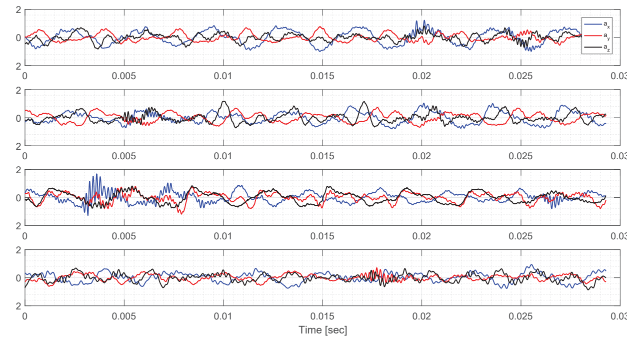

Fig. 7: Samples of the gear vibration data for each configuration and direction [14], [15]

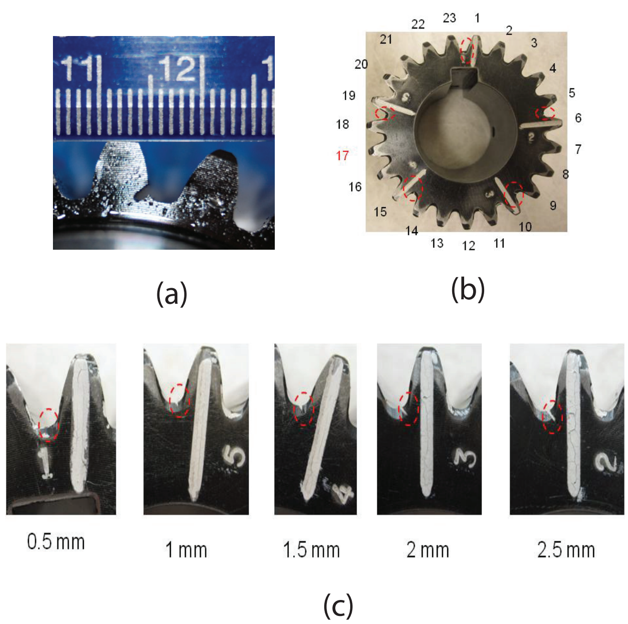

Fig. 6: Gear defects: a) crack on one tooth b) crack on 5 teeth c) crack sizes on 5 teeth [14], [15]

of the measured vibration for different gear conditions in the three directions are shown in Fig. 7.

3) Dataset Preparation: In order to analyze the vibration data measured by the accelerometer for various gear conditions, they were divided into multiple segments of revolutions. It was done using the tachometer signal that can provide the shaft phase information. The segment $x_k(i)$, for $i = 1, 2, ..., N$ is the vibration data of revolution number k of a total number of K revolutions. The total number of data segments that were obtained is $1,011$ segments in x, y and z directions. In each direction, 97 data segments were obtained for each healthy, single crack defect, and multiple crack defect conditions while 46 data segments were obtained for the missing tooth defect condition. These segments were compiled in a dataset for further exploration.

E. Application of CNN-LSTM to Fault Classification

A CNN-LSTM network was trained using this dataset for classification of input signal to healthy, single crack defect, multiple crack defect and missing tooth defect classes. Architecture of this network can be seen in Fig. 8. The data segments in x, y and z directions with dimensions of 500 by 1 pass in three parallel directions to CNNs; then, the extracted features are concatenated and are fed into the LSTM network for time series classification. This concatenation can be interpreted as a sensor fusion in which information from multiple time series are integrated. The dataset is divided with 80 percent for the training process and the rest for testing. Each CNN has three convolutional layers with 64 filters and kernel size of 20 with unit stride. Pooling layers are of max-pooling type. In LSTM part, the number of hidden units is 200 and the network is trained by the Adam optimization algorithm.

III. Results and Discussion

In order to evaluate the performance of the trained network, confusion analysis is performed. It can be represented by a confusion matrix in which the diagonal and off diagonal elements are the number or percentage of correct and incorrect classifications of unseen samples. They are here used to determine other performance criteria, e.g., overall accuracy, recall and precision.

Accuracy is one of the most important criteria that represents the rate of correct predictions and is defined as:

$$ACC = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

where numerator and denominator terms come from the confusion matrix elements. TF stands for true positive in which classification is correct for each defective condition. FP represents false positive in which mis-classifications happens for each defective condition. It is worth mentioning that mis-classification between defective conditions is also counted. Correct classification as faulty and incorrect classification as healthy are shown by TF and FN, respectively. Accuracy, at the best and worst conditions would be 1 and 0, respectively.

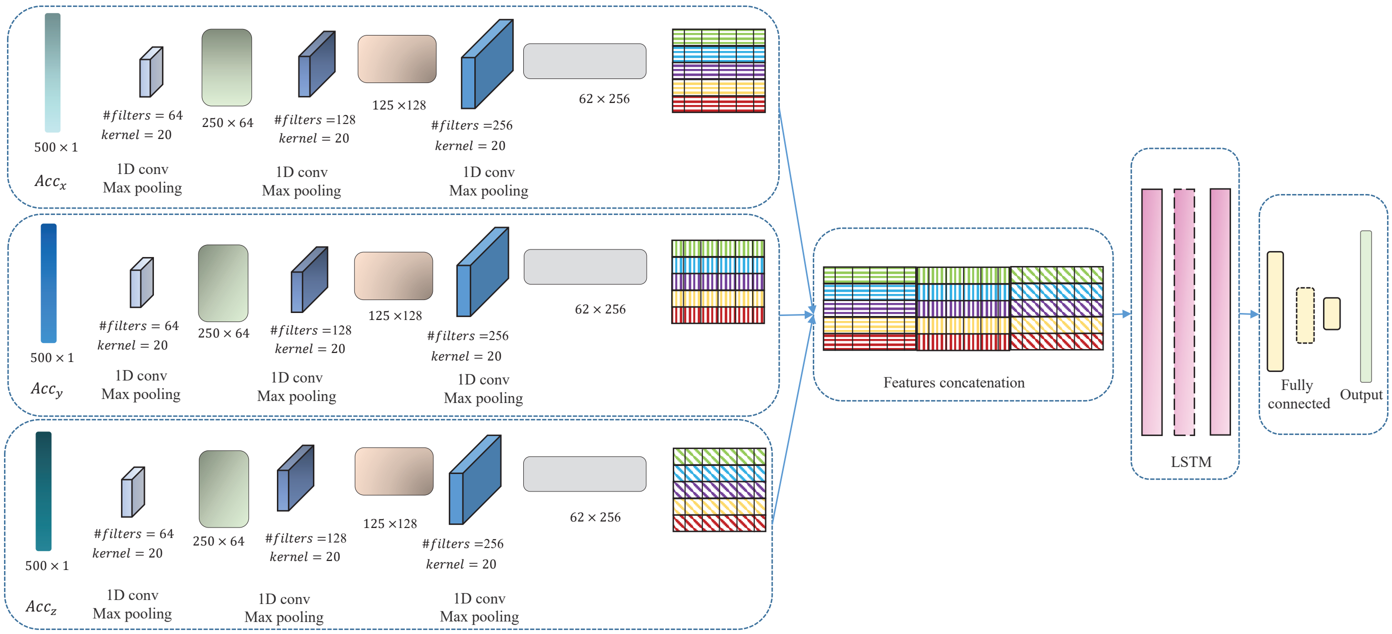

Fig. 8: CNN-LSTM architecture

Recall, which is also known as sensitivity or true positive rate, is another criterion that is defined as:

$$RCL = \frac{TP}{TP + FN} \tag{2}$$

The recall at the worst and best conditions would be 0 and 1, respectively.

Precision is the correct positive prediction divided by the total number of positive predictions and can be formulated as:

$$PREC = \frac{TP}{TP + FP} \tag{3}$$

Values of these criteria for the trained network can be seen in Table I in percentage. It shows 97% overall accuracy.

The precision of the classier in predicting both missing tooth and multiple crack conditions is maximum (100%), while it is 94.3% and 97% for the healthy and single crack conditions. High performance in classification between healthy and faulty classes is another notable result of this network that leads to early crack detection. This task has been challenging in the past because of the similarities between the corresponding signals [14], [15], [33]. In addition, given the number of classes, this network provides a generalization of previous approaches [14], [15], [33] for diagnostics of multiple cracks with various severities. In addition, the obtained results prove the capability of the proposed architecture for fusing multiple sensory data.

TABLE I: Performance criteria (percentage)

Class	Precision	Recall
H	94.3	97.3
SC	97.3	94.3
MC	100.0	100.0
MT	100.0	100.0
Overall Accuracy: 97.0		

IV. Conclusion

The current study has been concerned with the application of deep learning to fault classification of a mechanical system. A mockup of a helicopter gear box system was used with various faults. Four gear conditions including healthy, single crack, multiple crack, and missing tooth have been considered. Vibration signals along three principal directions have been recorded by the accelerometer to compile a dataset. A hybrid CNN-LSTM network was trained based on the dataset. In this hybrid network, the ability of CNN in extracting robust and informative features has been combined by that of LSTM in time series modeling. This network has been used for multi-class fault classification. The overall accuracy of this network is over 97%. Also, it is clear that the demonstrated network shows high performance in sensor fusion in which time series data recorded by three different accelerometers are integrated. In addition, the current study generalizes the gear fault diagnostics problem to a multiple fault problem with various severities. The framework of the current paper shows a very clear trend in application of state of the art deep learning techniques as powerful tools in health management of mechanical systems. Also, the CNN's ability in extracting informative features can be further enriched by adding external feature extraction techniques like phase space topology (PST) [17]. Also, this ability can be even further improved by integrating physics. In detail, although self-feature extraction process can explore deeper parameter spaces it may be computationally inefficient. Physics based domain modification can be an option for keeping the network away from exploring infeasible spaces with less possible bias. These ideas being pursued our ongoing studies.

ACKNOWLEDGMENT

This project was supported by a grant from Office of Naval Research (Grant No. N00014-19-1-2070). We are grateful to ONR and Capt. Lynn Petersen for recognizing the importance of this work. We are also grateful to United Technologies Research Center for providing the test data; in particular, we would like to acknowledge the willing help provided by Dr. Zaffir Chaudhry and Dr. Yan Chen.

REFERENCES

[1] A. K. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical systems and signal processing*, vol. 20, no. 7, pp. 1483–1510, 2006.

[2] K. Goebel, M. J. Daigle, A. Saxena, I. Roychoudhury, S. Sankararaman, and J. R. Celaya, *Prognostics: The science of making predictions*, 2017.

[3] E. F. Chad, "Synchronous sampling sideband orders from helical planetary gear sets," Master's thesis, Blacksburg, Virginia,, 1998.

[4] R. B. Randall, "A new method of modelling gear faults," *ASME Journal of Mechanical Design*, vol. 104(2), pp. 259–267, 1982.

[5] M. Serridge, *Ten Crucial Concepts Behind Trustworthy Fault Detection in Machine Condition Monitoring*. Dordrecht: Springer Netherlands, 1990, pp. 729–740. [Online]. Available: https://doi.org/10.1007/978-94-009-1914-3_42

[6] C. Bishop, *Pattern Recognition and Machine Learning*. Springer-Verlag, New York, 2006.

[7] G. Vachtsevanos, F. Lewis, M. Roemer, A. Hess, and B. Wu, Eds., *Intelligent Fault Diagnosis and Prognosis for Engineering Systems*. Wiley, 2006.

[8] W. Wang and P. McFadden, "Application of wavelets to gearbox vibration signals for fault detection," *Journal of sound and vibration*, vol. 192, no. 5, pp. 927–939, 1996.

[9] Z. Peng and F. Chu, "Application of the wavelet transform in machine condition monitoring and fault diagnostics: a review with bibliography," *Mechanical systems and signal processing*, vol. 18, no. 2, pp. 199–221, 2004.

[10] F. Combet and L. Gelman, "An automated methodology for performing time synchronous averaging of a gearbox signal without speed sensor," *Mechanical systems and signal processing*, vol. 21, no. 6, pp. 2590–2606, 2007.

[11] S. Hussain and H. A.Gabbar, "A novel method for real time gear fault detection based on pulse shape analysis," *MechanicalSystemsandSignal-Processing*, vol. 25, p. 1287–1298, 05 2011.

[12] J. Yuan, Z. He, and Y. Zi, "Gear fault detection using customized multiwavelet lifting schemes," *Mechanical Systems and Signal Processing*, vol. 24, no. 5, pp. 1509–1528, 2010.

[13] M. Samadani, T. Haj Mohamad, and C. Nataraj, "Feature Extraction for Bearing Diagnostics Based on the Characterization of Orbit Plots With Orthogonal Functions," *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. Volume 8: 28th Conference on Mechanical Vibration and Noise, 2016.

[14] T. H. Mohamad and C. Nataraj, "Gear fault diagnostics using extended phase space topology," in *Annual conference of the prognostics and health management society*, 2017.

[15] T. H. Mohamad, Y. Chen, Z. Chaudhry, and C. Nataraj, "Gear fault detection using recurrence quantification analysis and support vector machine," *Journal of Software Engineering and Applications*, vol. 11, no. 5, pp. 181–203, 2018.

[16] L. Hong, Y. Qu, J. S. Dhupia, S. Sheng, Y. Tan, and Z. Zhou, "A novel vibration-based fault diagnostic algorithm for gearboxes under speed fluctuations without rotational speed measurement," *Mechanical Systems and Signal Processing*, vol. 94, pp. 14–32, 2017.

[17] T. H. Mohamad, F. Nazari, and C. Nataraj, "A review of phase space topology methods for vibration-based fault diagnostics in nonlinear systems," *Journal of Vibration Engineering & Technologies*, pp. 1–9, 2019.

[18] H. Shao, H. Jiang, X. Zhang, and M. Niu, "Rolling bearing fault diagnosis using an optimization deep belief network," *Measurement Science and Technology*, vol. 26, no. 11, p. 115002, 2015.

[19] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 1, pp. 136–144, 2017.

[20] 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 Processing*, vol. 130, pp. 377–388, 2017.

[21] 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 Transactions on Industrial Electronics*, vol. 63, no. 5, pp. 3137–3147, 2016.

[22] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning*. MIT press Cambridge, 2016, vol. 1, no. 2.

[23] U.-P. Chong *et al.*, "Signal model-based fault detection and diagnosis for induction motors using features of vibration signal in two-dimension domain," *Strojniški vestnik*, vol. 57, no. 9, pp. 655–666, 2011.

[24] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-d convolutional neural networks," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 11, pp. 7067–7075, 2016.

[25] C. Lu, Y. Wang, M. Ragulskis, and Y. Cheng, "Fault diagnosis for rotating machinery: A method based on image processing," *PloS one*, vol. 11, no. 10, p. e0164111, 2016.

[26] M. A. Nielsen, *Neural networks and deep learning*. Determination press San Francisco, CA, 2015, vol. 25.

[27] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA transactions*, vol. 77, pp. 167–178, 2018.

[28] H. C. Cho, J. Knowles, M. S. Fadali, and K. S. Lee, "Fault detection and isolation of induction motors using recurrent neural networks and dynamic bayesian modeling," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 2, pp. 430–437, 2009.

[29] S. Nascimento and M. Valdenegro-Toro, "Modeling and soft-fault diagnosis of underwater thrusters with recurrent neural networks," *IFAC-PapersOnLine*, vol. 51, no. 29, pp. 80–85, 2018.

[30] A. S. Palau, K. Bakliwal, M. H. Dhada, T. Pearce, and A. K. Parlikad, "Recurrent neural networks for real-time distributed collaborative prognostics," in *2018 IEEE international conference on prognostics and health management (ICPHM)*. IEEE, 2018, pp. 1–8.

[31] F. O. Heimes, "Recurrent neural networks for remaining useful life estimation," in *2008 international conference on prognostics and health management*. IEEE, 2008, pp. 1–6.

[32] A. Abbasi, F. Nazari, and C. Nataraj, "Application of long short-term memory neural network to crack propagation prognostics," in *2020 IEEE International Conference on Prognostics and Health Management (ICPHM)*. IEEE, 2020, pp. 1–6.

[33] C. A. K. Kwuimy, P. K. Kankar, Y. Chen, Z. Chaudhry, and C. Nataraj, "Development of Recurrence Analysis for Fault Discrimination in Gears," *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. Volume 8: 27th Conference on Mechanical Vibration and Noise, 2015.