CHAPTER 4

Compound Fault Prediction of Rolling Bearing

4.1 Introduction

Catastrophic failure of mechanical systems due to faults occurring on the rolling bearing is still a great challenge. The operating state of rolling bearing significantly affects the accuracy, reliability, and useful life of any mechanical system. It is also important for plant safety and production efficiency [Hongbin *et al.,* 1995]. Thus, the health monitoring and fault diagnosis of a rolling bearing is a big task. Vibration signal detection is an effective method for fault diagnosis of rolling bearings. However, various factors, such as unknown source signals, the complexity of the transmission channel, restriction of sensor installation location and experimental cost problem, etc., have brought certain difficulties to the equipment health monitoring and fault diagnosis.Ideally, it is better if a vibration signal contains only one defect when it is measured by an acceleration sensor under low-noise condition. But in reality, these faults are of multiple types and are compounded in nature. In real situations, a fault signal in functional part appears as a spike sequence, such as inner-race fault, outer- race fault, and roller fault of rolling bearings. Hence the vast majority of these signals obey non-Gaussian

distribution. A compound fault signal usually consists of multiple characteristic signals and strong confusion noise, which makes it a tough task to separate weak fault signals from them.

To resolve the problem arisen above and to improve the detection of fault types and the health monitoring of rotating mechanical systems operating state, it is important to segregate the compound faults from acquired multimedia (acceleration sensor or acoustic) signals. The signal processing methods, such as Fast Fourier Transform (FFT) and Wavelet Transform (WT) has been used in such situations [Lou *et al.,* 2004]. Huang *et al.* (1998) proposed EMD as an adaptive and efficient method to decompose nonlinear and non-stationary signals into Intrinsic Mode Functions (IMF). A combined Independent Component Analysis (ICA) and Instantaneous Frequency (IF) method to detect simultaneous machinery faults using sound mixture emitted by machines have been proposed by Atmaja *et al.* (2009). Arifiant *et al.* (2011) for remote condition monitoring. The Sparse Support Vector Machine (SSVM) and Kernel Independent Component Analysis (KICA) were proposed as new approaches for complex industrial process monitoring and fault diagnosis by Ma *et al.* (2013).

Combined Mode Function (CMF) technique has been used to combine the nearby IMF to get high-frequency components and low-frequency components by Grasso *et al.,* (2016). The hybrid systems are better for compound fault diagnosis. To resolve the compound fault diagnosis problem of rolling bearings, Ensemble Empirical Mode Decomposition (EEMD) method [Huang *et al.,* 2009] along with ICA technique, has been used to some degree of success [Wang *et al.,* 2014]. However, if the frequency components in the signal are much complicated, it affects the decomposition results [Peng *et al.,*2005]. Therefore, CMF is useful as the pre-filter of EEMD to improve the effectiveness and accuracy of EEMD decomposition.

Machine learning methods like Artificial Neural Networks (ANN) based approach has been used for the detection of faults in machines quite earlier due to high feature extraction capability

[Stefano *et al.,* 1994]. Deep learning methods have recently proved their worth in handling big sized data. Convolution Neural Networks (CNN) have accomplished the state-of-the-art performances in recent past [Ciresan *et al.,* 2010], [Scherer *et al.,* 2010], [Krizhevsky *et al.,* 2010]. CNNs are deep neural networks that have both alternating convolution and sub-sampling layers. Convolution layers model the cells in the human visual cortex [Wiesel *et al.,* 1959]. CNN's are capable of performing automatic feature extraction and feature selection.

To resolve the compound fault diagnosis problem of rolling bearings by separation of multimedia signals (obtained from acoustic or acceleration sensors), ensemble empirical mode decomposition (EEMD) method along with some classifier (like independent component analysis (ICA) technique) has been used to some degree of success [Wang *et al.,* 2014]. But they are not found capable of detecting difficult faults existing on small balls of the bearing. In order to solve this problem, we are going to propose a new method based on use of Combined Mode Functions (CMF) for selecting the intrinsic mode functions (IMFs) instead of the maximum cross-correlation coefficient based EEMD technique, sandwiched with, Convolution Neural Networks (CNN), which are deep neural nets, used as fault classifiers. This composite CNN-CMF-EEMD method overcomes the deficiencies of other approaches, such as the inability to learn the complex non-linear relationships in fault diagnosis issues and fine compound faults like those occurring on small balls of the bearing. The difficult compound faults can be separated effectively by executing CNN-CMF-EEMD method, which extracts fault features easily and identifies them more clearly.

Bearing faults cause vibration at fault related frequencies. The frequencies corresponding to different bearing faults can be determined, if, bearing dimensions and shaft rotation are known. Ball fault frequency f_{BD} , is given by

$$f_{BD} = \frac{PD}{2} f_S \left(1 - \left(\frac{BD}{PD}\right)^2 cos^2(C)\right) \tag{4.1}$$

Outer race fault frequency, f_O, is given by

$$f_O = \frac{n}{2} f_S \left(1 - \frac{BD}{PD} cos(C)\right) \tag{4.2}$$

Inner race fault frequency f_I, is expressed as

$$f_I = \frac{n}{2} f_S \left(1 + \frac{BD}{PD} cos(C)\right) \tag{4.3}$$

Where f_S is the rotor speed in revolutions per second, BD is ball diameter, PD is pitch diameter, n is the number of balls, and the angle C is the contact angle which is zero for ball bearings.

4.2 The Basic Theory of EEMD, CMF, CNN, and Proposed Method

4.2.1 Empirical Mode Decomposition (EMD)

EMD is an efficient technique proposed by Huang to decompose nonlinear and nonstationary signals into intrinsic mode functions (IMF). The IMFs are a representation of the natural oscillatory mode in the given signal and behave like basis functions.

The following two conditions are satisfied by an IMF function:

(1) The number of zero-crossings and number of extrema must be either equal or it may differ by at most by one in the whole dataset.

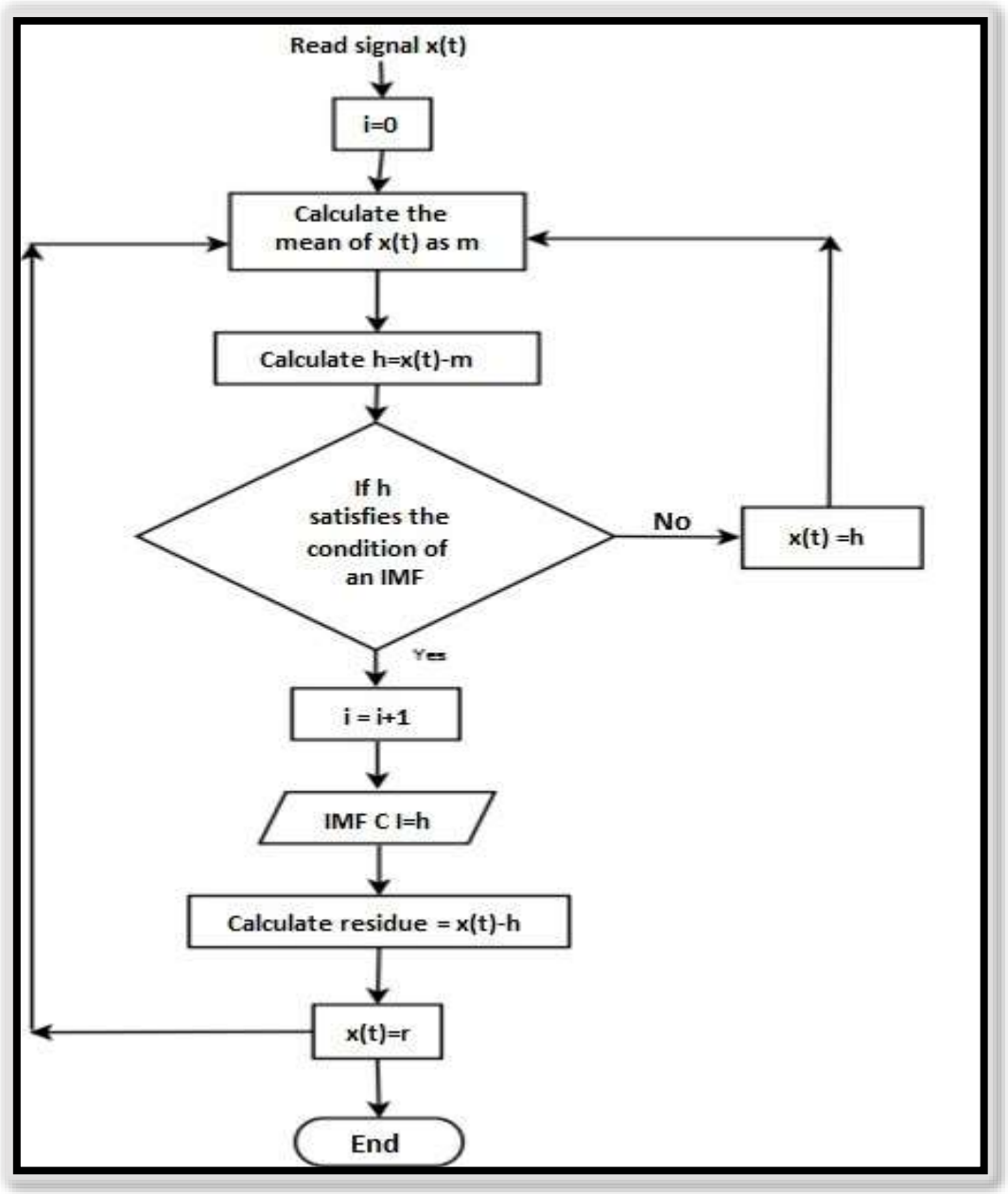

Figure 4.1: Flow chart of EMD

(2) The mean value of the envelope defined by the local minima local maxima is zero at any point

4.2.2 Ensemble Empirical Mode Decomposition

The principle of EEMD method utilizes the statistical characteristics of uniform frequency distribution. White noise is added to the original signal to make it continuous in different scales and, to avoid mode mixing. To avoid mode mixing, the EEMD method is implemented in following steps: First, a white noise of uniform scale and constant amplitude standard deviation is added to the original signal [Huang *et al.*, 2009]. Secondly, IMFs of ensemble are calculated as the final results of EEMD. The added white noise series present a uniform reference frame in the time–frequency and time-scale space for signals of comparable scales to collate in one IMF and then cancel itself out (via ensemble averaging) after serving its purpose; therefore, it significantly reduces the chance of mode mixing and represents a substantial improvement over the original EMD. The effect of the added white noise can be controlled according to the well-established statistical rule given as in equation (4.4) [Peng *et al.*, 2005].

$$\in_n = \frac{\in}{\sqrt{N}} \tag{4.4}$$

Where N is the number of ensemble members, $\in$ is the final standard deviation of error which is the difference between the input signal and the corresponding IMFs. In practice, the number of ensemble members is often set to 100 and the standard deviation of white noise series is set to 0.1 or 0.2.

4.2.3 Combined Mode Functions

The proposed approach is aimed at automatically reducing the n IMFs into a number K on of CMFs that are expected to better represent the multi-scale content of the multimedia signal.

The proposed approach for CMF computation involves four consecutive steps:

(1) Preliminary computation of sequential CMFs, denoted by $C_{s_k}(t)$, k=1… n,

(2) Computation of a dissimilarity index to determine a possible separation of IMFs into fewer CMFs, denoted by $C_{s_k}^*(t)$, k=1…, K^*,

(3) Iterative decomposition into different numbers $K^*$of CMFs, and

(4) Determination of the optimal number K<n of final CMFs,$C_{s_k}^*(t)$.

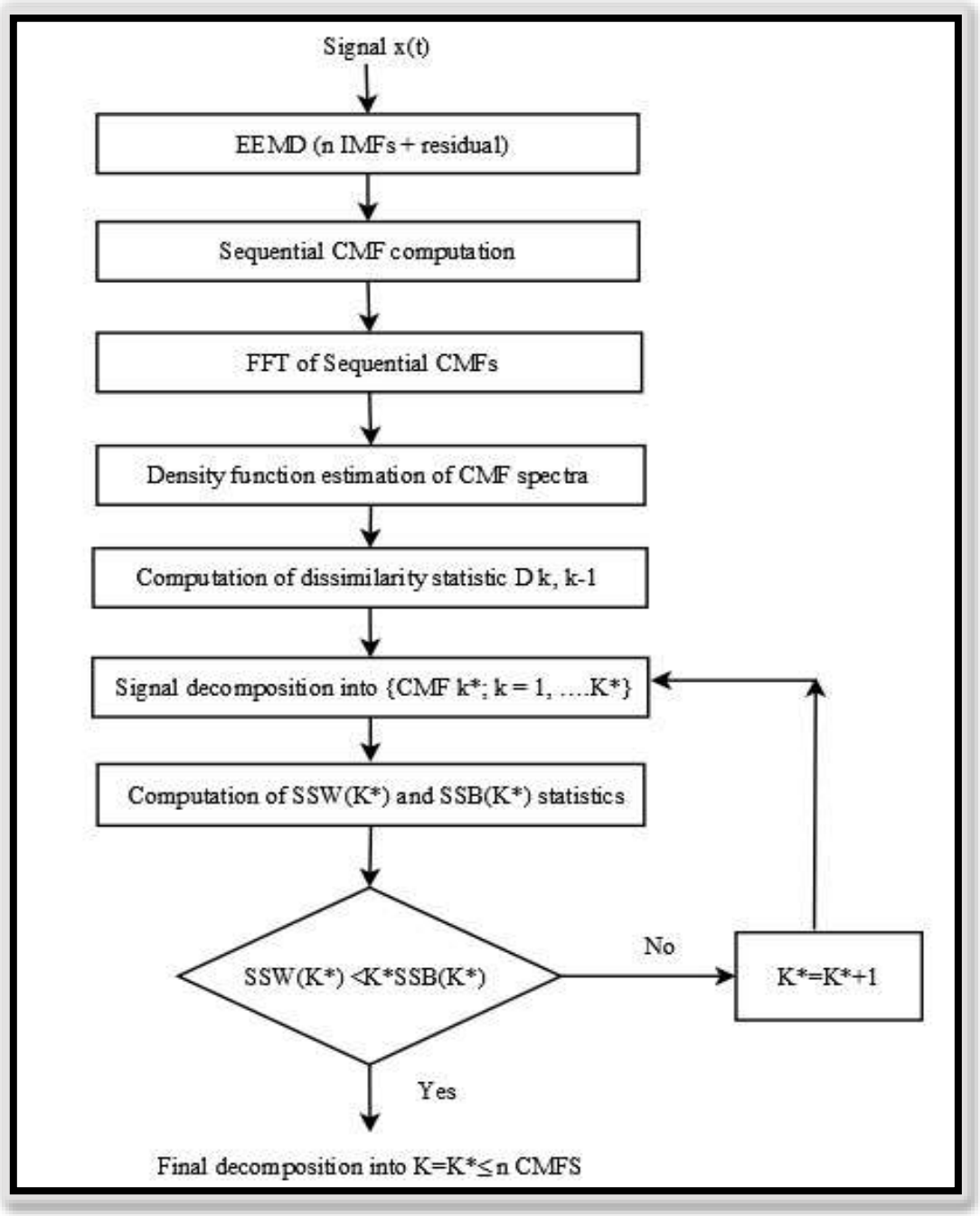

Figure 4.2: Flow chart of CMF

4.2.4 The CNN Theory

Convolutional Neural Network (CNN) consists of one or more convolutional layers (generally, with a sub-sampling step) and then followed by fully connected layers as in a standard multilayer neural network. A CNN consists of some convolutional and subsampling layers which are followed by fully connected layers. Then the input to a convolutional layer is an m × m × r image where r is the number of multimedia channels, which for RGB image has r=3. The convolutional layer will have k filters (or kernels) of size n × n × q where, n is smaller than the dimension of the image (m) and q can either be the same as the number of channels r or smaller and may vary for each kernel. Each map is then sub-sampled typically max pooling over p x p regions with p ranges between 2 to 5 for smaller and larger inputs respectively. The figure 4.3 illustrates a full layer in a CNN consisting of convolutional and sub-sampling sub-layers.

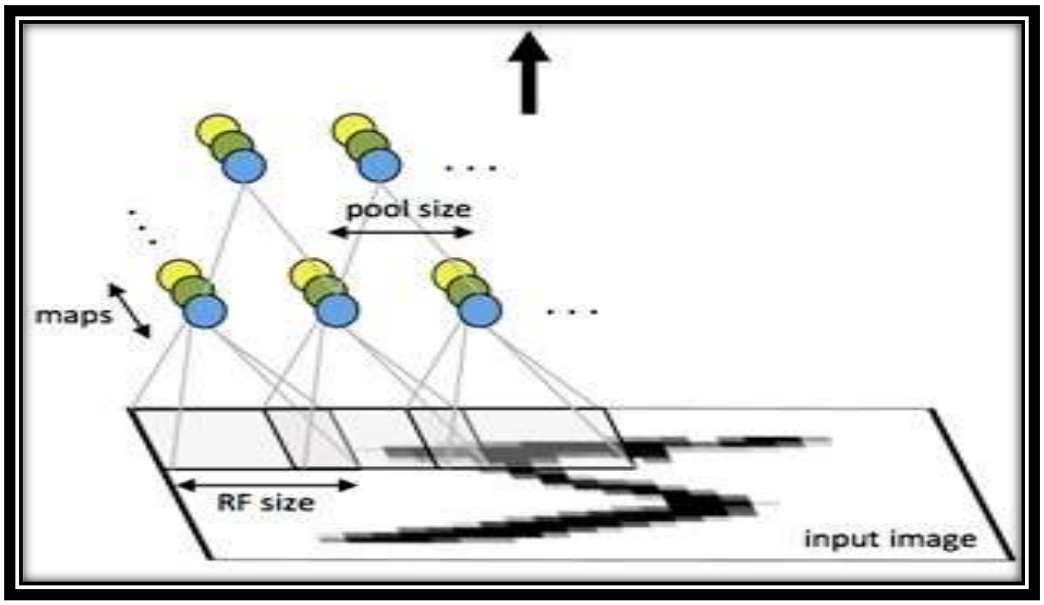

Figure 4.3: First layer of a convolutional neural network with pooling

Let δ^{l+1} be the error term for the $(l+1)^{st}$ layer in the network having a cost function $J(W, b; x, y)$ where (W, b) are the parameters and (x, y) are the data for training and label pairs. If the l^{th} layer is densely connected to the $(l+1)^{st}$ layer, then the error for the l-*th* layer is computed as

$$\delta^{(l)} = ((w^{(l)})^{T}\delta^{(l+1)}).f'(z^{(l)}) \qquad (4.5)$$

Where $(f'(z^{(l)})$ is derivative of the activation function

and the gradients are

$$\nabla_{W^{(l)}}J(W, b; x, y) = \delta^{(l+1)}(a^{(l)})^{T} \qquad (4.6)$$

If the l^{th} layer is a convolutional and subsampling layer, then the error is propagated through as

$$\delta_k^{(l)} = \text{unsample}\left(\left(W_k^{(l)}\right)^{T}\delta_k^{(l+1)}\right).f'(z_k^{(l)}) \qquad (4.7)$$

Where k indexes the filter number, and $f'(z_k^{(l)})$ is the derivative of the activation function.

Lastly, to calculate the gradient w.r.t to the filter maps, $\delta_k^{(l)}$ the same way we flip the filters in convolution layer

$$\nabla_{W_k^{(l)}}J(W, b; x, y) = \sum_{i=1}^{m} a_i^{(l)}\, \text{rot90}(\delta_k^{(l+1)}, 2) \qquad (4.8)$$

$$\nabla_{b_k^{(l)}}J(W, b; x, y) = \sum_{a,b}(\delta_k^{(l+1)})(\delta_k^{(l+1)})_{a,b} \qquad (4.9)$$

Where $\nabla_{W_k^{(l)}}$ is gradient of parameter W with respect to k^{th} filter, $\nabla_{b_k^{(l)}}$ is gradient of parameter b with respect to k^{th} filter, $a^{(l)}$ is the input to the l^{th} layer. The operation $a_i^{(l)} * (\delta_k^{(l+1)})$is the "valid" convolution between i^{th} input in the l^{th} layer and the error with respect to the k^{th}filter.

4.3 Data Description

4.3.1 The Training Data

The bearing data used here are provided by the Case Western Reserve University (CWRU) [Loparo, K.A *et al.*,2004]. The bearing data set was obtained from the experimental setup: (1) under normal condition (N), (2) with outer race fault (OF), (3) with inner race fault (IF) and (4) with roller fault (RF). The faults were introduced into the drive-end bearing of the motor with fault diameters of 0.18 mm, 0.36 mm and 0.54 mm, respectively. The detailed description of the datasets is shown in Table 4.1.

The designed CNN has five layers, in which the unit number of the input layer is determined by the dimension of the samples, the unit number of the hidden layers is 600, and the unit number of the output layer is determined by the number of the health conditions which are ten here. We convert them into four classes by merging the different sizes of the same fault type into one class. The active functions of the CNN are hyperbolic.

The weights of the CNN are initialized randomly, and the biases are initialized to zero.
The maximum training epoch is 20; the learning rate is 0.0001, and the momentum is 0.9.

4.3.2 The Testing Data

We take the compound faults of bearing roller and outer-race as the research object, and artificially made flaws by a wire-cutting machine for the tests of fault diagnosis. The sampling frequency is 100 kHz, and the sampling time is 10 s, and the rotating speed of a machine is 900 rpm. The bearing data used here are provided by the PloS One [Wang, H.*et al.*,2014]

Table 4.1: Data description

Datasets	Fault type	Load (hp)	The number of samples	Fault diameter (mm)	Classification label
A/B/C/D	Normal	1/2/3/1-3	200/200/200/600	0	1
	Roller		200/200/200/600	0.18	2
	Roller		200/200/200/600	0.36	3
	Roller		200/200/200/600	0.54	4
	Inner		200/200/200/600	0.18	5
	Inner		200/200/200/600	0.36	6
	Inner		200/200/200/600	0.54	7
	Outer		200/200/200/600	0.18	8
	Outer		200/200/200/600	0.36	9
	Outer		200/200/200/600	0.54	10

Table 4.2: Different fault classes

Class/fault	**Normal**	**Roller fault**	**Inner fault**	**Outer fault**
Level	0	1	2	3

Figure 4.4: The experiment set-up

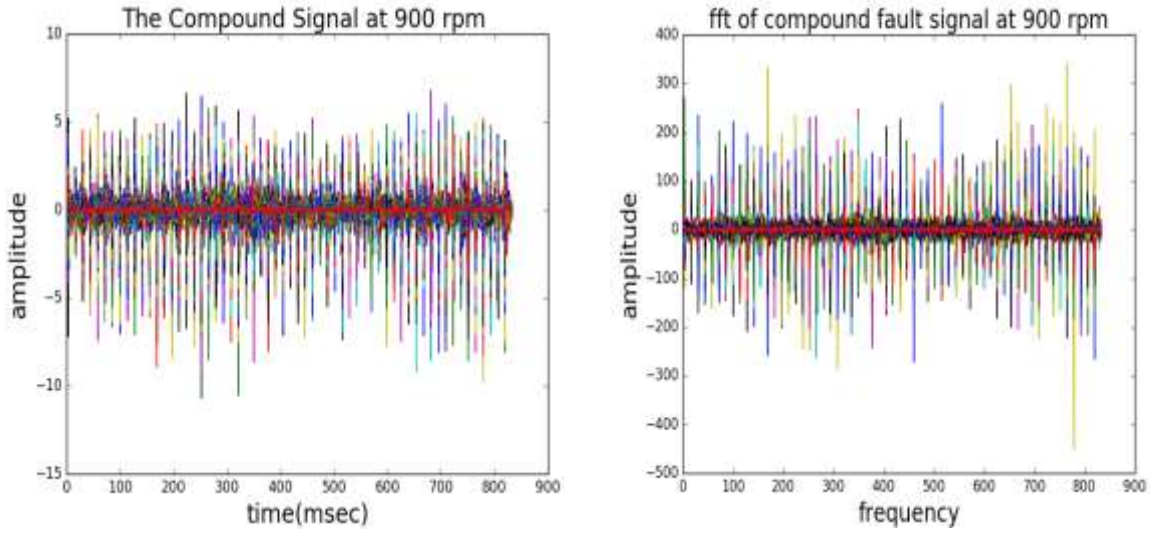

Figure 4.5: The compound fault signal

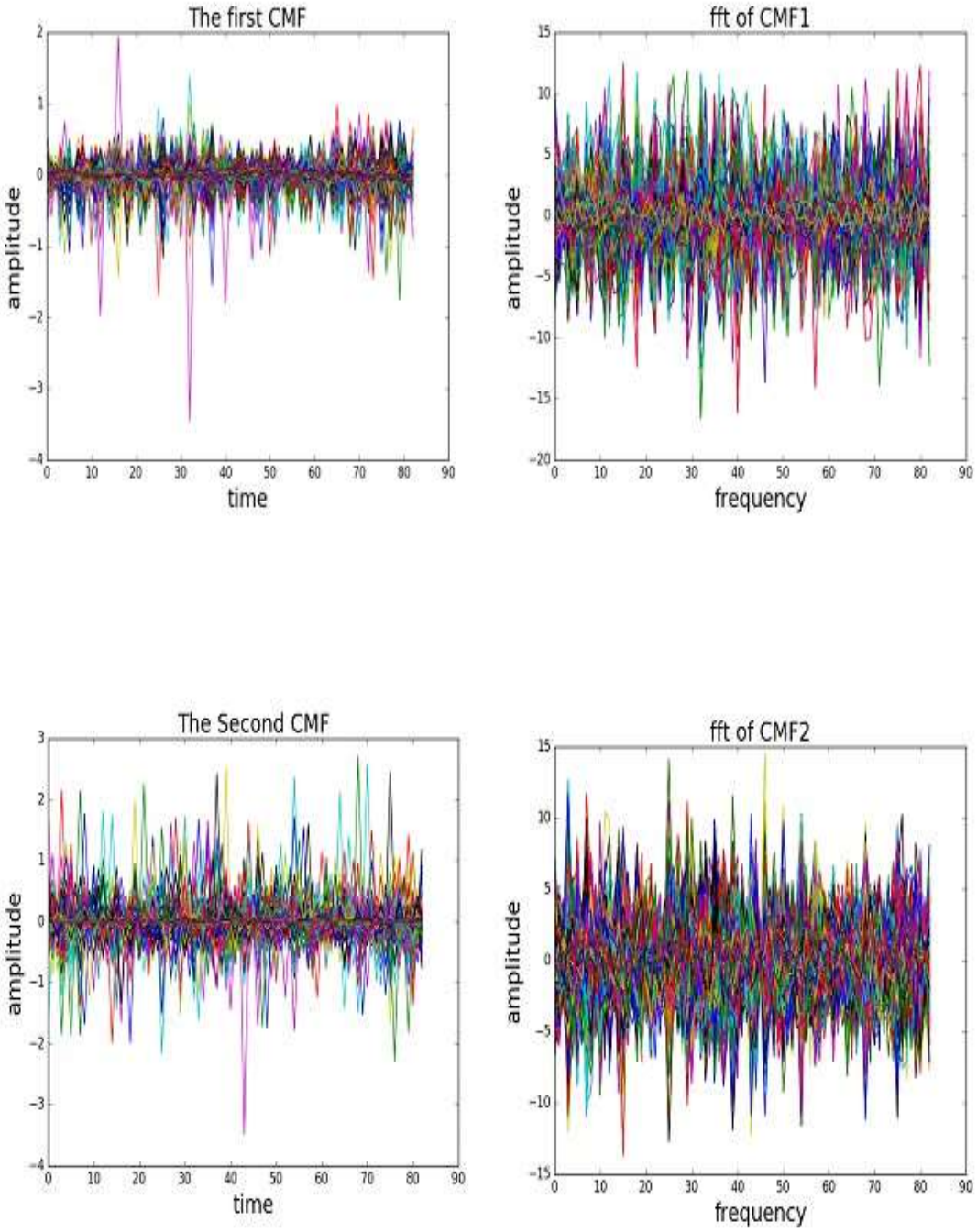

Figure 4.6: The two CMFs after merging IMFs

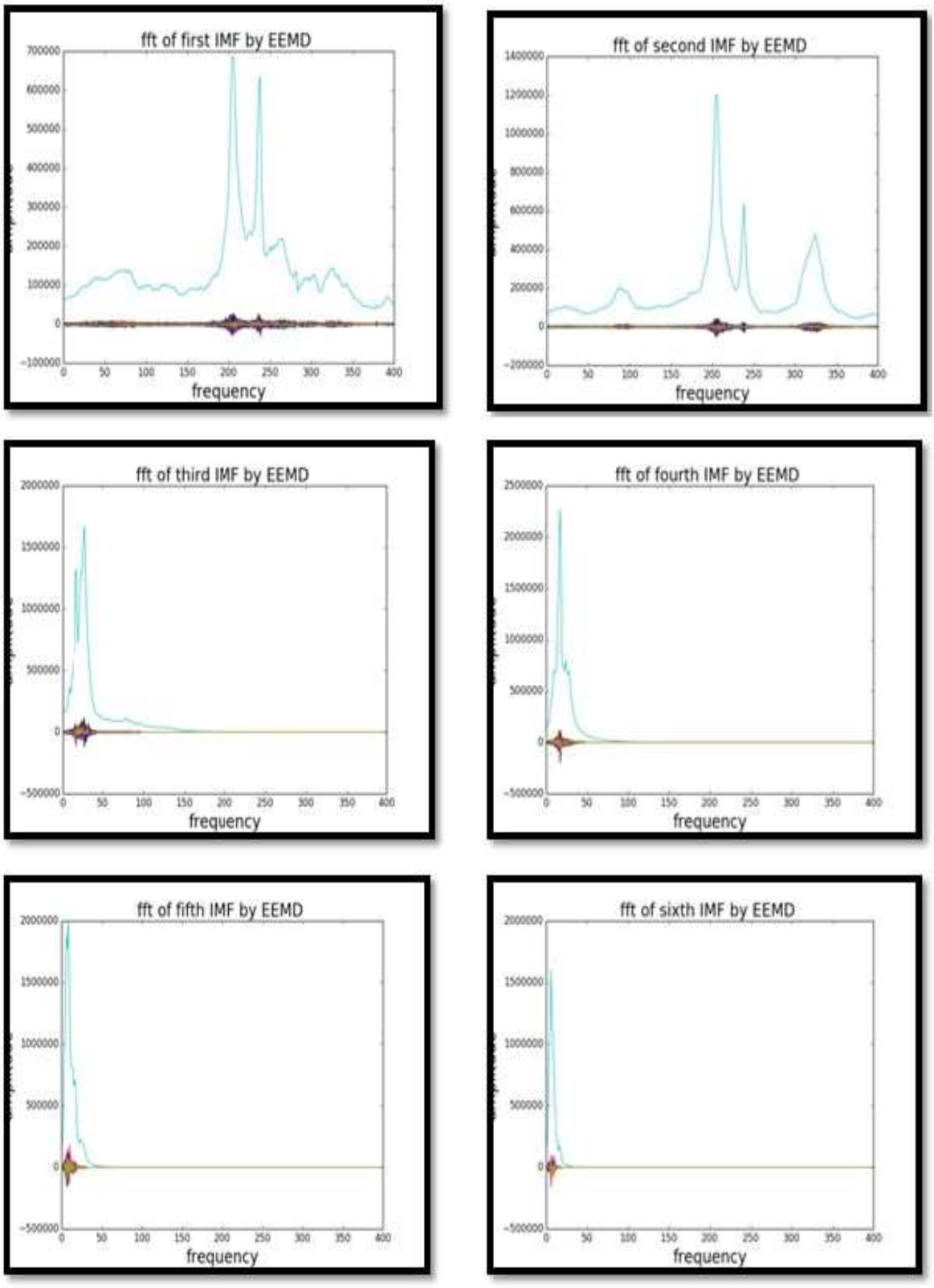

Figure 4.7: The IMFs obtained by decomposition of signal

4.4 Results and Analysis

In this chapter, we are testing the CNN-CMF-EEMD method and ANN- CMF-EEMD method for Fault Diagnosis using for fault classification.

A) Ensemble Empirical Mode Decomposition:

In this section, the above-mixed signal is firstly decomposed to IMF by EEMD method.

Then the cross-correlation coefficient of IMF and the original signal is calculated. The IMF can be observed in Figure 4.7 has been done using 900 rpm dataset [Wang, H.*et al.*,2014]. Figure 4.5 Shows the original input signal used. It can be seen clearly that the original compound fault signal is the combination of multiple weak signals and strong confusion noise(unwanted signals). The spectral analysis of Figure 4.7 shows impulsive peaks, which may be the fault frequency of strong noise signal. On decomposing the input signal using EEMD, we obtained 11 IMFs.

B) Combined Mode Function:

For the CMF analysis, the data is taken from the same 900 rpm data [Wang, H.*et al.*,2014]. The addition of all IMFs in sequential CMFs does not significantly change the spectra, apart from making more evident the contribution of multiples of the fundamental rotating frequency and other components that are not related to known defects. IMF selection seems to reduce the capability of detecting some embedded phenomena.

In the first iteration, we got our condition between SSW and SSB satisfied. Hence, we stop the loop there only. This resulted in 2 CMFs. Fourier transforms, and the CMFs are shown in Figure 4.6.

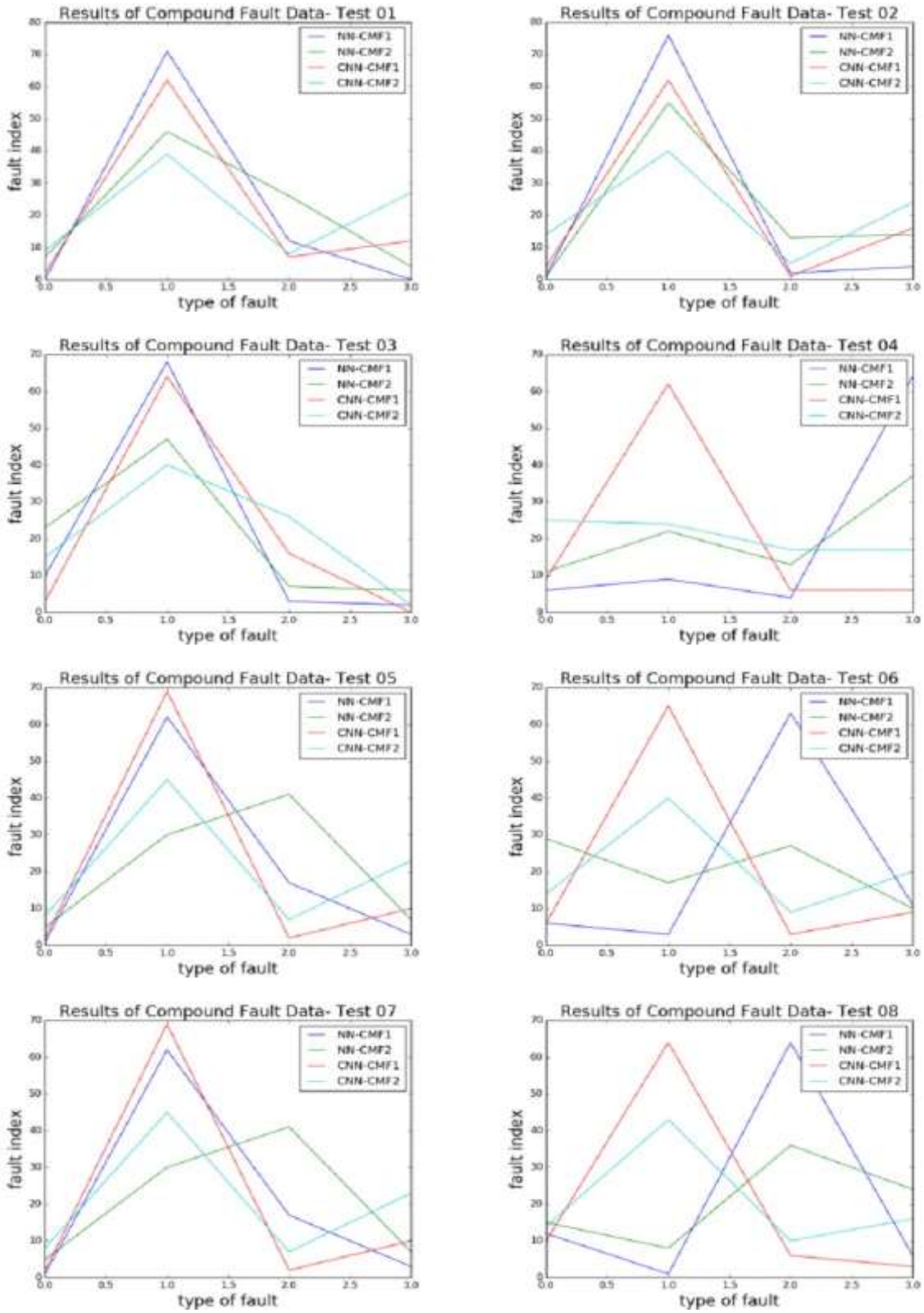

Figure 4.8: The faults observed at roller (1.0) and outer race (3.0) by the proposed method

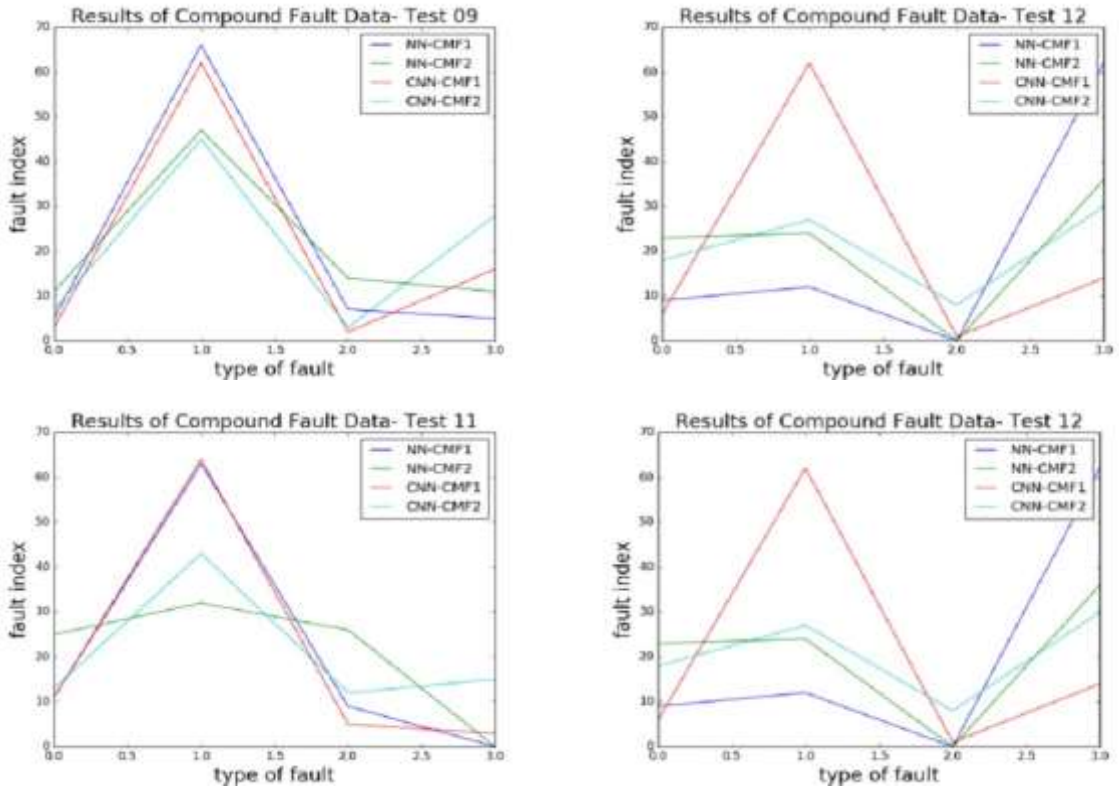

Figure 4.9: The faults observed at roller (1.0) and outer race (3.0) by the proposed method

This shows that the whole 11 IMFs can be represented with the help of 2 CMFs. As the two CMFs can provide the information present in the 11 IMFs of EEMD, we proceed forward for further classification with 2 CMFs only. The proposed approach works in a fully data-driven way by evaluating the role played by each IMF in determining the spectral property of the signal. The main idea of the approach is to compute the empirical probability density function of the CMFs frequency spectra and compute a dissimilarity index between density functions of adjacent IMF to cluster them. The enhanced computation of CMFs is expected to reduce the dimensionality of the problem and improve the interpretation of the system health conditions on other IMF selection methods commonly used in practice. The IMF's generated from the EEMD technique are combined to form CMF's, and such a combination can be interpreted as a new adaptive filter bank, which has the benefit of increasing the EEMD accuracy.

Table 4.3: The results of tests 1-6 for CNN-CMF-EEMD model

Fault		**ANN-CMF-EEMD**						**CNN-CMF-EEMD**					
		test1		test2		test3		test1		test2		test3	
Type	**Code**	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2
Healthy	0.0	01	01	00	07	10	23	04	14	2	9	03	15
Ball	1.0	76	55	71	46	68	47	62	40	62	39	64	40
Inner	2.0	02	13	12	26	03	07	01	05	07	08	16	26
Outer	3.0	04	14	00	04	02	06	16	24	12	27	00	02
Accuracy		85.43		86.95		87.59		91.65		90.53		93.33	
No of Tested samples		83	83	83	83	83	83	83	83	83	83	83	83

Fault		**ANN-CMF-EEMD**						**CNN-CMF-EEMD**					
		test4		test5		test6		test4		test5		test6	
Type	**Code**	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2
Healthy	0.0	06	11	01	05	06	29	09	25	02	08	06	14
Ball	1.0	04	22	62	30	03	17	62	24	69	45	65	40
Inner	2.0	04	13	17	41	63	27	06	02	02	07	03	09
Outer	3.0	64	37	03	07	11	10	06	17	10	23	09	20
Accuracy		80.12		84.18		77.03		94.68		92.86		93.14	
No of Tested samples		83	83	83	83	83	83	83	83	83	83	83	83

Table 4.4: The results of tests 7-12 for CNN-CMF-EEMD model

Fault		**ANN-CMF-EEMD**						**CNN-CMF-EEMD**					
		test7		test8		test9		test7		test8		test9	
Type	**Code**	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF2
Healthy	0.0	00	02	12	15	05	11	05	19	10	14	03	07
Ball	1.0	66	40	01	08	66	47	64	34	64	43	62	45
Inner	2.0	07	13	64	36	07	14	12	23	06	10	02	03
Outer	3.0	10	28	06	24	05	11	02	07	03	16	16	28
Accuracy		87.03		90.12		91.07		94.76		94.23		93.73	
No of Tested samples		83	83	83	83	83	83	83	83	83	83	83	83

Fault		**ANN-CMF-EEMD**						**CNN-CMF-EEMD**					
		Test10		Test11		Test12		Test10		Test11		Test12	
Type	**Code**	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF 2	CMF 1	CMF2
Healthy	0.0	08	16	11	25	09	23	64	42	11	13	06	18
Ball	1.0	73	57	63	32	12	24	00	04	64	43	62	27
Inner	2.0	01	04	09	26	00	00	02	02	05	12	01	08
Outer	3.0	01	06	00	00	62	36	17	35	03	15	14	30
Accuracy		90.12		89.75		80.56		90.34		94.6		95.94	
No of Tested samples		83	83	83	83	83	83	83	83	83	83	83	83

(C) Convolution Neural Network (CNN): The datasets used for training the neural networks are obtained from Case Western University (CWU) [158]. The training dataset is classified into

following types: The first class contains the normal healthy dataset, the second, third and the fourth classes contain roller, inner, and outer racer faults respectively. The test data has 83 samples in each test and each sample have 1200 data points. With CNN-CMF-EEMD, we achieved an accuracy of ~ 94 percent, with a learning rate of 0.0001, 600 hidden units, input 1x30x40, conv1 14x28x38 pool1 14x14x19, conv2 16x13x18, and the 4-class classification. For other combination of learning parameters, its result is degraded, i.e., it is showing lower accuracy. These parameters were selected by purely trial and error basis under suggested guidelines, as we generally do in case of neural networks for a different type of data sets.

When CNN is implemented with CMF and EEMD, 11 out of 12 results are of roller fault are detected precisely (Figure 4.8 and Figure 4.9). The outer race fault detection varied with less degree of fault indices. The Table 4.3 and Table 4.4 also show that the proposed CNN-CMF-EEMD method has shown its capability for diagnosis of roller faults which are supposed to be most difficult to get diagnosed in rolling bearings. It is able to resolve it in a better way as compared to other available techniques.

In general, Deep Neural Networks (DNNs) are trained under one of two general tasks: regression and classification. In a regression task, the network learns to generate a real-valued output that matches the ground-truth. In a classification task, the network learns to categorize an input to one of the training classes. To train a multi-class single-label classification network, SoftMax cross-entropy loss is by far the most popular loss function for the training regime, where the ground-truth is a binary vector consisting of a value one at the correct class index, and 0s everywhere else. During training, the objective is to minimize the negative log-likelihood of the loss by multiplying the network's predictions to the binary ground truth vectors. In deep learning, existing CNN's are typically trained with a soft-max cross-entropy loss which considers the ground-truth class by maximizing the predicted probability of the correct label. This cross-entropy loss sometimes ignores the intricate inter-class relationships

that exist in the data. This ignorance is responsible for misinterpretation of inter-class relationship and wrong classification of CNN due to improper learning, which affects the performance of CNN-CMF-EEMD method.

4.5 Conclusion

The Proposed method is a composite of EEMD, CMF, and Convolution Neural Network techniques. In this chapter, we have used the EEMD technique to generate IMF for compound fault detection in the roller bearings along with CMF algorithm as selection criteria. The CMFs generated are used as input to CNN for fault diagnosis. The IMF's generated from the EEMD technique are combined to form CMF's, and such a combination is interpreted as a new adaptive filter bank, which has the benefit of increasing the EEMD accuracy. The proposed approach works in a fully data-driven way by evaluating the role played by each IMF in determining the spectral property of the signal. The minimal number of final CMFs is eventually determined by applying a criterion that inherits the cluster validity principle used in unsupervised classification. The application of the method showed that the method is suitable to reduce the number of relevant modes from many IMF to few CMFs and, simultaneously, to enhance the interpretation and characterization of multiscale phenomena of interest. This study showed the investigation of the nature and possible causes of bearing defects. The diagnosis of compound faults is improved by CNN-CMF-EEMD approach due to its extraordinary capability of detecting roller faults These faults were supposed to be most difficult to detect due to combined spin and the circular motion of rollers. For improving these classification results from CNN techniques, a known classified single-faulty dataset is diagnosed alongside with the unknown compound-faulty dataset, and the training of CNN is assumed to be correct only if it predicts the known fault accurately. Only those predictions of compound fault by CNN-CMF-EEMD are considered which predicted known faults precisely. The credibility of

CNN-CMF-EEMD technique is thus verified to classify the compound faults in the rolling bearing.