GMM-Aided DNN Bearing Fault Diagnosis Using Sparse Autoencoder Feature Extraction

Andrei Maliuk, Zahoor Ahmad, and Jong-Myon Kim(✉)

University of Ulsan, Ulsan 44610, South Korea
jmkim07@ulsan.ac.kr

Abstract. Deep learning techniques are gaining popularity due to their ability of feature extraction, dimensionality reduction, and classification. However, one of the biggest challenges in bearing fault diagnosis is reliable feature extraction. When using the bearing fault vibration spectrum, the deep neural network (DNN) model can learn the relationships in data that are unrelated to the task. In this work, a simple approach to bearing fault diagnosis using the elimination of unrelated data artifacts for DNN is proposed. The proposed fault diagnosis pipeline is explained and the comparison with popular fault diagnosis methods is performed.

Keywords: Fault diagnosis · Bearing · Vibration · Deep learning

1 Introduction

The rolling-element bearings are the most common industrial equipment which is necessary for rotary motion machines. In electric motors, bearings are mounted on both ends of the rotor shaft to provide smooth rotation. Especially in traction motors, bearings have very high requirements for reliability due to the sensitive areas of applications where a failure of the electric motor can lead to damage of the plant or even human casualties. Electric motors are gaining attraction over internal combustion engines due to their simple construction, however, a mechanical fault in the electric motor can result in catastrophic failure. Specifically, the rolling-element bearing faults are responsible for 45% of all the electric motor faults occurrences [1]. For this reason, condition monitoring of the bearings in electric motors is of primary importance.

The most common types of faults in bearing are outer and inner race faults. When a bearing is in the process of operation, rolling elements pass the damaged areas on the surface of the bearing race and create vibration impulses with a certain rate defined as fundamental defect frequency. In the case of the outer or inner race faults, the frequencies are the ball pass frequency of the outer race (BPFO) and the ball pass frequency of the inner race (BPFI) [2]. The definitions of these frequencies are given in Eqs. (1–2).

$$BPFO = N \times \left(\frac{S_{sh}}{2}\right)\left(1 - \frac{d_r}{D_p}\cos\phi\right), \tag{1}$$

$$BPFI = N \times \left(\frac{S_{sh}}{2}\right)\left(1 + \frac{d_r}{D_p}\cos\phi\right), \tag{2}$$

A. J. Pinho et al. (Eds.): IbPRIA 2022, LNCS 13256, pp. 555–564, 2022.
https://doi.org/10.1007/978-3-031-04881-4_44

where S_{sh} is a shaft speed in RPMs, d_r is the diameter of the rolling element, D_p is the pitch diameter.

In recent years bearing fault diagnosis methods can mostly be described as an application of Machine Learning techniques to the bearing operation history data such as vibration, current or acoustic emission data. The usual pipeline for these methods is signal preprocessing, feature extraction, feature selection, and classification. Signal processing methods commonly used for bearing fault diagnosis are envelope analysis, which has proven to be useful in analyzing low-amplitude high-frequency broadband signals containing bearing fault characteristics, Spectral kurtosis [3], which is capable of extracting the transient components masked by a noisy signal and can find the locations of transients in the frequency domain. Another technique is Wavelet Packet Transform based on Wavelet Transform and is used for signal multi-band filtering and denoising [4].

Following the signal preprocessing step, the feature extraction step allows to represent a time, frequency, or time-frequency domain signal using several statistical parameters of the signal to facilitate the work of the classifier. The feature selection process in its turn is used to choose the best features out of the feature pool and lower the feature space dimensionality. Xie and Zhang performed a comparative study of the most popular feature selection techniques for fault classification tasks [5]. Separately using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) with SVM classifier the performance of two was compared and both significantly reduced data dimensions and showed improvement for classification accuracy, however, LDA had better results due to consideration of interclass and intraclass correspondence.

After the feature pool is constructed, it is provided to a machine learning classification algorithm with labels. Classification algorithms such as Support Vector Machine, Decision trees, and k-Nearest Neighbour are of the highest prevalence in rotating machinery fault diagnosis field these days [6–8].

Gaining huge popularity during the last decade, Deep learning is a technology that can automatically learn representative features from the data and perform classification tasks. It significantly reduced the dependency on manual feature selection. Deep Neural Network architecture can acquire hidden relationships contained in the original data and amplify the meaningful interclass differences while suppressing irrelevant information that can cause interference.

However, working with real data it is impossible to guarantee the absence of discriminant fault-unrelated information, which helps in the improvement of classification performance. This redundant information in the data may have no relationship to the observed phenomena and in other terms can be referred to as artifacts. Thus, expert data preprocessing for intrinsic fault-related information extraction is of primary concern. This is true for both conventional feature extraction and DNN methods.

The diagnosis of the bearing faults traditionally relies on the analysis of the bearing characteristic frequencies which appear at certain stages of the bearing fault formation. These four main stages of bearing fault formation are presented in Fig. 1. In Stage 1 the emerging subsurface microcracks start to appear at the ultrasonic frequencies from 20 kHz to 350 kHz in the range D. The progressing wear results in ringing natural frequencies that start to appear in the range C at 500–2000 Hz together with higher signal amplitude in the range D. Bearing defects become visible in Stage 3 when fundamental

frequencies, accompanied by the well-formed sidebands, start to appear in the frequency range B. At this stage, the fault diagnosis can be performed, and the damaged component can be found. Stage 4 is characterized by decreased amplitude of characteristic frequencies and the presence of broadband random vibration. It is caused by the increased rotor vibration which is a result of the growth of the bearing damaged area. At Stage 4 the bearing has to be replaced immediately.

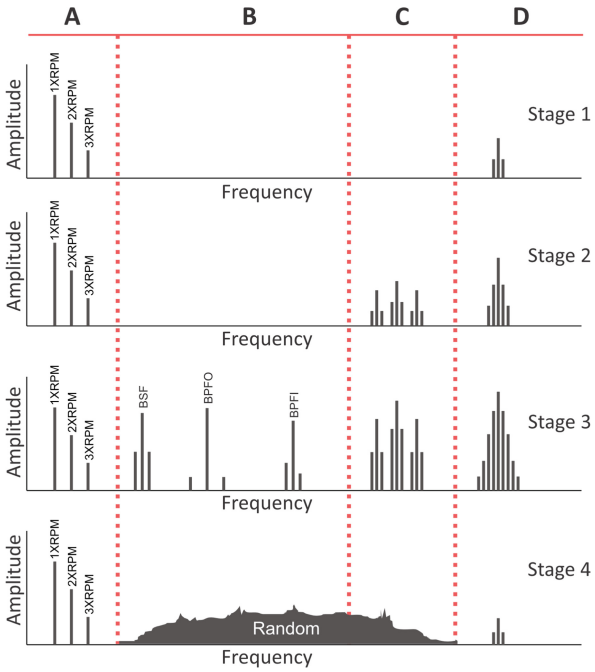

Fig. 1. Bearing fault formation stag.

From the above, it is possible to conclude that for bearing fault diagnosis the amount of useful vibration spectrum information can be scarce. So, for the successful analysis and classification of the bearing fault, this information must be precisely selected to improve the performance of the classification algorithms.

In this paper, a solution for this problem is proposed as a supervised learning model using expert-selected frequency bands. Here, from the full envelope frequency spectrum of vibration signal, fault characteristic frequency bands are isolated and selected using Gaussian windows. It is done to neutralize the chances for DNN to learn the bearing resonance frequencies or normal frequency component information that is present in the envelope spectrum as meaningful information. Feature extraction is performed using sparse autoencoders trained specifically for each selected band and the features from each autoencoder bottleneck layer are used to create the feature pool. After that, the obtained feature pool is provided to Deep Neural Network for classification.

The rest of the paper is organized as follows. The experimental setup and data collection are described in Sect. 2. The proposed method is described in Sect. 3. Section 4 contains results and discussion. The conclusion is made in Sect. 5.

2 Experimental Setup

The data used in this work is obtained from a public dataset made by Kat-DataCenter of the Chair of Design and Drive Technology, Paderborn University, Germany [9]. The test rig used for the experiment is shown in Fig. 2. The test rig has an electric motor, measuring shaft, replaceable bearing module, flywheel, and load motor on one shaft. PMSM 425W drive motor Type SD4CDu8S-009 is controlled by KEB Combivert 07F5E 1D-2B0A industrial inverter at a switching frequency of 16 kHz.

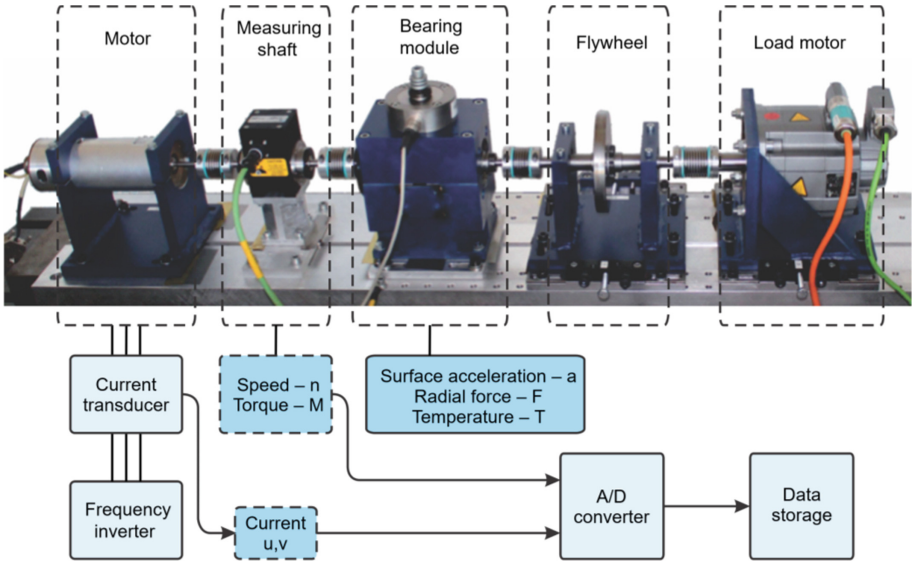

Fig. 2. Modular test rig.

Vibration data was collected using an accelerometer (Model No. 336C04, PCB Piezotronics, Inc.) and a charge amplifier with a 30 kHz Low-Pass filter. The sensor was attached to the top end of the test rig bearing module. The signal was digitalized and saved with a sampling rate of 64 kHz.

In the experiment, healthy and faulty bearings were used. The damages of the bearing were generated by accelerated lifetime tests. In this work in total 17 different bearing signals were used, where 6 bearings are healthy, 5 have a fault in the outer ring and 6 have a fault in the inner ring. The dataset structure used in this work with bearing codes, types of faults and labels are presented in Table 1.

Table 1. Bearing state codes.

Bearing code	Type of fault	Extent of damage	Class label
K001	Healthy	Run-in period > 50 h	0
K002	Healthy	Run-in period = 19 h	0
K003	Healthy	Run-in period = 1 h	0
K004	Healthy	Run-in period = 5 h	0
K005	Healthy	Run-in period = 10 h	0
K006	Healthy	Run-in period = 16 h	0
KA04	Outer ring	1	1
KA15	Outer ring	1	1
KA16	Outer ring	2	1
KA22	Outer ring	1	1
KA30	Outer ring	1	1
KI04	Inner ring	1	2
KI14	Inner ring	1	2
KI16	Inner ring	3	2
KI17	Inner ring	1	2
KI18	Inner ring	2	2
KI21	Inner ring	1	2

Besides the different bearing damage extent, for higher reliability of methods developed with this data, the test rig was operated at 4 different operational conditions with changing rotational speed, load torque, and radial force. The operating parameters of the test rig are shown in Table 2. Bearing vibration data with all operation parameters are included in the dataset used in this research work. Time-domain plots of vibration signal of healthy bearing, bearing with outer ring fault, and bearing with inner ring fault are presented in Fig. 3.

Table 2. Test rig operation parameters.

No.	Rotational speed [rpm]	Load torque [Nm]	Radial force [N]
0	1500	0.7	1000
1	900	0.7	1000
2	1500	0.1	1000
3	1500	0.7	400

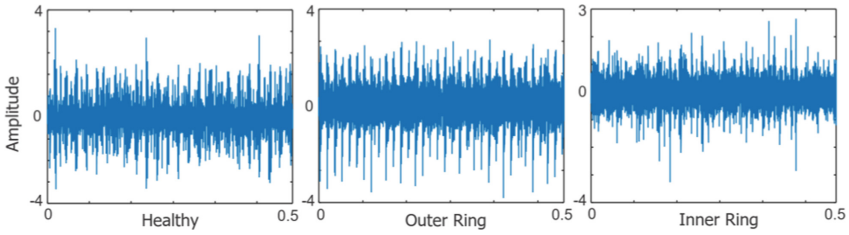

Fig. 3. Time-domain vibration signal plots.

3 Proposed Method

The pipeline of the proposed methodology is shown in Fig. 4. As it can be seen from Fig. 4, in the first step time-domain bearing vibration signals are going through envelope analysis and FFT to obtain the envelope frequency spectrum of the vibration signal. It allows extracting the fault frequencies which are amplitude-modulated to the high-frequency region that cannot be distinguished in the raw frequency spectrum.

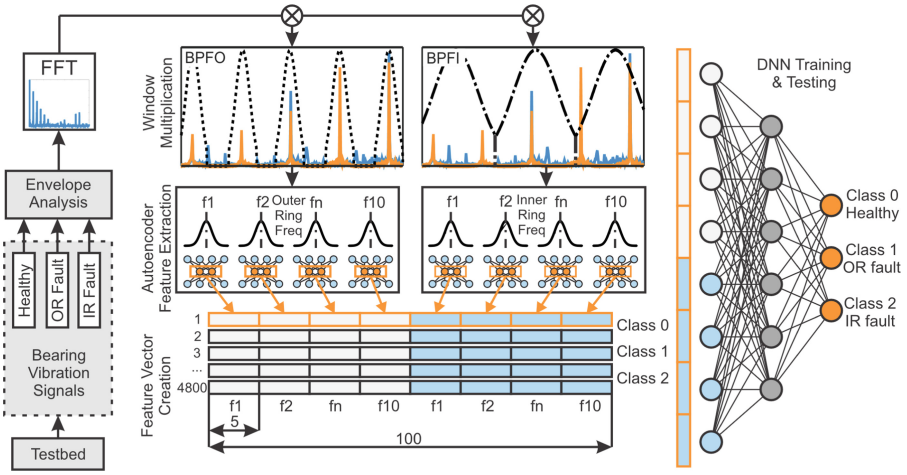

Fig. 4. A pipeline of the proposed method.

Secondly, the vibration envelope spectrum of every one-second sample is multiplied with the Gaussian windows focused on the characteristic frequency components of the outer ring and inner ring faults. The fault bands around inner ring characteristic frequency harmonics tend to be wider than in the case with outer ring faults due to the inner ring rotation at the shaft speed. Therefore, Gaussian window shapes depend on the rotational speed and the expected shapes of the fault signature frequency band. The Gaussian windows means are placed at the harmonics of fault characteristic frequencies which values are calculated using Eqs. 1 and 2. Ten windows are deployed for inner ring fault harmonics selection and in a similar way ten windows are deployed for outer ring harmonics.

Following that, for each of 20 selected bands, a unique three-layer sparse autoencoder is trained in an unsupervised manner for feature extraction with the dimension of the bottleneck layer equal to 5, while the input and output dimensions depend on the width of each frequency band. Consequently, when the feature extraction process is done, the data from each autoencoders bottleneck layer is gathered in the feature vector, where each of the 10 bands obtained from multiplication with IR windows and each of the 10 bands obtained from the multiplication with OR windows is characterized by 5 features from autoencoder. As a result, each envelope spectrum of a 1-s vibration signal sample is now represented by 100 features.

This number of features can be too high and lead to classification performance decrease for conventional Machine Learning techniques, therefore in this work Deep Neural Network architecture was chosen to perform the classification task. A DNN with a 100-neurons input layer, one 25-neurons hidden layer, and a 3-neuron output layer is constructed for classification.

Before training the dataset is balanced among three classes using the stratified sampling technique and is normalized between zero and one. The activation function for the input layer and a hidden layer of the DNN is ReLu and the output layer activation function is SoftMax. Cross entropy loss function with Adam optimizer are used for DNN training.

4 Experimental Results and Analysis

Figure 5 shows network loss and accuracy plot for one training iteration. Here the network was trained for 51 epochs at which the loss and the accuracy tended to converge, and the training process was automatically early stopped with the training accuracy reaching 0.9976, and loss reaching 0.099.

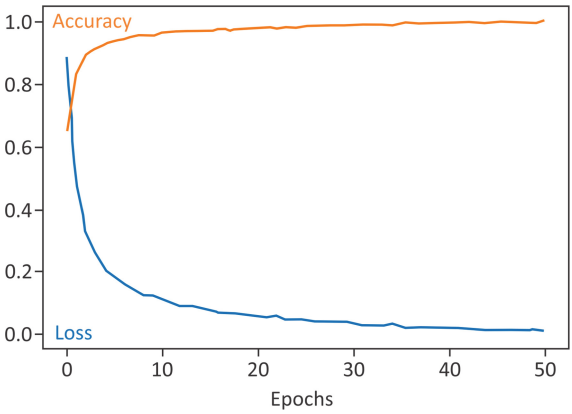

Fig. 5. Loss-accuracy plot for DNN training.

To evaluate the performance of the proposed method, a holdout validation approach was chosen. For this, prior to training the network, the dataset was randomly split into training and testing parts in 80/20 manner with the same proportion of the samples for each class. The network was trained using the training data and then tested using previously unseen testing data. To verify the stable performance of the proposed method, the neural network was trained and tested 10 times using the unique data split for each iteration. Figure 6 shows the confusion matrix obtained by averaging of the results from 10 training and testing iterations. The accuracy of the proposed method is also calculated as the average of 10 iterations.

Performance comparison with other bearing fault diagnosis methods trained and tested using Paderborn bearing vibration data is given in Table 3.

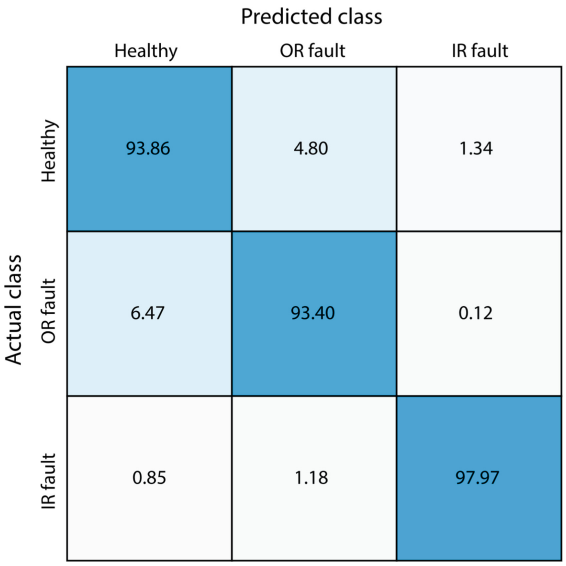

Fig. 6. Confusion matrix.

Table 3. Performance comparison.

Method	Accuracy %
Proposed	95.07
GMM-WBBS [10]	93.10
WPT-BE-MSVM [11]	91.40
WPT-PCA-MSVM [11]	91.46

The comparison shows that the accuracy of the proposed method is generally higher than the accuracy of the comparison methods. This difference can be mainly explained by the usage of the Sparse Autoencoder for feature extraction due to its capability to create an effective characteristic representation of the data.

The Sparse Autoencoder and DNN parts were implemented using TensorFlow with Python 3.8. All signal processing was performed in MATLAB R2021b.

5 Conclusions

In this work, a novel bearing fault diagnosis method was proposed using sparse autoencoder for feature extraction and Deep Neural Network for classification. The workflow of the process was described and on an overall idea of the fault diagnosis method and its structure was given. Method performance was compared to a set of state-of-the-art methods.

The GMM-based spectrum selection for fault characteristic frequency bands was intended to neglect the possible negative effects of the model learning bearing resonance frequencies and bearing normal vibration components along with gear mesh frequency components as related to the bearing fault. Thanks to this, the proposed approach allows reducing the possibility of the negative influence of the data artifacts.

Although the proposed bearing fault diagnosis algorithm had access exclusively to narrowband meaningful data instances selected using Gaussian window series, it showed very high accuracy results compared with techniques that utilized the whole range of the data thanks to the highly effective feature selection performance of Sparse Autoencoders.

Acknowledgements. This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea (No. 20192510102510). This work was also supported by the Technology development Program (S3126818) funded by the Ministry of SMEs and Startups (MSS, Korea).

References

1. Bazurto, A.J., Quispe, E.C., Mendoza, R.C.: Causes and failures classification of industrial electric motor. In: 2016 IEEE ANDESCON, Arequipa, Peru, pp. 1–4, October 2016. https://doi.org/10.1109/ANDESCON.2016.7836190
2. Nandi, A.K., Ahmed, H.: Condition Monitoring with Vibration Signals: Compressive Sampling and Learning Algorithms for Rotating Machines. Wiley-IEEE Press, Hoboken, NJ, USA (2019)
3. Sawalhi, N., Randall, R.B., Endo, H.: The enhancement of fault detection and diagnosis in rolling element bearings using minimum entropy deconvolution combined with spectral kurtosis. Mech. Syst. Sign. Process. **21**(6), 2616–2633 (2007). https://doi.org/10.1016/j.ymssp.2006.12.002
4. Rajeswari, C.: Bearing fault diagnosis using wavelet packet transform, hybrid pso and support vector machine. Procedia Eng. **97**, 12 (2014)
5. Xie, Y., Zhang, T.: A fault diagnosis approach using SVM with data dimension reduction by PCA and LDA method, p. 6

6. Soualhi, A., Medjaher, K., Zerhouni, N.: Bearing health monitoring based on hilbert-huang transform, support vector machine, and regression. IEEE Trans. Instrum. Meas. **64**(1), 52–62 (2015). https://doi.org/10.1109/TIM.2014.2330494
7. Demetgul, M.: Fault diagnosis on production systems with support vector machine and decision trees algorithms. Int. J. Adv. Manuf. Technol. 12 (2013)
8. Lee, C.-Y., Huang, Y.K.-Y., Shen, X., Lee, Y.-C.: Improved weighted k-nearest neighbor based on PSO for wind power system state recognition. Energies **13**(20), 5520 (2020). https://doi.org/10.3390/en13205520
9. Lessmeier, C., Kimotho, J.K., Zimmer, D., Sextro, W.: Condition Monitoring of Bearing Damage in Electromechanical Drive Systems by Using Motor Current Signals of Electric Motors: a Benchmark Data Set for Data-Driven Classification, p. 17 (2016)
10. Maliuk, A.S., Prosvirin, A.E., Ahmad, Z., Kim, C.H., Kim, J.M.: Novel bearing fault diagnosis using Gaussian mixture model-based fault band selection. Sensors **21**(19), 6579 (2021). https://doi.org/10.3390/s21196579
11. Rapur, J.S., Tiwari, R.: Experimental fault diagnosis for known and unseen operating conditions of centrifugal pumps using MSVM and WPT based analyses. Measurement **147**, 106809 (2019). https://doi.org/10.1016/j.measurement.2019.07.037
12. Xu, W.: Research on bearing fault diagnosis base on deep learning. In: 2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD), Chengdu, China, pp. 261–264, May 2021. https://doi.org/10.1109/ICAIBD51990.2021.9459073