Multi-label fault diagnosis based on Convolutional Neural Network and Cyclic Spectral Coherence

Zhuyun Chen[1], Alexandre Mauricio[2,3], Weihua Li[1], Konstantinos Gryllias[2,3]

[1]School of Mechanical and Automotive Engineering, South China University of Technology, 510640, GuangZhou, China
(mezychen@gmail.com; whlee@scut.edu.cn)
[2]Department of Mechanical Engineering, KU Leuven, Celestijnenlaan 300, Leuven, 3001, Belgium
[3]Dynamics of Mechanical and Mechatronic Systems, Flanders Make, Leuven, Belgium
(alex.ricardomauricio@kuleuven.be; konstantinos.gryllias@kuleuven.be)

Abstract

Rotating machines are widely used in manufacturing industry, where sudden failures of key components such as bearings may lead to unexpected breakdown of machines and cause economic loss and human casualties. In addition, machines usually are operating under different working conditions leading to the dynamic changes of fault characteristic, thus presenting big challenges of reliable and accurate fault diagnosis. Data-driven based Deep Learning (DL) fault diagnosis methods are powerful tools to capture hierarchical features from raw input to classify fault patterns by stacking multiple non-linear transformation layers. It constructs and trains deep models relying on huge historical data and requiring less expert knowledge to obtain decision-making. These techniques present effectiveness and advantages in many intelligent fault diagnosis tasks. However, many DL methods are developed for the diagnosis of single fault type without considering the correlations of fault modes. In this paper, we develop a novel fault diagnosis method based on cyclostationary tool and Convolutional Neural Networks (CNN) to tackle these problems. The proposed method presents three characteristics: 1) Cyclic Spectral Coherence (CSCoh) is adopted to provide bearing discriminative patterns for specific type of faults. 2) A fault occurred on the same component (fault pattern), but having different fault severity levels can be regarded a multi-label classification problem, where the fault pattern and the fault severity level are considered to be two specific faults. 3) A novel CNN is constructed by introducing a sigmoid activation output and binary-cross entropy loss function to conduct the multi-label classification task. Specifically, CSCoh is adopted to capture correlation features of periodic phenomenon in the frequency domain. CSCoh is a bi-variable map of two frequency values, which could be used to enhance signatures masked by strong noise, characterizing the fault vibration signals obtained from the rotating machinery under different operating conditions. Then a CNN is developed for multi-label fault classification, which includes fault patterns and fault severity levels identification. The proposed method is evaluated in the experimental study of rolling element bearing fault diagnosis, in which data are collected under different working conditions. The experiment results demonstrate that the proposed method could presents good classification performance and superiority compared with other approaches.

1 Introduction

Rotating machines play important role in manufacturing industry. Rolling element bearings, as the key components of machines easily suffer from the sudden failures due to the long operating under the harsh conditions. This unexpected breakdown of machines may lead to economic loss and even human casualties. Thus, it is essential to develop the condition monitoring techniques for the early and accurate defect detection of such components.

Recently, data driven based-DL intelligent fault diagnosis methods have achieved increasing attentions, due to the powerful feature leaning capability from raw input. DL algorithms refer to deep neural networks, where multiply non-linear transformation layers are stacked to construct the hierarchical architectures. Each layer can be regarded as a data pre-processing unit, where the input is converted into abstract features. With the increase of the layers, the high-level layer can learn more discriminative representations which are helpful for the diagnosis tasks [1]-[2]. The typical DL algorithms, such as Auto Encoder (SAE), Deep Belief Network (DBN), Convolutional Neural Network (CNN) and Long Short-Term Memory network (LSTM), have been applied for the fault diagnosis and detection [3]-[5].

Wang [6] took advantages of the CNN to learn features automatically from the raw vibration signals. Then, Hidden Markov Models (HMM) were employed as strong stability tools to classify rolling element bearing faults. Chen [7] proposed a SAE-DBN method for fault diagnosis by utilizing the multiply sensor information. In the first step, eighteen statistic indexes were extracted from the raw vibration signals, which were fed into sparse autoencoders for feature fusion. Finally, the fused features were input into DBN for fault diagnosis. Janssens [8] extracted the frequency spectra from two vibration accelerometers, and then a 2D CNN with one convolutional layer was designed to learn useful features for bearing fault detection. The proposed method demonstrated its advantages compared to that with hand-crafted features. Sun [9] presented a sparse Deep Stacking Network (DSN) to improve motor diagnosis performance, where the output label of DSN was coded as binary 0 and 1, which leads to more accurate and robust classification results. Chopra [10] adopted SAE for unsupervised features extraction from the engine data, and the majority voting based criteria was used to determine the engine fault type. Althobiani [11] utilized both the Teager Kaiser energy operator and statistical measures to reveal the fault patterns contained in collected signals, and then further adopted DBN for diagnosis of reciprocating compressor. The proposed method provided highly reliable and applicable. Tamilselvan [12] applied DBN for aircraft engine health diagnosis and electric power transformer health diagnosis, which obtained high classification accuracy and presented good generalization performance. In addition, Ince [13] developed a 1D CNN to conduct end-to-end motor fault diagnosis from raw signal input. Jia [14] proposed a normalized CNN for improving the bearing diagnosis performance under imbalanced data by embedding normalized layers and weighted Softmax loss.

From those works mentioned above, different measurements, such as raw time-series signals, frequency spectra, time domain and frequency domain statistical indexes, were adopted as the input of the DL algorithms, which obtained high diagnosis results. However, most of the studies are focusing on distinguishing different fault patterns while ignoring the diagnosis of fault severity levels. The fault severity identification is meaningful to detect the early fault occurrences and guide the decision-making.

Gan [15] adopted the Wavelet Packet Transform (WPT) to extract representative features and then designed a two-layer Hierarchical Diagnosis Network (HDN) for rolling bearing faults. In this study, different DBNs are stacked together, respectively, for classifying fault patterns and fault severity levels. However, multiply DBNs should be designed and trained for meeting the classification tasks. In addition, Wen [16] proposed a Hierarchical Convolutional Neural Network (HCNN), which can be used to recognize the fault patterns and fault severity levels at the same time. In this work, two fully-connected branches are designed at the end of pooling layer to conduct diagnosis. The first branch is used for the diagnosis of fault patterns, and the second branch is used for the recognition of fault severity levels. However, the drawback is that each branch should be constructed with an independently loss function, and a parameter was introduced to balance the trade-off of two losses of HCNN, which requires much expertise knowledge and computational cost for training.

Inspired by previous works, a novel approach integrating CNN and CSCoh is proposed for the multi-label fault classification of rolling element bearings. Firstly, CSCoh is adopted to capture correlation features of periodic phenomenon in the frequency domain, which provides a good discriminative input for CNN. Then a novel CNN is constructed for implementing multi-label fault classification by introducing a novel activation function and loss function. Compared to the other methods, the proposed method only requires a CNN to obtain the fault patterns and fault severity levels at the same time. In addition, compared to other methods, the proposed only replaces the activation function and loss function, without introducing extra parameters and computational cost, which is more suitable for the real industrial applications.

The remaining part of the paper is organized as follows. In Section 2, the theory of the cyclic spectral analysis and the CNN are provided. The diagnosis procedure using the proposed method is introduced in

Section 3. In Section 4, a comprehensive experimental description and the analytical are introduced. Section 5 describes the conclusion of this paper.

2 Introduction to cyclic spectral analysis and Convolutional Neural Networks

2.1 cyclic spectral analysis

In rotating machines, the bearing defects usually generate modulated signals by the characteristic frequencies of the bearings. Such signal, though not periodic, usually can be described as cyclostationary, whose statistical properties vary periodically with time [17]-[18]. The common spectral analysis technique is. Fast Fourier Transform (FFT). It is based on the assumption that the analysed signals are stationary, which can not accurately describe the cyclostationary. To deal with the problems, the cyclic spectral analysis techniques are developed to detecting and identify the hidden periodic behaviour of signals [19]-[20].

For a cyclostationary signal $x(t)$, the second-order moment of cyclostationarity can be defined as an instantaneous AutoCorrelation Function (ACF) with a cyclic T, which is defined as:

$$R_{xx}(t,\tau) = R_{xx}(t+T,\tau) = E\{x(t+\tau/2)x(t-\tau/2)^*\} \tag{1}$$

Then, the second-order statistical descriptor of cyclostationarity, called the Cyclic Spectral Correlation (CSC), can be estimated by implementing the double Fourier transform on the ACF, which is given by:

$$\mathrm{CSC}(\alpha, f)=\int\int R(t,\tau)e^{-j2\pi(\alpha t+f\tau)}dtd\tau \tag{2}$$

The CSC is a bi-variable map of two frequency values. The parameters f and α are called the spectral frequency and cycle frequency, respectively. Contrary to the classic spectral analysis, it provides an additional frequency dimension, revealing both the carriers and their modulations. Spectral frequency f is linked to the carrier component, and the cyclic frequency α is linked to its modulation. It can be observed that for α is equal to zero, it is the classical power spectrum. Furthermore, for α is not equal to zero, it is the power spectrum for that specific cyclic component. Then the Cyclic Spectral Coherence (CSCoh) can be used to measure the degree of correlation between two spectral components given estimated by:

$$\mathrm{CSCoh}(\alpha, f) = \frac{\mathrm{CSC}(\alpha, f)}{\mathrm{CSC}(0, f)\,\mathrm{CSC}(0, f-\alpha)} \tag{3}$$

The CSCoh can be interpreted as the CSC of a whitened signal, which tends to equalize regions with very different energy levels, magnifying weak cyclostationary signals [20].

2.2 Convolutional Neural Network

CNN as a category of multi-layer neural network has achieved great success in areas such as image recognition, image classification, object detections, recognition faces [21]. A typical CNN usually is constructed by the four main operations, convolutional layer, activation layer, pooling or sub sampling layer, and fully-connected layers. Different kinds of layers play different roles. By stacking multiply convolutional, pooling and fully-connected layers, CNN can learn from low-level features to high-order or more abstract features. The layer types considered in this work are introduced.

2.2.1 Convolutional layer

Convolution is the first layer of CNN. The primary purpose of convolution is to extract the features by implementing the convolution operation on raw input data with learned convolution kernel/weights [22]. For each input $\mathbf{x}_i$ and convolution kernel $\mathbf{k}_j$, the output feature map can be calculated as follows,

$$\begin{gathered}\mathbf{y}_{i,j} = f(\mathbf{b}_j + \textstyle\sum_i \mathbf{k}_j * \mathbf{x}_i) \\ f(\mathbf{x})=\max(0,\mathbf{x}),\ \mathbf{x}>0\end{gathered} \tag{4}$$

where, * denotes the convolution operation, **k** and **b** are the value of the kernel and the bias. $f(\ )$ is the activation function, which is usually selected as the Rectified Linear Unit (ReLU) to enable better training of CNN.

2.2.2 Pooling layer

In the second step, the pooling layer is followed, which is used to reduce the spatial dimension and gain computation performance and some translation invariance. This is achieved by summarizing the feature responses in a region of neurons in the previous layer. For an input feature map $\mathbf{x}_i$, the output feature map is obtained,

$$\mathbf{y}_i = \max_{r \times r}(\mathbf{x}_i) \tag{5}$$

where r is the pooling size, and the common pooling operation adopted is known as max-pooling, which slides a window, and gets the maximum on the window as the output.

2.2.3 Fully-connected layer

In the fully-connected layer, the neurons are fully connection to all activations in the previous layer, and a Softmax classifier is usually attached to compute the class score. For the input vector $\mathbf{z}_i$ (i=1, 2, …, N), where N is the number of samples. the Softmax computes the exponential of the given input vector, and the sum of exponential values of all the values in the inputs. Then the ratio of the exponential of the input value and the sum of exponential values is the output of the Softmax function, which can be defined as,

$$Softmax(\mathbf{z}_i) = \frac{\exp(\mathbf{z}_i)}{\sum_j \exp(\mathbf{z}_i)} \tag{6}$$

The output corresponds to the probabilities of each class, and the target class will have the high probability. Softmax will enforce that the total sum of all the probabilities equals to one. That means, in order to increase the probability of a particular class, the module will correspondingly decrease the probability of at least one of the other classes. Thus, the final output will only have one true label. In order to effectively update the neural network, the Cross-Entropy (CE) loss can be adopted by minimizing the loss function between the probability output and the true target class, which is defined as,

$$Loss_{CE} = \sum_{i=1}^{C} \mathbf{y}_i (\log(\hat{\mathbf{y}}_i)) \tag{7}$$

where $\mathbf{y}$ is the true label of the data set, $\hat{\mathbf{y}}$ is the Softmax output, and C is the number of class.

3 The proposed CSCoh-CNN fault diagnosis framework

3.1 The architecture of the proposed CNN

In this section, the architecture of the proposed CNN is designed. Compared to the traditional CNN architecture, the proposed architecture introduces a new activation function in the output layer and a new loss function of CNN.

3.1.1 Sigmoid activation function

In the traditional CNN, the Softmax is usually regarded as the final fully-connected layer to predict the classes. While in the proposed CNN, it is replaced with Sigmoid activation function, which can be defined as,

$$Sigmoid(\mathbf{z}_i) = \frac{1}{1+\exp(-\mathbf{z}_i)} \tag{8}$$

For each value of the Sigmoid input, the Sigmoid function returns an independently real-valued output, which can be used to estimate the true output. For the Softmax output, the high value will have the higher probability than other values. That means for a classification problem, there is only one right class output, the outputs are mutually exclusive. While for the Sigmoid output, since the output are independently, it allows to have high probability for all of the classes, and the high value will have the high probability but not the higher probability. That means, for a multi-label classification, Sigmoid can output multiply correct classes, once a probability of one of the output nodes is above the threshold which is usually set to 0.5.

In order to better explain the differences of Softmax and Sigmoid, a fault diagnosis case is taken for example, presented in figure 1. When a Ball Fault with defect diameters of 14 mil (BF14) of rolling element bearing occurs, it can be observed that the traditional Softmax can correctly predict the BF fault with a probability of 85%. But it can only provide a true class output, which fails to diagnosis the severity level at the same time. On the contrary, for the Sigmoid activation output, it not only can obtain the fault pattern: BF with a probability of 92%, but also the fault severity level which is estimated with a probability value of 84%.

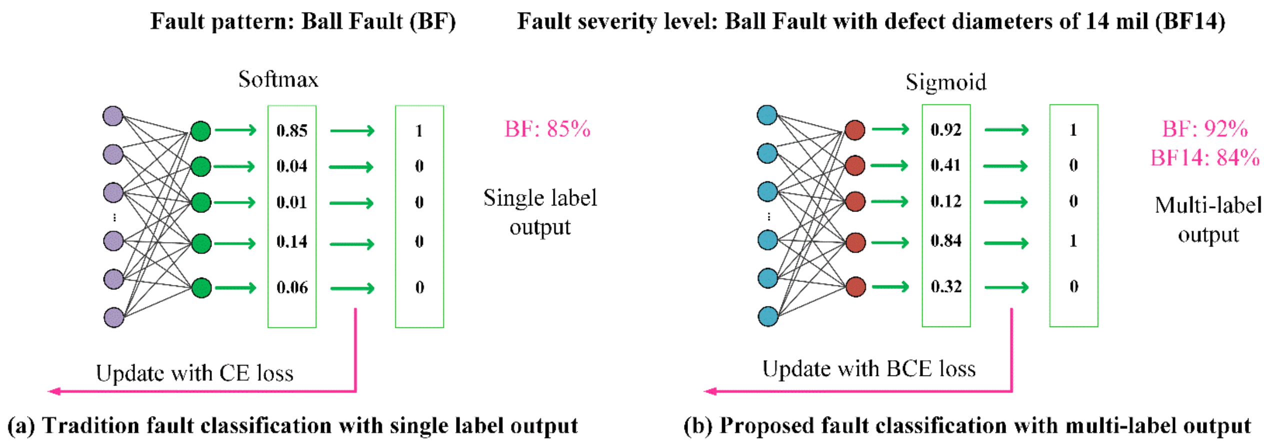


Figure 1: The differences between Softmax and Sigmoid activation function

3.1.2 Binary Cross-Entropy Loss

In order to optimize the CNN with multi-label classification tasks, the Binary Cross-Entropy (BCE) loss is adopted by splitting a multi-label classification problem in C binary classification problems. Unlike the CE loss, BCE is independently for each class, meaning that the loss computed for every output component is not affected by other output class. The loss function can be defined as,

$$Loss_{BCE} = \sum_{i=1}^{C} \mathbf{y}_i \log(\hat{\mathbf{y}}_i) + (1-\mathbf{y}_i)\log(1-\hat{\mathbf{y}}_i) \tag{9}$$

where the update of the weight can be easily implemented by Back-Propagation (BP) algorithm, which is the same as that of the traditional CNN.

3.2 Fault diagnosis based on the proposed CSCoh-CNN

In this section, a multi-label fault diagnosis framework combining CSCoh and CNN is constructed as presented in figure 2. Inspired by the typical LeNet-5 [22], the proposed CNN architecture is designed by stacking two convolutional layers, two pooling layers, one fully-connected layer, and one Sigmoid classification layer.

In the convolutional layer configurations, a small kernel size (3 x 3) is applied in each convolution layer to capture the detail information and reduce the number of parameters. The convolution stride is fixed to 1 pixel. The number of filters is set equal to 6 in the first convolutional layer and the second one is doubled (12) to increase the feature learning capability. In the pooling layers, the max pooling is carried on the input over a (2 x 2) pixel window with stride 2. Therefore, the feature map size is halved to reduce the time complexity. Then the fully-connected architectures is set to 100 neurons. The number of Sigmoid output nodes correspond to the number of predicted classes.

The specifically training procedure can be summarized as follows,

Step 1: The raw vibration data are collected from the test rig, which are pre-processed by cyclic spectral analysis to obtain 2D CSCoh maps. The maps are further downsized to size 112×112 by balancing the computational cost and accuracy.

Step 2: The CNN is constructed by stacking multiply convolutional and pooling layers. Especially, Sigmoid activation function is adopted to predict the independently probability of each class. Accordingly, the BCE is adopted for measuring the distribution between the multi-label output and the target output.

Step 3: CNN network is updated by minimizing the BCE loss to improve the performance of the model in each epoch. The training procedure is the same as that of the traditional CNN.

Step 4: At the testing phase, the testing samples are fed into the trained CNN model to obtain the final diagnosis result.

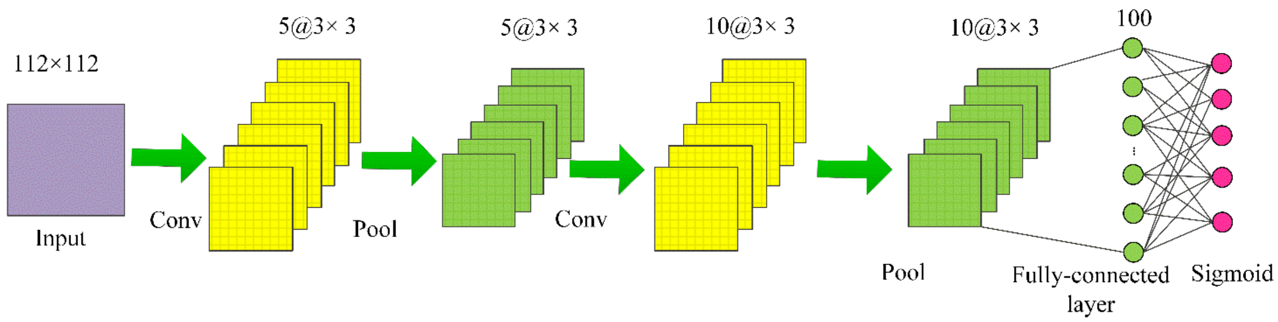


Figure 2: The fault diagnosis framework of the proposed method

4 Experiment verification

4.1 Experiment Setup

The experimental data of rolling element bearings have been acquired from the public bearing data center provided by Case Western Reserve University (CWRU), which is regarded as a benchmark dataset. The test rig is mainly composed of an induction motor, a transducer and a dynamometer. The vibration data are collected near the driving end of motor with a sampling frequency of 48 kHz. The motor bearings were seeded with faults using electro-discharge machining (EDM). In addition to the Normal Condition (NC), bearings with defect diameters of 7 mil, 14 mil and 21 mil have been introduced separately at the inner raceway, ball and out raceway. Each bearing is tested under four different loads (0, 1, 2 and 3 hp). Thus, the faults with two different diagnosis levels ('Level 1' and 'Level 2') can be detected. 'Level 1' means to discriminate the bearing with different fault patterns. While 'Level 2' denotes to further diagnosis the bearing with different severity levels, such as BF7 and BF17 cases, which is more challenge. The detailed description of data is listed in Table 1.

Level 1: Fault pattern	Level 2: Fault severity levels (mil)	Class encoding
Normal Condition (NC)	0	[1,0,0,0,0,0,0,0,0,0,0,0,0]
Ball Fault (BF)	7	[0,1,0,0,1,0,0,0,0,0,0,0,0]
	14	[0,1,0,0,0,1,0,0,0,0,0,0,0]
	21	[0,1,0,0,0,0,1,0,0,0,0,0,0]
Inner race Fault (IF)	7	[0,0,1,0,0,0,0,1,0,0,0,0,0]
	14	[0,0,1,0,0,0,0,0,1,0,0,0,0]
	21	[0,0,1,0,0,0,0,0,0,1,0,0,0]
Outer race Fault (OF)	7	[0,0,0,1,0,0,0,0,0,0,1,0,0]
	14	[0,0,0,1,0,0,0,0,0,0,0,1,0]
	21	[0,0,0,1,0,0,0,0,0,0,0,0,1]

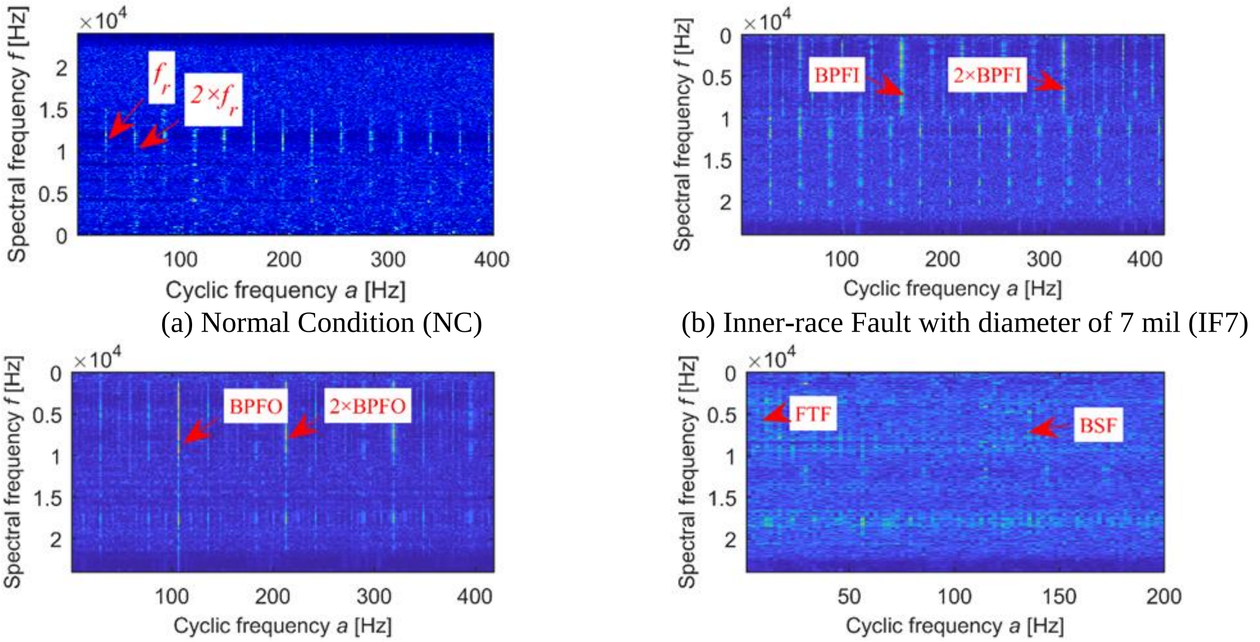
Table 1: Description of the bearing health conditions

4.2 Analysis of the CSCoh 2D maps

In order to obtain the 2D CSCoh maps from the vibration measurements, 24000 data points (time duration of 0.5 seconds) are considered from the time-series signals to form one sample. There are 20 samples obtained from each health condition under every working load. In addition, it should be noted that, due to the limited sampling time, there are only 14 samples obtained for IF14 under the load 1.

When defects of bearings occur, the bearing fundamental fault frequencies can be detected to analyse their dynamic behaviors. In order to validate the effectiveness of CSCoh in revealing the discriminative information of different fault types, four health conditions including the NC, IF with defect diameter of 7 (IF7), OF with defect diameter of 7 (OF7) and BF with defect diameter of 14 (BF14) are presented in figure 3.

It can be seen that the CSCoh maps provide unique representations for given fault types. In figure 3 (a), the fundamental shaft frequency (f_r) and its harmonic presents in the lower frequency are clearly observed, which is consistent with the dynamic behavior of the normal condition. In figure 3 (b) and figure 3 (c), Ball Pass Frequency of Inner-race (BPFI) and the Ball Pass Frequency of Outer-race (BPFO) and its harmonic can be clearly captured respectively, corresponding to the occurrence of the specific faults. It should be noted that in the case of BFs, presented in figure 4 (d), the weak amplitude of the Fundamental Train Frequency (FTF) and the Ball Spin Frequency (BSF) can be detected only in a few of samples of BF14, which reveal the existence of the ball fault. This demonstrates that the proposed is able to provide a good discriminative features when defects of bearings occurred.



(a) Normal Condition (NC) (b) Inner-race Fault with diameter of 7 mil (IF7)

(c) Outer-race Fault with diameter of 7 mil (OF7) (d) Ball Fault with diameter of 7 mil (BF14)

Figure 3: The 2D CSCoh maps of different health conditions

4.3 A Fault diagnosis using the proposed method

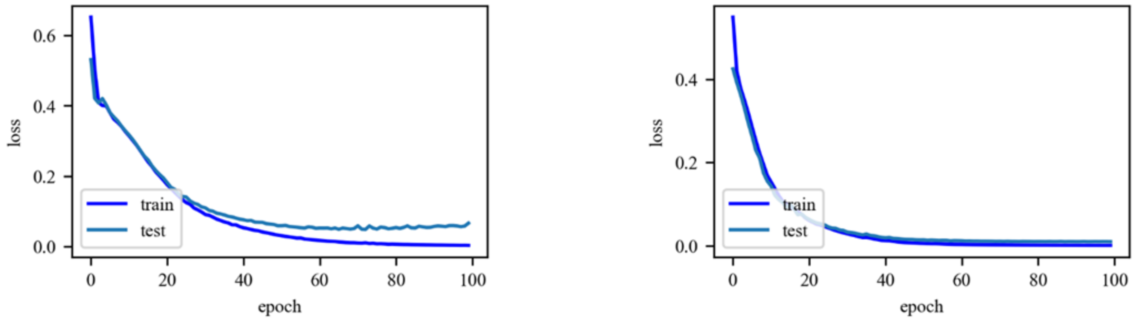
4.3.1 Effect of training sample ratio on classification performance

In order to study the influence of the training sample size on the classification rate, two datasets (dataset A and dataset B) are constructed to evaluate the performance. In dataset A, 20% of the samples are taken as the training data, while the rest for testing. In dataset B, 50% of samples are taken as training, and the rest for testing. Dataset A is constructed to simulate the insufficient training for the network. While dataset B is designed to sufficient training of network.

For CNN training, Adam algorithm is utilized to adjust the network weights with a batch size of 50. The epoch is set to 100, Ten trials are implemented to reduce the random. And the loss curves are shown in figure 4. From the figure 4, it can be seen that the training losses in both datasets are smooth, and remain stable, when they reach a certain number of iterations. It reveals that the models are well trained under the two training data. In the test stage, the loss curve in dataset A decreases slowly than that of the train stage, and it is close to a fixed value, and keep stable. It is possible that CNN with a large number of parameters trained on the insufficient training data suffers from the overfitting problem. By adding more training samples, the loss curve as shown in figure 4(b), are obviously decreasing.

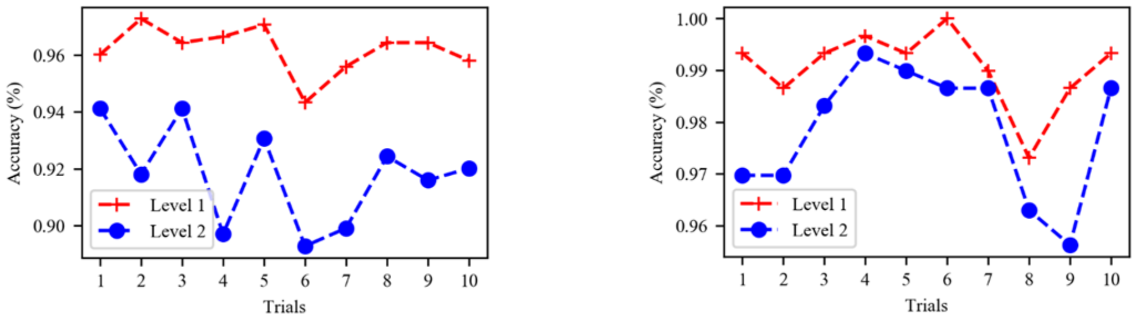
In addition, the diagnosis results of ten trials are presented in figure 5. “Level 1” denotes the classification accuracy of the fault patterns, where NC, BF, IF and BF are correctly recognized. “Level 2” reflects the total recognition accuracy, where the fault patterns and fault severity levels are all correctly classified.

From the results, it can be observed that, the results of all ten trials present relative high accuracy in both datasets. In addition, ‘Level 1’ is much higher than ‘Level 2’, since the former only needs to diagnosis the specific fault patterns, while the later requires to discriminative the fault severity levels of each fault patterns. Moreover, CNN with dataset A is much lower than that of dataset B, especially in ‘Level 2’. This is because that ‘Level 2’ contains more discriminative classes, which is more difficult for diagnosis. Therefore, the results of dataset B is able to obtain better classification performance compared to that of dataset A, since more of the training samples are contained.



(a) Loss curve of CNN using dataset A (b) Loss curve of CNN using dataset B

Figure 4: The loss curve of CNN with different datasets



(a) Results of 10 trials using dataset A (b) Results of 10 trials using dataset B

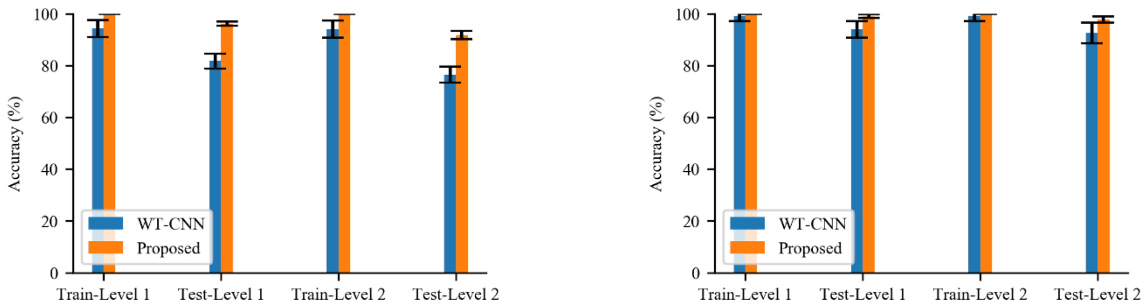
Figure 5: The results of ten trials of different datasets

4.3.2 Comparison with other methods

Wavelet Transform (WT), which decomposes the signals into wavelets, is usually considered as an effective tool for pre-processing nonstationary and transient signals [23]. The two-dimensional time frequency representation of WT provides a high resolution in both the time-domain and frequency-domain, which provides good information about the health conditions of rotating machinery.

In this section, a comparison of the WT scalograms and the CSCoh maps is carried out. Morelet wavelet basis is adopted to implement the time-frequency transformation. In order to provide a relative fair comparison, all the pre-processing data are fed into CNN for training, and the results are conducted with ten trials. Final results are averaged. The training and testing accuracy of different methods are shown in figure 6.

It can be seen that, the training accuracies of the proposed method in both datasets are 100.0%. While the training accuracies of WT are relative lower, and present larger standard deviations. In addition, the testing accuracy, especially, in ‘Level 2’, the proposed method also obviously performs better than that of WT-CNN.



(a) Result comparison of different methods using dataset A (b) Result comparison of different methods using dataset B

Figure 6: Result comparison of different methods

Furthermore, the testing accuracy of each types under different severity levels are further listed in Table 2 and Table 3, respectively. It can be observed that WT-CNN achieves average testing accuracy of 76.7% and 92.8%, in dataset A and dataset B, respectively. On the contrast, the proposed method achieves average testing accuracy of 92.2% and 97.9%, respectively, which is better that that of the WT-CNN. In addition, the diagnosis accuracies of the BF7 and the BF14 are obviously lower that of the other fault types. This is because that the characteristic frequencies in those cases are not obvious in the 2D CSCoh maps, which makes CNN difficulty to obtain good classification performance. From the analysis, it can be concluded that the proposed method is effective in extracting discriminative features and conducting the multi-label classification tasks.

Methods	Accuracy (%) of each fault severity level using dataset A									
	NC	BF7	BF14	BF21	IF7	IF14	IF21	IF7	IF14	IF21	AVG
WT-CNN	100	52.7	71.7	64.2	100.0	59.8	99.5	93.4	39.5	86.5	76.7
Proposed	100	100.0	79.0	58.4	100	96.1	95.9	100.0	96.4	96.5	92.2

Table 2: The average testing accuracy of each fault severity level using dataset A

Methods	Accuracy (%) of each fault severity level using dataset B										
	NC	BF7	BF14	BF21	IF7	IF14	IF21	IF7	IF14	IF21	AVG
WT-CNN	100.0	95.7	85.0	82.3	100.0	92.0	100	95.4	79.5	98.0	92.8
Proposed	100.0	100.0	95.7	89.6	100.0	96.5	98.3	100.0	99.0	99.4	97.9

Table 3: The average testing accuracy of each fault severity level using dataset B

Conclusion

In this work, a new DL-based fault diagnosis framework, combining CSCoh and CNN is proposed for multi-label fault classification. Firstly, CSCoh is considered, as a pre-processing step, to reveal the fault nature of each fault types. Then, a novel CNN is constructed for conducting fault classification with multiply labels by introducing the Sigmoid activation function and BCE loss function. The proposed method is verified on the data collected from the CWRU motor bearing test rig. Two different datasets including the insufficient training and sufficient training data are designed to evaluate the effectiveness of the methodology. It has been demonstrated that the proposed method not only achieves high classification performance, but also presents better generalization performance compared to WT-CNN fault diagnosis method.

References

[1] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, *Deep learning and its applications to machine health monitoring*, Mech. Syst. Signal Process. 115 (2019) 213-237.

[2] S. Khan, T. Yairi, *A review on the application of deep learning in system health management*, Mech. Syst. Signal Process. 107 (2018) 241-265.

[3] F. Jia, Y. Lei, J. Lin, X. Zhou, 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. 72-73 (2016) 303-315.

[4] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, M. Modarres, *Deep learning enabled fault diagnosis using time-frequency image analysis of rolling element bearings*. Shock and Vibration, vol. 2017, Oct. 2017.

[5] B. Zhang, S. Zhang, W. Li, *Bearing performance degradation assessment using long short-term memory recurrent network*, Computers in Industry. 106 (2019) 14-29.

[6] S. Wang, J. Xiang, Y. Zhong, Y. Zhou, *Convolutional neural network-based hidden Markov models for rolling element bearing fault identification*, Knowledge-Based Systems, 144 (2018) 65-76.

[7] Z. Chen, W. Li, *Multisensor Feature Fusion for Bearing Fault Diagnosis Using Sparse Autoencoder and Deep Belief Network*, Ieee T Instrum Meas. 66 (2017) 1693-1702

[8] O. Janssens, V. Slavkovikj, B. Vervisch, K. Stockman, M. Loccufier, S. Verstockt, R. Van de Walle, S. Van Hoecke, *Convolutional Neural Network Based Fault Detection for Rotating Machinery*, J. Sound Vibrat. 377 (2016) 331-345.

[9] C. Sun, M. Ma, Z. Zhao, X. Chen, *Sparse Deep Stacking Network for Fault Diagnosis of Motor*, IEEE Transactions on Industrial Informatics. 14 (2018) 3261-3270.

[10] P. Chopra and S. K. Yadav, *Fault detection and classification by unsupervised feature extraction and dimensionality reduction*, Complex Intell. Syst., vol. 1, no. 1, pp. 25–33, Jan. 2015.

[11] F. Althobiani and A. Ball, *An approach to fault diagnosis of reciprocating compressor valves using Teager–Kaiser energy operator and deep belief networks*, Expert Syst. Appl., vol. 41, no. 9, pp. 4113–4122, Dec. 2013.

[12] [P. Tamilselvan and P.Wang, *Failure diagnosis using deep belief learning based health state classification*, Rel. Eng. Syst. Safety, vol. 115, pp. 124–135, Jul. 2013.

[13] Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, *Real-time 751 motor fault detection by 1-d convolutional neural networks*, IEEE Trans. 752 Ind. Electron., vol. 63, no. 11, pp. 7067–7075, Nov. 2016

[14] F. Jia, Y. Lei, N. Lu, S. Xing, *Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization*, Mech. Syst. Signal Process. 110 (2018) 349-367.

[15] M. Gan, C. Wang, C.a. Zhu, *Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings*, Mech. Syst. Signal Process. 72-73 (2016) 92-104.

[16] L. Wen, X. Li, L. Gao, Y. Zhang, *A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method*, IEEE Trans. Industr. Electron. 65 (2018) 5990-5998.

[17] P. Borghesani, J. Antoni, *A faster algorithm for the calculation of the fast spectral correlation*, Mech. Syst. Signal Process. 111 (2018) 113-118.

[18] A. Mauricio, J., Qi, K. Gryllias, 2019. *Vibration-Based Condition Monitoring of Wind Turbine Gearboxes Based on Cyclostationary Analysis*. Journal of Engineering for Gas Turbines and Power, 141(3), p.031026.

[19] K. Gryllias, A. Mauricio, J. Qi, 2018, September. *Advanced cyclostationary-based analysis for condition monitoring of complex systems*. In 2018 26th European Signal Processing Conference (EUSIPCO) (pp. 385-389). IEEE.

[20] A. Mauricio, J. Qi, W. Smith, R. Randall, K. Gryllias, *Vibration Based Condition Monitoring of Planetary Gearboxes Operating Under Speed Varying Operating Conditions Based on Cyclo-non-stationary Analysis*, 61 (2019) 265-279.

[21] Y. LeCun, G. Bengio, Hinton, *Deep learning*, Nature, 521 (2015) 436-444.

[22] Y. LeCun, L. Bottou, Y. Bengio, Y, P. Haffner, P. (1998). *Gradient-based learning applied to document recognition*. Proceedings of the IEEE, 86(11), 2278-2324.

[23] Yan, R.X. Gao, X. Chen, *Wavelets for fault diagnosis of rotary machines: A review with applications*, Signal Processing, 96 (2014) 1-15.