Master of Science

Bearing Fault Diagnosis under Variable Speed using Machine Learning & Digital Signal Processing Techniques

The Graduate School of the University of Ulsan

Department of Electrical/Electronic & Computer Engineering

Tra Viet

Bearing Fault Diagnosis under Variable Speed using Machine Learning & Digital Signal Processing Techniques

Supervisor: Kim Jong-Myon

A Dissertation Submitted to

The Graduate School of the University of Ulsan

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

By

Tra Viet

Department of Electrical/Electronic & Computer Engineering

Ulsan, Korea

December, 2017

Bearing Fault Diagnosis under Variable Speed using Machine Learning & Digital Signal Processing Techniques

This certifies that the Master's thesis of Tra Viet is approved.

Young-Keun Kwon

Committee Chair Dr.

Jong-Myon Kim

Committee Member Dr.

Ki-Chang Im

Committee Member Dr.

Department of Electrical/Electronic & Computer Engineering

Ulsan, Korea

December, 2017

Declaration of Authorship

I, TRA VIET, declare that this thesis titled, "Bearing Fault Diagnosis under Variable Speed using Machine Learning & Digital Signal Processing Techniques" and the work presented in it is my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed and I have acknowledged all main sources of help.
- Where I have quoted from the work of others, the source is always given. Except for such quotations, this thesis is entirely my work.

Approved by:

Prof. Young-Keun Kwon, Committee Chair

Prof. Jong-Myon Kim, Committee Member

Prof. Ki-Chang Im, Committee Member

University of Ulsan
Smart HSE (Health, Safety, Environment) Laboratory
Department of Electrical/Electronic & Computer Engineering, University of Ulsan, South Korea

Abstract

Bearing Fault Diagnosis under Variable Speed using Machine Learning & Digital Signal Processing Techniques

BY TRA VIET

Rolling element bearings are crucial components of rotating machines, and are the leading cause of failure in essential industrial equipment such as rolling machines in paper mills, wind turbines, and induction motors. A faulty bearing may cause equipment breakdown, which can lead to unscheduled and costly downtime for an entire industry; thus, the detection of incipient bearing faults is essential for maintaining production schedules and minimizing costs. Various techniques have been used for fault detection in bearings including the analysis of vibration acceleration signals and the motor stator current. These techniques are useful in diagnosing bearing defects at high rotational speeds; however, incipient bearing defects at low rotational rates, which are characterized by their small energy acoustic emissions (AE), can be more efficiently diagnosed with AE-based methods. These methods can diagnose bearing defects before they appear on the bearing surface. In this thesis report, three AE-based techniques are presented for diagnosing faults in bearing. First is a method for diagnosing incipient bearing defects under variable speed conditions, by extracting features from different subbands of the inherently non-stationary AE signal, and then classifying bearing defects using a weighted committee machine, which is an ensemble of support vector machines and artificial neural networks. Second is a novel method for diagnosing bearing abnormalities under variable operating speeds using convolution neural networks (CNNs). The CNNs use the energy distribution maps of the

AE signal spectrum as inputs, and automatically extract the optimal features that can be used to diagnose various single and compound bearing defects under variable speed conditions. In the third, we improve the performance of above CNNs-based method by training CNNs using Stochastic Diagonal Levenberg-Marquardt algorithm, a robust training algorithm that combines the Gauss-Newton and the steepest descent methods to exploit the speed advantage of the former and the stability of the later. It yields better convergence results, even for complex non-quadratic error functions. The proposed methods using data as AE signals are generated by the experimental testbed of Smart HSE Laboratory to validate the presented methodologies. Experimental results demonstrate that the proposed methods yield better diagnostic performance in comparison to state-of-the-art AE-based methods, by can extract optimal features and enhance the performance of classifiers…

Acknowledgement

Firstly, I offer my gratitude to my dear parents who guided me and supported me morally and financially at every stage of my life to achieve this success. Moreover, I appreciate the kind and affectionate supervision of my research advisor Prof. JongMyon Kim, which enabled me to complete my degree in a sophisticated way and instilled in me profound research potentials. I am thankful to all my family members, friends and colleagues who're well wishes and support helped me to get to this milestone

Next, I would like to thank the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (NRF-2012R1A1A2043644) for their financial support. Moreover, I am thankful to Korean government to provide me with Brain Korea 21 (BK21) scholarship which enables me to pursue my studies, join local and international conferences and publish my research manuscripts in international journals. I would like to acknowledge my committee members for their valuable comments and for their broad perspective in redefining the ideas in this dissertation.

TRA VIET

Ulsan, November 2017

Contents

List of Figures

List of Tables

Dedicated to my dear Parents, whose upbringing, guidance, support and peerless love have enabled me to achieve this success. **T**hank you, my wife, for always being there for me, love me, care for me, without you this thesis may have been incomplete.

1 Introduction

1.1 Motivation

Induction motors are massively used as a low-speed rotating machine to support heavy loads at steady rotational speed in industries. These stresses result in different types of failures that occur in various parts of industrial motors. Sudden failures lead to unexpected shutdown in the industrial system and result in enormous economic losses as well as safety threat (Berriri *et al.*, 2012; Cabal-Yepez *et al.*, 2013).

Several types of failures (i.e. rotor related, stator related, bearing related and others) arise in an industrial motors. Among these failures, bearing related faults are significant in number and make it almost half of the overall failures. According to the survey of faults on induction motors conducted by Electric Power Research Institute (EPRI) and the Motor Reliability Working Group of IEEE-IAS, bearing related failures are more than 50 percent of all failures in rotating motor (Thorsen and Dalva, 1999). A faulty bearing may cause equipment breakdown, which can lead to unscheduled and costly downtime for an entire industry; thus, the detection of incipient bearing faults is essential for maintaining production schedules and minimizing costs.

1.2 Background of Bearing Fault Diagnosis

Various techniques have been used for fault detection in bearings including the analysis of vibration acceleration signals (Bediaga *et al.*, 2013; Immovilli *et al.*, 2013; Jin *et al.*, 2014; Seshadrinath *et al.*, 2014) and the motor stator current (Frosini and Bassi, 2010; Lau and Ngan, 2010; Elbouchikhi *et al.*, 2017). These techniques are effective in diagnosing bearing defects at high rotational speeds; however, incipient bearing defects at low rotational speeds, which are characterized by their low energy acoustic emissions (AE), can be more effectively diagnosed with AE-based methods (Caesarendra *et al.*, 2013; Niknam *et al.*, 2013; Kang *et al.*, 2015d; Nguyen *et al.*, 2015; Kang *et al.*, 2016). These methods can diagnose bearing defects before they appear on the bearing surface (Tandon and Choudhury, 1999). AE-based methods have predominantly utilized either envelope analysis (Kang *et al.*, 2015d) or feature-extraction-based techniques for bearing fault diagnosis (Kang *et al.*, 2015c; Kang *et al.*, 2016; Tra *et al.*, 2017).

Feature extraction-based methods for bearing fault diagnosis usually consist of three steps: signal acquisition, extraction of features from the acquired signal, and classification of various defect types based upon the extracted features. The extracted features consist primarily of statistical properties of the time and frequency domain AE signal, along with quantities calculated through the complex envelope analysis of the AE signal. Feature extraction is of great significance in these methods, as discriminative features yield discriminative models and hence good diagnostic performance. However, designing discriminative features not only requires significant domain knowledge, more often than not, feature selection algorithms are also required to eliminate irrelevant and redundant features that degrade the performance of these methods (Rauber *et al.*, 2015; Kang *et al.*, 2016). Thus, the diagnostic performance of these methods relies heavily on the quality of features used to construct discriminative models by the classifiers.

Envelope analysis-based methods diagnose bearing defects by detecting peaks at the characteristic bearing defect frequencies (BDFs) in the power spectrum of the complex envelope signal (Randall and Antoni, 2011; Bediaga *et al.*, 2013; Kang *et al.*, 2015b; Kang *et al.*, 2015d; Nguyen *et al.*, 2015). When a bearing is rotating at constant speed, the repeated excitation of a localized bearing defect results in transient impulses in the waveform of the AE signal that occur periodically or quasi-periodically. The frequency of these impulses is referred to as the BDF, and is dependent upon the location of the bearing defect, the bearing geometry, and its rotational speed. The repeated impacts of a bearing's moving parts against bearing defects excite the resonant frequencies of adjacent components. These resonant frequencies have a high bandwidth and are modulated by the BDF, thus the resonant frequencies act as carrier frequencies, and the BDF acts as the modulation frequency (Randall and Antoni, 2011; Wang *et al.*, 2015). Through envelope analysis, the recorded AE signals are first demodulated, and then their power spectra are analyzed to detect the BDF associated with each type of localized defect.

1.3 Purpose and Goals of Research and Key points of Contribution

In practice, bearing operation is mostly subject to variations in speed as, for example, in wind turbines. Both envelope analysis and feature extraction-based methods are susceptible to variations in the rotational speed of the bearing and hence are of limited utility under such conditions. As BDFs are functions of a bearing's rotational speed, any variation in the operating speed of a bearing results in translation of the power spectrum of the envelope signal along the frequency axis. Hence, traditional envelope analysis-based methods, which rely on detecting peaks at precise frequencies, do poorly under variable-speed conditions (Lacey, 2008). Similarly, feature extraction-based methods diagnose bearing defects by exploiting the variation between the features of the AE signals for specific defect types. For a given defect type at a constant rotational speed, the features extracted from the AE signals recorded at various times do not differ significantly, and serve to distinguish the defect type in question from AE signals for other defect types. However, variations in a bearing's speed can affect the amplitude of the impulses generated by repeated impact of moving parts with the defective areas. This can lead to significant variation in features, even for the same defect type (Lacey, 2008). Thus, AE signals captured at different rotating speeds will have different attributes, resulting in inconsistent feature vectors, hence limiting the direct application of these methods under variable speed conditions. Moreover, order analysis (OA) based methods have also been proposed to diagnose bearing defects under variable speed (Fyfe and Munck, 1997; Wang *et al.*, 2014a; Wang *et al.*, 2014b; Wang *et al.*, 2015), which use characteristic fault orders instead of BDFs to diagnose different bearing defects. These methods can be categorized into tachometer-based and tacho-less OA methods. Tachometer-based OA methods require additional speed sensors on the machine making them costlier, whereas tacho-less methods are susceptible to noise as these methods estimate the speed from the acquired signals (Lu *et al.*, 2016). The authors in (Lu *et al.*, 2016) propose a tacho-less OA method that uses a high speed camera to estimate the rotating speed of the bearing; this technique also requires specialized hardware, which is costly and may not be suitable for general use. The proposed methodologies offer solution for diagnosing bearing defects under variable speed conditions.

The contributions of this thesis are summarized as follow:

- We proposed a method that decomposes the raw AE signal into different subband signals using the discrete wavelet packet transform (DWPT), which is more effective for the multiresolution analysis of non-stationary signals. A feature vector is extracted from each subband signal, resulting in multiple feature vectors from the raw AE signal. The feature vectors extracted from different subbands of the original AE signal result in models of the fault data that are discriminatory, even under variations in the rotational speed. The weighted committee machine (WCM), which is an ensemble of support vector machines (SVMs) and artificial neural networks (ANNs), is employed to build discriminatory models of the feature vectors for different bearing defects.

- We proposed a novel method for diagnosing bearing defects under variable speed conditions. The proposed method does not require any specialized hardware, rather it employs convolutional neural networks to mine the energy distribution maps for discriminative features, which can be used to diagnose bearing defects under variable operating speeds. This work proposes the use of energy distribution maps of the AE signal spectrum as inputs to the CNNs for the diagnosis of bearing defects under variable operating speeds. The use of energy distribution maps as inputs to the CNNs is justified by a detailed analysis of the effects of speed on the AE signals, and a comparison of different types of inputs used to train the CNNs.

- We also investigated various training algorithms for CNNs and proposed the use of S-DLM algorithm for training the CNNs as it results in faster convergence and a better diagnostic performance.

1.4 Organization of Master Thesis

Research work for the Master thesis are organized in the given manner presented in Fig. 1-1

Diagnosis of Rolling Element Bearings

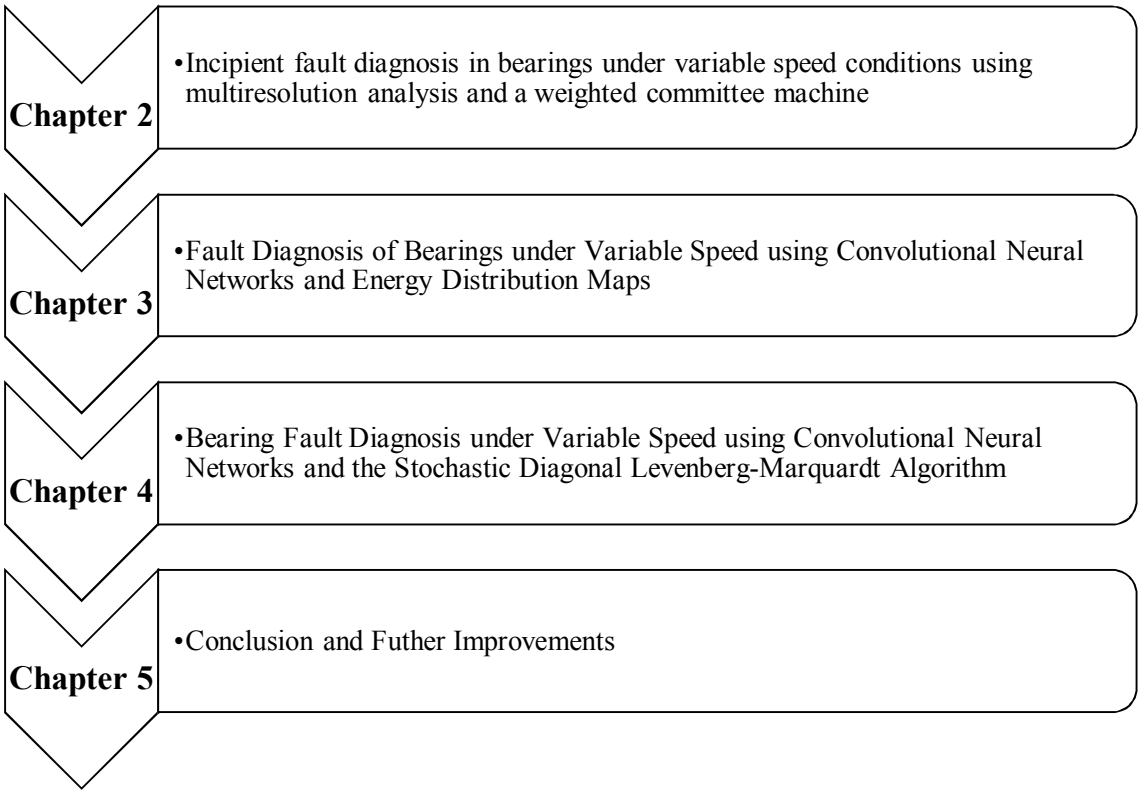


FIGURE 1-1: Organization of research thesis report

2 Incipient fault diagnosis in bearings under variable speed conditions using multiresolution analysis and a weighted committee machine

Incipient defects in bearings are traditionally diagnosed either by developing discriminative models for features that are extracted from raw acoustic emission (AE) signals, or by detecting peaks at characteristic defect frequencies in the envelope power spectrum of the AE signals. Under variable speed conditions, however, such methods do not yield the best results. This letter proposes a technique for diagnosing incipient bearing defects under variable speed conditions, by extracting features from different subbands of the inherently non-stationary AE signal, and then classifying bearing defects using a weighted committee machine, which is an ensemble of support vector machines and artificial neural networks. The proposed method also improves the generalization performance of the neural networks to enhance their classification accuracy, particularly with limited training data.

2.1 Introduction

Bearings are the leading cause of failure in induction motors, and 51% of failures are attributed to a defective bearing (Thorsen and Dalva, 1999). These failures can cause unscheduled and costly shutdowns. The early detection of these defects is helpful in preventing such abrupt shutdowns. The initiation and propagation of cracks in the bearing

material is associated with the sudden release of energy, which propagates as structure-borne waves or acoustic emissions (AEs). AE activity can be used to detect material defects in bearings; thus, acoustic emissions are commonly used to monitor the condition of bearings, especially for the diagnosis of incipient faults (ISO/TC 108, 2007).

Localized defects in bearings, including cracks on the roller and inner/outer raceways, can be diagnosed via envelope analysis of the AE signals. This involves the detection of peaks in the power spectrum of the AE envelope signal at the characteristic frequencies associated with each defect type, such as the ball spin frequency ($2 \times BSF$), the ball pass frequency over the outer raceway (BPFO), and the ball pass frequency over the inner raceway (BPFI) (Kang *et al.*, 2015a). However, since these are kinematic quantities, these characteristic frequencies depend on the rotation speed and bearing geometry. Therefore, envelope analysis-based methods are of limited utility in the diagnosis of bearing faults under variable speed conditions. Khan et al. proposed a vibration signal imaging-based method for rotational speed-invariant fault diagnosis in bearings (Khan and Kim, 2016). In their technique, they render the vibration acceleration signal as a grayscale image and use texture descriptors, which are based on the local binary operator, for the diagnosis of different faults. The high rotational speed of bearings and the lower sampling rate (12 kHz) resulted in clearly distinguishable textures and easy to handle image dimensions for the data (Khan and Kim, 2016). However, acoustic emissions are sampled at very high sampling rates (such as 250 kHz for this study) because their frequency content lies within the range of 20 kHz to 1 MHz (ISO/TC 108, 2007). Imaging-based methods would result in very large images with sparsely distributed textures. Moreover, AE signals are usually used for incipient fault diagnosis at low rotational speeds (Kang *et al.*, 2015a), resulting in lower impulses in the AE data; thus, AE signals would be rendered as barely visible textures by an imaging-based method.

Feature extraction based methods usually extract features from the raw condition monitoring signals, i.e., vibration acceleration or AE, and then use those features to develop discriminatory models of the data to classify different bearing defects (Rauber *et al.*, 2015). However, under variable speed conditions, features extracted from the raw AE signal, result in overlapped feature spaces that result in poor discriminatory models. In this letter, an

improved feature extraction based method is proposed to diagnose incipient bearing defects. The proposed method decomposes the raw AE signal into different subband signals using the discrete wavelet packet transform (DWPT), which is more effective for the multiresolution analysis of non-stationary signals (Yang *et al.*, 2009). A feature vector is extracted from each subband signal, resulting in multiple feature vectors from the raw AE signal. In contrast, traditional feature extraction based methods extract a single feature vector from the raw AE signal. The feature vectors extracted from different subbands of the original AE signal result in models of the fault data that are discriminatory, even under variations in the rotational speed, as shown by the results in Section 2.4. The weighted committee machine (WCM), which is an ensemble of support vector machines (SVMs) and artificial neural networks (ANNs), is employed to build discriminatory models of the feature vectors for different bearing defects.

2.2 The acquisition of acoustic emission data

For this study, acoustic emission data for both normal and defective bearings are captured using a wide-band AE sensor; these are recorded at a sampling rate of 250 kHz using a PCI-2 system. These signals are then divided into two datasets based upon the defect size, as shown in Table 2-1. Each dataset has signals for a bearing in normal condition (BNC) and bearings with seven types of localized defects, including a crack on the outer raceway (BCO); crack on the inner raceway (BCI); crack on the roller (BCR); cracks on the inner and outer raceways (BCIO); cracks on the outer raceway and roller (BCOR); cracks on the inner raceway and roller (BCIR); and cracks on the inner raceway, outer raceway, and roller (BCIOR). For each of these eight bearing conditions, 90 AE signals are recorded at five different rotational speeds each, as given in Table 2-1.

The length of each AE signal is five seconds. Each of the two datasets contains a total of $\sum_{Fault\_classes=1}^{8} \sum_{RPM=1}^{5} 90\langle AE\ Signals \rangle$, i.e., 3600 AE signals.

TABLE 2-1. Details of the acoustic emission datasets for two different crack sizes.

Single and compound seeded bearing failures (90 AE signals for each bearing condition), *fs*=250 kHz	Average rotational speed (RPM)	Crack size		
		length	width	depth
Dataset 1	AE signals for 5 RPMs*	3 mm	0.35 mm	0.30 mm
Dataset 2	AE signals for 5 RPMs*	6 mm	0.49 mm	0.50 mm

**Each dataset contains AE signals recorded at five different rotational speeds of 300 rpm, 350 rpm, 400 rpm, 450 rpm, and 500 rpm.*

2.3 The proposed methodology for speed-invariant fault diagnosis

Fig. 2-1 presents the proposed methodology for incipient fault diagnosis in bearings under variable speed conditions. The proposed method employs the discrete wavelet packet transform (DWPT) for the effective multiresolution analysis (MRA) of a non-stationary AE signal (Liu and Pan, 2010). Fig. 2-2 shows measured AE signals, envelope signals via analytical analysis, envelope power spectra for a defect-free bearing (i.e., normal) and defective bearings (i.e., a bearing with a crack on its outer race (BCO), a bearing with a crack on its inner race (BCI), and a bearing with a crack on its roller (BCR)). Bearing characteristic frequencies (i.e., BPFO, BPFI, and BSF) can be computed as follows:

$$\text{BPFO} = \frac{n_{roller} \cdot f_r}{2}\left(1 - \frac{B_d}{P_d}\cos\alpha\right),$$

$$\text{BPFI} = \frac{n_{roller} \cdot f_r}{2}\left(1 + \frac{B_d}{P_d}\cos\alpha\right), \text{ and}$$

$$\text{BSF} = \frac{P_d \cdot f_r}{2 \cdot B_d}\left(1 - \left(\frac{B_d}{P_d}\cos\alpha\right)^2\right) \tag{2.1}$$

where BPFO is the ball pass frequency of the outer race, BPFI is the ball pass frequency of the inner race, and BSF is the ball spin frequency, nroller is the number of rolling elements, B_d is the diameter of rolling elements, P_d is the diameter of a pitch, f_r is the rotating speed, and α is the contact angle between the rolling element line and the inner raceway of the bearing (Kang *et al.*, 2015a).

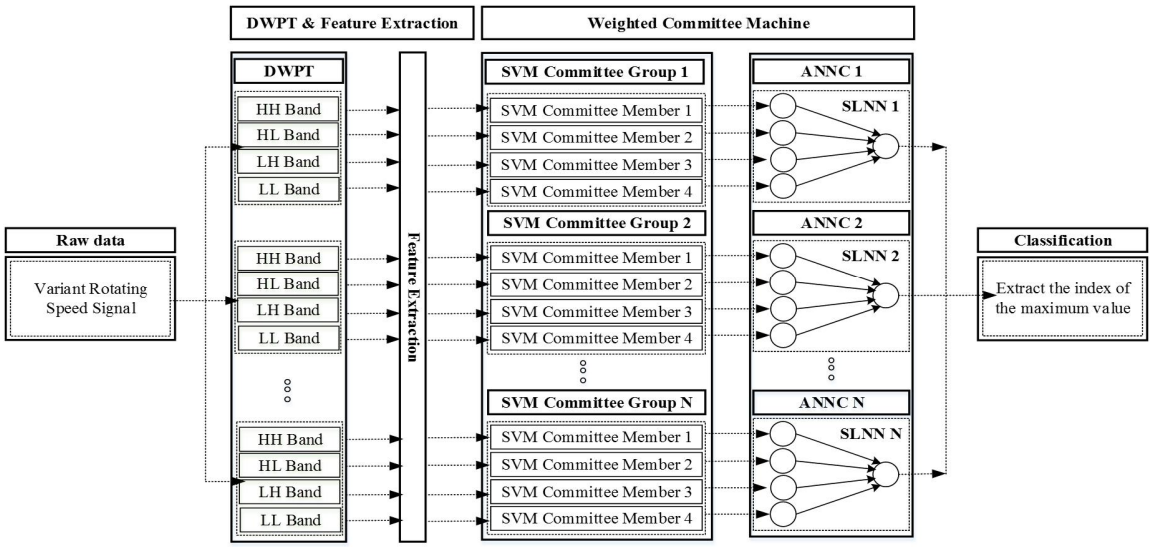


FIGURE 2-1. The proposed method for incipient fault diagnosis under variable speed conditions using a weighted committee machine.

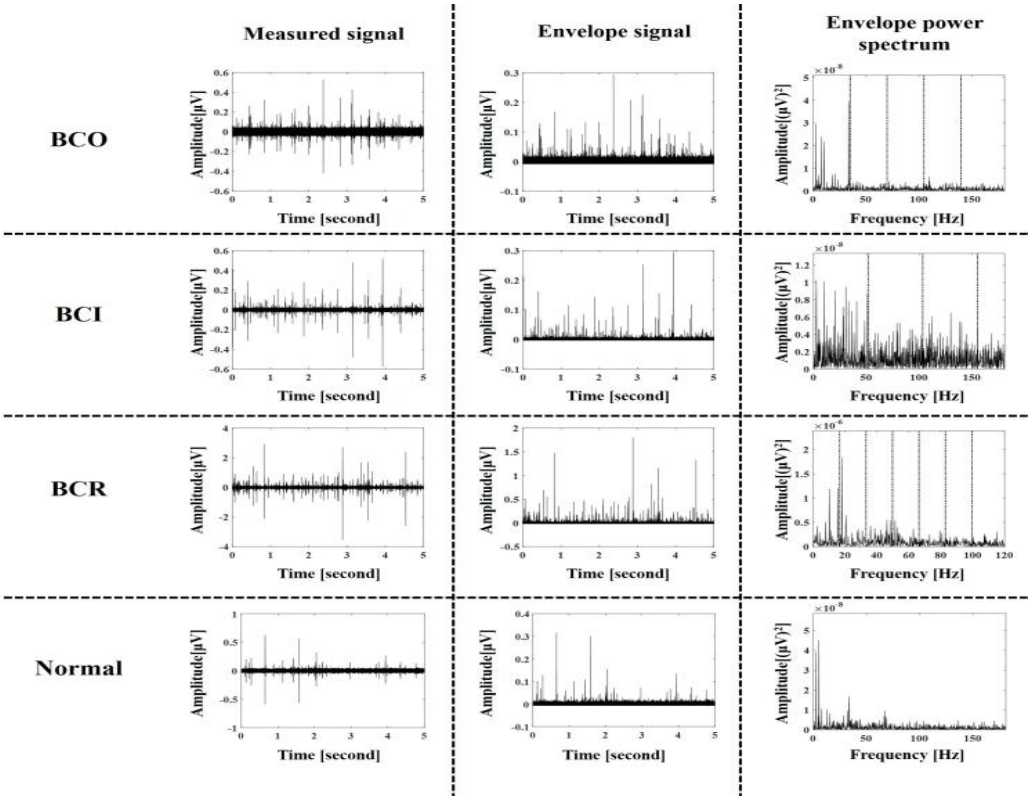


FIGURE 2-2. Measured 5-second AE signals, envelope signals, and envelope power spectra for a defect-free bearing and defective bearings.

The proposed method employs the discrete wavelet packet transform (DWPT) for the effective multiresolution analysis (MRA) of a non-stationary AE signal (Liu and Pan, 2010). Using the Daubechies 15 mother wavelet, the AE signals are decomposed into two resolutions to generate four subband signals, where the frequency range of each subband according to the Nyquist theorem are 0 ~ 31.25 KHz, 31.25 ~ 62.5 KHz, 62.5 ~ 93.75 KHz, and 93.75 ~ 125 KHz in the sampling rate of 250 KHz. For each subband signal, 19 statistical quantities are calculated, as given in Tables 2-2 and Tables 2-3. Thus, for each AE signal, a set of four feature vectors is extracted (i.e., one from each of its four subbands). The number of subbands is determined empirically to extract sufficient discriminatory fault information for rotational speed-invariant fault diagnosis.

TABLE 2-2. The three statistical features of the frequency-domain AE subband signal.

Feature	Equation	Feature	Equation	Feature	Equation
Frequency center (FC)	$\frac{1}{N_{tfreqbin}}\sum_{n=1}^{N_{tfreqbin}} S(f)$	RMS frequency (RMSF)	$\sqrt{\frac{1}{N_{tfreqbin}}\sum_{n=1}^{N_{tfreqbin}} S(f)^2}$	Root variance frequency (RVF)	$\sqrt{\frac{1}{N_{tfreqbin}}\sum_{n=1}^{N_{tfreqbin}} (S(f)-FC)}$

Here, S(f) is the magnitude response of the FFT of x(n) and $N_{tfreqbin}$ is the total number of frequency bins.

Traditional feature extraction based methods for bearing fault diagnosis, extract features from the raw fault signal only, i.e., vibration acceleration or AE signal. However, as demonstrated by the results in Table 2-4, it results in poor discriminatory models for different fault types, under variable speed conditions. The discriminative performance of these models improves when the raw AE signal is decomposed into different subbands, and feature vectors extracted from those subbands are used to construct discriminatory models for different types of bearing faults or defects. The AE signals are decomposed into four subbands only, as further decomposition does not significantly improve the diagnostic performance of the proposed method and only adds to the complexity of the WCM.

These feature vectors are used to train the WCM and subsequently classify unknown AE signals. The WCM is an ensemble of two classifiers: support vector machine committee groups (SVMCGs) in the first stage and artificial neural networks as a combiner (ANNC) in the second stage. The SVMCGs in the first stage are used as a prior classifier, whereas the ANNC in the second stage combines the outputs of individual members of each

SVMCG. In this study, the first stage of the WCM has eight SVMCGs (i.e., one for each fault class). There are four SVM committee members in each SVMCG, which are trained on the feature vectors from the corresponding subband signals. All of the SVMs in an SVMCG are trained using the one-against-all multi-class (OAA MC) SVM framework. Thus, each SVMCG is used to create a discriminatory model that can identify a single fault type, independent of the rotational speed, using feature vectors extracted from the four subbands of the original AE signal.

TABLE 2-3. The 16 statistical features of the time-domain AE subband signals.

Feature	Equation	Feature	Equation	Feature	Equation
Peak (x_{peak})	$max\left(\lvert x(n)\rvert\right)$	Shape factor	$\frac{x_{rms}}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$	Mean ($\bar{x}$)	$\frac{1}{N}\sum_{n=1}^{N}x(n)$
Root-mean-square (x_{rms})	$\sqrt{\frac{1}{N}\sum_{n=1}^{N}x(n)^2}$	Entropy	$-\sum_{n=1}^{N}p(n).log_2\,p(n)$	Shape factor square-mean-root	$\frac{x_{smr}}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$
Kurtosis	$\frac{1}{N}\sum_{n=1}^{N}\left(\frac{x(n)-\bar{x}}{\sigma}\right)^4$	Skewness	$\frac{1}{N}\sum_{n=1}^{N}\left(\frac{x(n)-\bar{x}}{\sigma}\right)^3$	Peak-to-peak (PP)	$max\left(x(n)\right)$ $-min\left(x(n)\right)$
Crest factor	$\frac{max\left(\lvert x(n)\rvert\right)}{x_{smr}}$	Square-mean-root(s_{rms})	$\left(\frac{1}{N}\sum_{n=1}^{N}\sqrt{\lvert x(n)\rvert}\right)^2$	Kurtosis factor	$\frac{Kurtosis}{\left(\frac{1}{N}\sum_{n=1}^{N}x(n)^2\right)^2}$
Clearance factor	$\frac{x_{peak}}{x_{smr}}$	5th normalized moment	$\frac{1}{N}\sum_{n=1}^{N}\left(\frac{x(n)-\bar{x}}{\sigma}\right)^5$		
Impulse factor	$\frac{max\left(\lvert x(n)\rvert\right)}{\frac{1}{N}\sum_{n=1}^{N}\lvert x(n)\rvert}$	6th normalized moment	$\frac{1}{N}\sum_{n=1}^{N}\left(\frac{x(n)-\bar{x}}{\sigma}\right)^6$		

Here, x is an input signal, N is the total number of samples, $\sigma=\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x(n)-\bar{x}\right)^2}$ *, and* $p(n)=x(n)^2\Big/\sum_{n=1}^{N}x(n)^2$

The outputs of each SVMCG (i.e., the four decision values of its member SVMs) are then used to train the artificial neural networks in the second stage, which act as a combiner for the first stage outputs. The WCM labels an unknown AE signal using the index of the maximum value in the ANNC outputs.

TABLE 2-4. Average classification accuracies and sensitivities for single and multiple combined bearing defects using different levels of decomposition of the original AE signal.

Datasets	No of Subbands	Average Sensitivity for each fault type								ACA
		BCI	BCO	BCR	BCIO	BCIR	BCOR	BCIOR	BNC	(%)
Dataset 1	1	100.0	80.00	100.0	91.11	86.66	97.77	86.66	97.77	92.50
	2	95.55	100.0	97.77	88.88	91.11	91.11	91.11	100.0	94.44
	4	100.0	90.00	100.0	100.0	98.88	98.88	95.55	97.77	97.64
	8	96.66	100.0	100.0	97.77	98.89	98.88	93.33	96.66	97.77
Dataset 2	1	93.33	97.77	93.33	95.55	82.22	100.0	91.11	97.77	93.88
	2	97.77	97.77	95.55	100.0	95.55	97.77	93.33	100.0	97.22
	4	98.88	98.88	95.55	100.0	95.55	100.0	98.88	100.0	98.47
	8	100.0	100.0	100.0	97.77	95.55	97.77	100.0	100.0	98.89

These feature vectors are used to train the WCM and subsequently classify unknown AE signals. The WCM is an ensemble of two classifiers: support vector machine committee groups (SVMCGs) in the first stage and artificial neural networks as a combiner (ANNC) in the second stage. The SVMCGs in the first stage are used as a prior classifier, whereas the ANNC in the second stage combines the outputs of individual members of each SVMCG. In this study, the first stage of the WCM has eight SVMCGs (i.e., one for each fault class). There are four SVM committee members in each SVMCG, which are trained on the feature vectors from the corresponding subband signals. All of the SVMs in an SVMCG are trained using the one-against-all multi-class (OAA MC) SVM framework. Thus, each SVMCG is used to create a discriminatory model that can identify a single fault type, independent of the rotational speed, using feature vectors extracted from the four subbands of the original AE signal. The outputs of each SVMCG (i.e., the four decision values of its member SVMs) are then used to train the artificial neural networks in the second stage, which act as a combiner for the first stage outputs. The WCM labels an unknown AE signal using the index of the maximum value in the ANNC outputs.

The SVM is a supervised binary classification algorithm that finds the largest margin hyperplane in the feature space of two classes. This hyperplane is then used to label test

samples from both classes. It can solve nonlinear classification problems by mapping them to high-dimensional feature spaces using kernel functions (Ma et al., 2015). In this study, we use the Gaussian radial basis function (RBF) as a kernel function. Using Lagrange multipliers, the SVM algorithm can be reduced to solve the following optimization problem:

$$\begin{aligned} &\arg\max_{\alpha_i}\left\{\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j K(x_i,x_j)\right\}, \\ &\textit{subject to} \qquad \sum_{i=1}^{n}\alpha_i y_i=0, 0\le\alpha_i\le C \qquad \forall i=1,2,...,n \end{aligned} \tag{2.2}$$

Here, x_i and x_i represent samples from the training dataset, α_i represents the Lagrange multiplier, and C is a penalty variable for tuning the generalization performance of the SVM. The corresponding classification function for the SVM is as follows:

$$\begin{aligned} &F(x)=\operatorname{sgn}\{f(x)\} \\ &\textit{where}\ \ f(x)=\sum_{i=1}^{n}\alpha_i^* y_i^* K(x_i^*,x)+b \end{aligned}, \tag{2.3}$$

Here, α_i^* represents the Lagrange multiplier corresponding to the support vector x_i^* . The decision value $f(x)$ ranges between $-\infty$ to $+\infty$ and represents the signed distance of an unknown observation x from the decision boundary. A positive decision value for a class indicates that x is predicted to be in that class, whereas a negative value indicates otherwise (Avidan, 2004).

In the second stage of the WCM, single-layer ANNs with only one neuron in the output layer are trained independently as a binary classifier in the OAA strategy; this is done using the Levenberg-Marquardt (LM) backpropagation algorithm (Liu, 2010). Single-layer ANNs are used due to their simplistic structure and their ability to classify linearly-separable patterns. The output neurons use the sigmoid activation function; thus, the output of each ANNC lies between 0 and +1. The class label for the unknown sample $\bar{x}$ is determined using the decision function in Eq. (2.4), where y_i is the output of the ith ANNC.

$$F(\bar{x},y_1,y_2...y_8)=\arg\max(y_i) \tag{2.4}$$

Overfitting can reduce the generalization performance of an ANN, but this can be

improved through a variety of techniques, including early stopping, regularization, and retraining (Hagiwara, 2002). In this letter, we utilize these methods to improve the generalization of the ANNC, as illustrated in Fig. 2-3. The ANNC is trained using the decision values from SVMCG, which are divided into training and validation subsets at ratios of 0.7 and 0.3, respectively. The performance function of the ANNC is modified by adding the mean of the sum of squares of the network weights and biases, msw, as given in Eq. (2.5).

$$msereg = \gamma * msw + (1-\gamma) * mse \tag{2.5}$$

Here, γ is the performance ratio and mse is the mean of the sum of squares for the network errors. This modified performance function smooths the network response, making it less likely to overfit; this is done by forcing it to have smaller weights and biases. The training and validation steps are repeated multiple times for different values of γ to find the optimal value in the interval [0 1]. The above process is repeated to train several ANNCs and find the one with the best generalization.

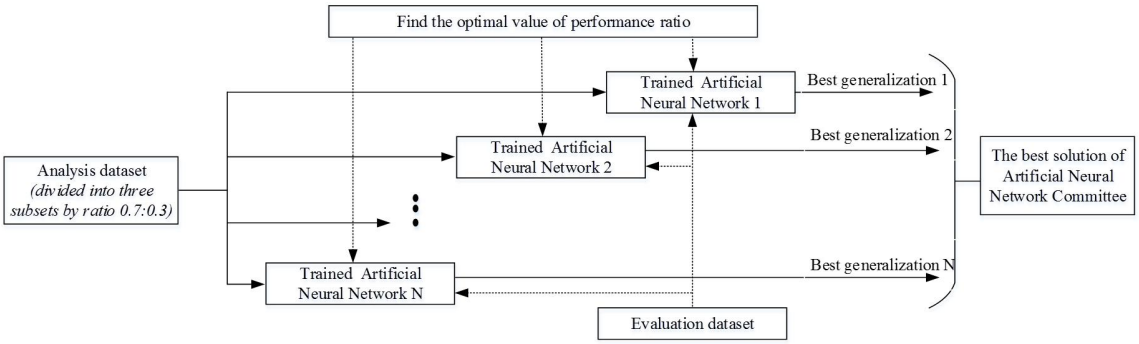


FIGURE 2-3. The algorithm for optimizing the artificial neural network committee.

2.4 Experimental results and discussion

The proposed method for speed-invariant fault diagnosis in bearings is validated using the datasets presented in Table 2-1. Each dataset is divided into two subsets: one for analysis and the other for evaluation. For each bearing condition, 50% of the AE data signals are used to construct the analysis dataset, while the remaining 50% are used to set up the evaluation dataset. The AE signals in the analysis set are decomposed into two, four and eight subband signals each. For each level of decomposition, 19 features are extracted

from each subband signal. These feature vectors are used to train the SVMCGs of the WCM using the OAA framework. The SVMCG reduces the 19 features into a single decision value that is more discriminative; this is then used to train the ANNC in the second stage, as discussed in Section 2.3. Figure 2-4 shows the four decision values from SVMCG 1 for the 240 AE signals in the analysis set, when each AE signal is decomposed into four subband signals.

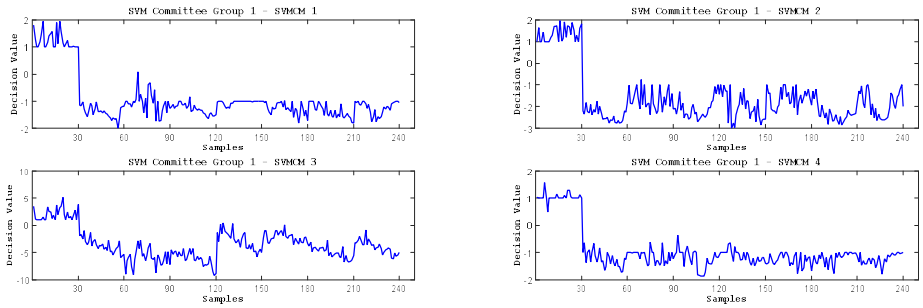


FIGURE 2-4 . The decision values of SVMCMs in SVMCG 1.

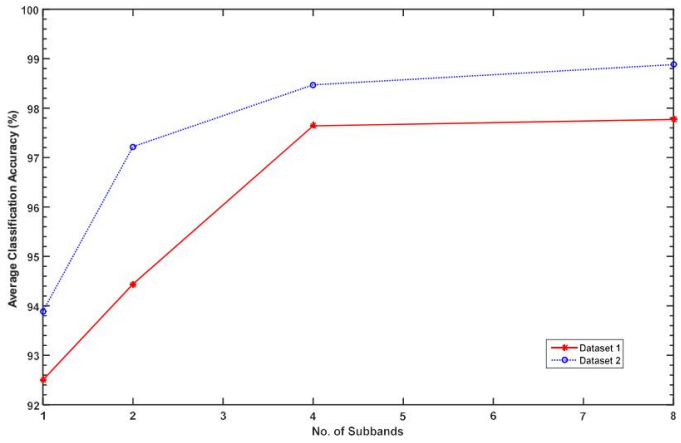


FIGURE 2-5. The performance of different number of subbands for datasets 1 and 2

The trained WCM is then used to classify the feature vectors extracted from different subbands of unknown AE signals in the evaluation dataset. The results in Table 2-4 indicate that the diagnostic performance of the proposed method, in terms of classification accuracy and sensitivity (i.e., the probability of correctly detecting a given fault), improves

significantly when the WCM is trained using features vectors that are calculated for the original AE signal at multiple resolutions. The proposed method correctly detects incipient bearing faults of two different dimensions with average classification accuracies of 97.64% and 98.64%, respectively, when the original AE signal is decomposed into four subbands. However, further decomposition of the AE signal does not result in any significant improvement in diagnostic performance, as shown in Fig. 2-5.

2.5 Conclusions

Existing AE-based methods for the diagnosis of incipient defects in bearings, especially techniques that use envelope analysis, are not invariant to changes in a bearing's rotational speed. This is the case because they rely on the detection of peaks at characteristic defect frequencies in the envelope power spectrum of the AE signals, and these defect frequencies vary with the bearing's rotational speed. In this letter, an improved feature extraction based method is proposed for the diagnosis of incipient defects in bearings, under variable speed conditions. The proposed method decomposes the non-stationary AE signal into different subbands using DWPT. It extracts 19 statistical features from each subband signal to form feature vectors, which are then used to train the weighted committee machine. The WCM is an ensemble of support vector machines and artificial neural networks. Our experimental results show that the proposed method can effectively diagnose both single and combined bearing defects under variable speed conditions. Moreover, the advantages of decomposing the AE signal into different subbands are not unlimited, i.e., the improvement in diagnostic performance wanes when the AE signal is decomposed into more than four subbands for feature extraction.

2.6 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. 20162220100050 and No. 20161120100350). It was also funded in part by The Leading Human Resource Training Program of Regional Neo industry through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (NRF-2016H1D5A1910564), in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the

Ministry of Education (2016R1D1A3B03931927), and in part by the development of a basic fusion technology in the electric power industry (Ministry of Trade, Industry & Energy, 201301010170D).

3 Fault Diagnosis of Bearings under Variable Speed using Convolutional Neural Networks and Energy Distribution Maps

Incipient bearing defects are primarily diagnosed by searching for peaks at bearing defect frequencies (BDFs) in the envelope power spectrum of acoustic emission (AE) signals, or by developing discriminative models for features extracted from the AE signals through time and frequency domain analysis. These methods perform poorly under variable speed conditions, as variation in a bearing's speed can drastically change BDFs, and result in bad discriminative models for the extracted features. This chapter proposes a novel method for diagnosing bearing defects under variable operating speeds using convolution neural networks (CNNs). The CNNs use the energy distribution maps of the AE signal spectrum as inputs, and automatically extract the optimal features that can be used to diagnose various single and compound bearing defects under variable speed conditions. The energy distribution map approximates the shape or the relative distribution of energy among different frequency bands of the AE spectrum for each defect type, and is not affected by variation in a bearing's speed. Experimental results demonstrate that the proposed method yields better diagnostic performance in comparison to state-of-the-art AE-based methods.

3.1 Introduction

Rolling element bearings are crucial components of rotating machines (Lu *et al.*, 2014), and are the leading cause of failure in essential industrial equipment such as rolling machines in paper mills, wind turbines and induction motors, where they account for 51% of total failures (Thorsen and Dalva, 1999; Lu *et al.*, 2016). A faulty bearing may cause equipment breakdown, which can lead to unscheduled and costly downtime for an entire industry; thus, the detection of incipient bearing faults is essential for maintaining production schedules and minimizing costs.

Various techniques have been used for fault detection in bearings including the analysis of vibration acceleration signals (Bediaga *et al.*, 2013; Immovilli *et al.*, 2013; Jin *et al.*, 2014; Seshadrinath *et al.*, 2014) and the motor stator current (Frosini and Bassi, 2010; Lau and Ngan, 2010; Elbouchikhi *et al.*, 2017). These techniques are effective in diagnosing bearing defects at high rotational speeds; however, incipient bearing defects at low rotational speeds, which are characterized by their low energy acoustic emissions (AE), can be more effectively diagnosed with AE-based methods (Caesarendra *et al.*, 2013; Niknam *et al.*, 2013; Kang *et al.*, 2015d; Nguyen *et al.*, 2015; Kang *et al.*, 2016). These methods can diagnose bearing defects before they appear on the bearing surface (Tandon and Choudhury, 1999). AE-based methods have predominantly utilized either envelope analysis (Kang *et al.*, 2015d) or feature-extraction-based techniques for bearing fault diagnosis (Kang *et al.*, 2015c; Kang *et al.*, 2016; Tra *et al.*, 2017).

Feature extraction-based methods for bearing fault diagnosis usually consist of three steps: signal acquisition, extraction of features from the acquired signal, and classification of various defect types based upon the extracted features. The extracted features consist primarily of statistical properties of the time and frequency domain AE signal, along with quantities calculated through the complex envelope analysis of the AE signal. Feature extraction is of great significance in these methods, as discriminative features yield discriminative models and hence good diagnostic performance. However, designing discriminative features not only requires significant domain knowledge, more often than not, feature selection algorithms are also required to eliminate irrelevant and redundant features that degrade the performance of these methods (Rauber *et al.*, 2015; Kang *et al.*, 2016). Thus, the diagnostic performance of these methods relies heavily on the quality of features used to construct discriminative models by the classifiers.

Envelope analysis-based methods diagnose bearing defects by detecting peaks at the characteristic bearing defect frequencies (BDFs) in the power spectrum of the complex envelope signal (Randall and Antoni, 2011; Bediaga *et al.*, 2013; Kang *et al.*, 2015b; Kang *et al.*, 2015d; Nguyen *et al.*, 2015). When a bearing is rotating at constant speed, the repeated excitation of a localized bearing defect results in transient impulses in the waveform of the AE signal that occur periodically or quasi-periodically. The frequency of

these impulses is referred to as the BDF, and is dependent upon the location of the bearing defect, the bearing geometry, and its rotational speed. The repeated impacts of a bearing's moving parts against bearing defects excite the resonant frequencies of adjacent components. These resonant frequencies have a high bandwidth and are modulated by the BDF, thus the resonant frequencies act as carrier frequencies, and the BDF acts as the modulation frequency (Randall and Antoni, 2011; Wang *et al.*, 2015). Through envelope analysis, the recorded AE signals are first demodulated, and then their power spectra are analyzed to detect the BDF associated with each type of localized defect.

In practice, bearing operation is mostly subject to variations in speed as, for example, in wind turbines. Both envelope analysis and feature extraction-based methods are susceptible to variations in the rotational speed of the bearing and hence are of limited utility under such conditions. As BDFs are functions of a bearing's rotational speed, any variation in the operating speed of a bearing results in translation of the power spectrum of the envelope signal along the frequency axis. Hence, traditional envelope analysis-based methods, which rely on detecting peaks at precise frequencies, do poorly under variable-speed conditions (Lacey, 2008). Similarly, feature extraction-based methods diagnose bearing defects by exploiting the variation between the features of the AE signals for specific defect types. For a given defect type at a constant rotational speed, the features extracted from the AE signals recorded at various times do not differ significantly, and serve to distinguish the defect type in question from AE signals for other defect types. However, variations in a bearing's speed can affect the amplitude of the impulses generated by repeated impact of moving parts with the defective areas. This can lead to significant variation in features, even for the same defect type (Lacey, 2008). Thus, AE signals captured at different rotating speeds will have different attributes, resulting in inconsistent feature vectors, hence limiting the direct application of these methods under variable speed conditions. Moreover, order analysis (OA) based methods have also been proposed to diagnose bearing defects under variable speed (Fyfe and Munck, 1997; Wang *et al.*, 2014a; Wang *et al.*, 2014b; Wang *et al.*, 2015), which use characteristic fault orders instead of BDFs to diagnose different bearing defects. These methods can be categorized into tachometer-based and tacho-less OA methods. Tachometer-based OA methods require additional speed sensors on the machine making them costlier, whereas tacho-less methods

are susceptible to noise as these methods estimate the speed from the acquired signals (Lu *et al.*, 2016). The authors in (Lu *et al.*, 2016) propose a tacho-less OA method that uses a high speed camera to estimate the rotating speed of the bearing; this technique also requires specialized hardware, which is costly and may not be suitable for general use.

Spectrum analysis of the AE signals suggests that although the BDFs and their amplitudes depend upon the bearing's rotational speed, the overall shape and relative distribution of energy across the spectrum, which is primarily defined by the relative position and amplitude of the BDF and its harmonics, is determined solely by the defect type. A bearing's rotational speed has only scaling and translational effects on the shape and relative distribution of energy in the spectrum. Hence, this study proposes using convolutional neural networks (CNNs) to extract distinguishing features from the energy distribution maps of the spectrum of an AE signal, and then using those automatically learned features to diagnose various bearing defects. CNNs are used for multiple reasons: First, they are unaffected by translation or local distortion of the inputs, they are shift invariant, and they exploit local structures and correlation in the input data by requiring the receptive fields of hidden units to be local (LeCun *et al.*, 1998). Although CNNs are traditionally used for automatic representation learning and classification of images, the strong local structure of the AE signal spectrum due to the high correlation between the amplitudes of nearby frequencies, and the scaling and translational effects of a variable rotating speed justify their use in bearing fault diagnosis. In this study, an energy distribution map of the frequency spectrum of an AE signal is constructed and then applied as input to a LeNet-5 architecture that automatically learns unique representations to classify various bearing defects under variable speed conditions. The contributions of this chapter are as follows:

1) A novel method is proposed for diagnosing bearing defects under variable speed conditions. The proposed method does not require any specialized hardware, rather it employs convolutional neural networks to mine the energy distribution maps for discriminative features, which can be used to diagnose bearing defects under variable operating speeds.

2) This work proposes the use of energy distribution maps of the AE signal spectrum as inputs to the CNNs for the diagnosis of bearing defects under variable operating speeds. The use of energy distribution maps as inputs to the CNNs is justified by a detailed analysis of the effects of speed on the AE signals, and a comparison of different types of inputs used to train the CNNs.

The rest of this chapter is organized as follows: Section 3.2 presents the experimental testbed and the data acquisition system used to collect the acoustic emission data. Section 3.3 presents the proposed method, introduces CNNs and the LeNet-5 architecture. In Section 3.4, the proposed method is validated using the AE data collected through the experimental testbed and its performance is compared with three state-of-the-art AE-based methods. Finally, conclusions are offered in Section 3.5.

3.2 The Experimental Testbed and the Acoustic Emission Data for Seeded Bearing Defects

The experimental testbed that is used to generate acoustic emission (AE) data for seeded bearing defects at variable operating speeds is shown in Fig. 3-1(a). There are two shafts in this setup, a drive end shaft (DES) and a non-drive end shaft (NDES). Both shafts are connected through a gearbox with a reduction ratio of 1.52:1, and are fastened at both ends using rolling elements bearings (FAG NJ206-E-TVP2). The DES is powered by a three-phase induction motor, which is operated at six different speeds, i.e., 250, 300, 350, 400, 450, and 500 revolutions per minute (RPM). The operating speed is measured using a displacement transducer installed on the NDES. The AE data is collected from bearings fastened to the NDES. These bearings are seeded with single and multiple-combined defects. A fan with adjustable blades is also coupled to the NDES to load the bearings; however, this study does not consider the effects of variations in load. The acoustic emissions are captured using a wide-band AE sensor that is coupled to the bearing housing at the NDES at a distance of 21.48 mm. Since, the AE activity is mostly in the ultrasonic range therefore the AE signals are recorded at a sampling rate of 250 kHz (ISO/TC 108, 2007) using a PCI-2 data acquisition system, as shown in Fig. 3-1(b).

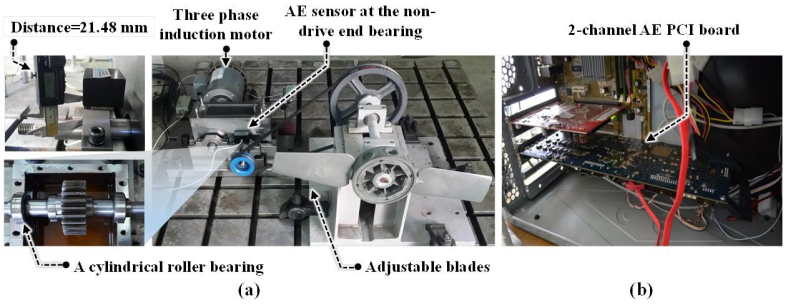


FIGURE 3-1. (a) The experimental testbed. (b) The data acquisition system

AE signals are recorded for a normal, defect-free bearing and bearings seeded with seven types of localized defects of two sizes (3 mm and 12 mm), as given in Table 3-1. The seven seeded defect types contain both single and multiple-combined defects: outer raceway crack (BCO), inner raceway crack (BCI), roller crack (BCR), inner and outer raceway cracks (BCIO), outer and roller cracks (BCOR), inner and roller cracks (BCIR), and inner, outer, and roller cracks (BCIOR), as shown in Fig. 3-2.

TABLE 3-1. Description of the datasets used to evaluate the proposed method

fs=250 kHz		Operating speed (RPM)	Crack size (mm)		
			length	width	depth
Dataset 1	Training subset	300, 400, 500	3.0	0.35	0.30
	Testing subset	250, 350, 450		0.35	0.30
Dataset 2	Training subset	300, 400, 500	12.0	0.49	0.50
	Testing subset	250, 350, 450		0.49	0.50

90 five-second AE data samples for each bearing condition (a defect-free bearing and 7 defective bearing s) at each RPM

The recorded AE signals are divided into two datasets, one for each crack size: 3 mm and 12 mm. Thus, each dataset has signals for eight bearing conditions, which are recorded at six operating speeds of 250, 300, 350, 400, 450, and 500 revolutions per minute (RPM). For the purpose of training and testing the CNNs, each dataset is divided into training and testing subsets, as given in Table 3-1. The training subset includes AE signals acquired at shaft speeds of 300, 400, and 500 RPM, while the testing subset includes the AE signals recorded at shaft speeds of 250, 350, and 450 RPM. For every operating speed, 90 AE signals are recorded for a duration of five seconds each. Hence, each dataset contains $N_{RPM} \times N_{Classes} \times N_{Signals}$, or 4320 AE signals, where N_{RPM} is the number of operating speeds for which the AE signals are recorded, $N_{RPM} = 6$; $N_{Classes}$ is the total number of

defect types or bearing conditions, $N_{Classes} = 8$; and, $N_{Signals}$ is the total number of AE signals recorded for each bearing condition at each shaft speed, $N_{Signals} = 90$.

FIGURE 3-2. The single and compound seeded bearing defects considered in this study, (a) BCI, (b) BCO, (c) BCR, (d) BCIO, (e) BCIR, (f) BCOR, and (g) BCIOR.

3.3 The Proposed Method for Bearing Fault Diagnosis under Variable Speed Conditions

The proposed method for the diagnosis of bearing faults under variable speed conditions is illustrated in Fig. 3-3. It consists of an offline training process in which the CNNs are trained using the training subset, and an online fault diagnosis process where a trained CNNs are used to diagnose bearing defects. The terms online and offline are not used to imply the proposed scheme's performance in terms of time, rather these terms indicate the availability of the CNNs for bearing fault diagnosis. As discussed in Section 3.2, the experimental testbed is used to acquire 1-D AE signals, which are recorded at ultra-high sampling rates (i.e., sampling rate of 250 kHz). Mining these raw AE signals for meaningful and distinctive features would require prohibitively large CNNs. Hence, in the proposed scheme the raw AE signals are first converted into 2-D energy distribution maps, which are then used as inputs to the CNNs. These maps show the distribution of energies in various frequency bands of the AE signal and roughly approximate the shape of the AE spectrum. The energy distribution maps are constructed by first multiplying the AE signal with a Hanning window function, and then computing its fast Fourier transform (FFT). The resulting 1-D spectrum of the AE signal is split into 1024 frequency bands, and then the root mean square (RMS) value is calculated for each of these frequency bands. The RMS value approximates the energy carried by each frequency band (Abu-Mahfouz, 2005). These RMS values are then stacked on top of each other to create a 2-D array of size 32 x 32. This 2-D array shows the distribution of energies across the entire spectrum of the AE signal, and hence it is called the energy distribution map. In the offline training process,

these energy distribution maps of the AE signal spectrum are used as inputs to train the CNNs. The CNNs can automatically learn higher order discriminative features from its inputs and then use them to for classification. In the online phase, the proposed method employs the trained CNNs to diagnose various single and compound bearing defects under variable operating speeds using the energy distribution maps of the AE signals from the testing subset as inputs.proposed method for the diagnosis of bearing faults under variable speed conditions is illustrated in Fig. 3-3.

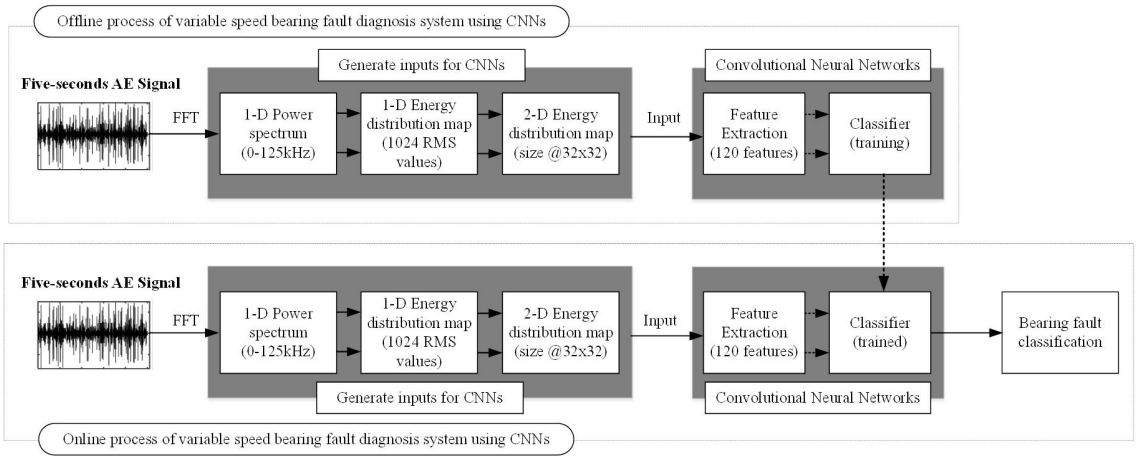


FIGURE 3-3. The proposed method for bearing fault diagnosis under variable speed conditions.

It consists of an offline training process in which the CNNs are trained using the training subset, and an online fault diagnosis process where a trained CNNs are used to diagnose bearing defects. The terms online and offline are not used to imply the proposed scheme's performance in terms of time, rather these terms indicate the availability of the CNNs for bearing fault diagnosis. As discussed in Section 3.2, the experimental testbed is used to acquire 1-D AE signals, which are recorded at ultra-high sampling rates (i.e., sampling rate of 250 kHz). Mining these raw AE signals for meaningful and distinctive features would require prohibitively large CNNs. Hence, in the proposed scheme the raw AE signals are first converted into 2-D energy distribution maps, which are then used as inputs to the CNNs. These maps show the distribution of energies in various frequency bands of the AE signal and roughly approximate the shape of the AE spectrum. The energy distribution maps are constructed by first multiplying the AE signal with a Hanning window function, and then computing its fast Fourier transform (FFT). The resulting 1-D spectrum of the AE

signal is split into 1024 frequency bands, and then the root mean square (RMS) value is calculated for each of these frequency bands. The RMS value approximates the energy carried by each frequency band (Abu-Mahfouz, 2005). These RMS values are then stacked on top of each other to create a 2-D array of size 32 x 32. This 2-D array shows the distribution of energies across the entire spectrum of the AE signal, and hence it is called the energy distribution map. In the offline training process, these energy distribution maps of the AE signal spectrum are used as inputs to train the CNNs. The CNNs can automatically learn higher order discriminative features from its inputs and then use them to for classification. In the online phase, the proposed method employs the trained CNNs to diagnose various single and compound bearing defects under variable operating speeds using the energy distribution maps of the AE signals from the testing subset as inputs.

3.3.1 Effect of Speed on the Bearing Fault Signals

Bearing faults are localized mechanical defects in a bearing's material, and their initiation and growth are associated with the sudden release of energy in the form of acoustic emissions (ISO/TC 108, 2007). AE activity occurs whenever the rolling elements of a bearing pass over a localized defect. For example, a defect on a bearing's outer raceway would generate high energy impulses at the same rate as the ball pass frequency of the outer raceway (BPFO). As a bearing's outer raceway is always stationary, the amplitude of these impulses will not vary significantly over time, and hence these periodic impulses would appear as a single peak in the frequency spectrum.

Likewise, a localized defect on a bearing's inner raceway would produce high energy impulses at the same rate as the ball pass frequency of the inner raceway (BPFI). However, because the inner raceway is a rotating element, the defect periodically enters and leaves the load zone. This leads to variation in the contact force between the rolling elements and the inner raceway, resulting in impulses with high amplitudes at maximum contact force, and impulses with lower amplitudes when the contact force decreases. Thus, the amplitude of the impulses fluctuates with the rotational frequency of the inner raceway. More precisely, the BPFI is amplitude modulated at the rotational frequency of the inner raceway. Similarly, the ball spin frequency (BSF) is amplitude modulated at the fundamental train frequency (FTF), as a roller defect moves in and out of the load zone at the FTF, resulting

in sidebands around the BSF (Lacey, 2008; Randall and Antoni, 2011; Bediaga *et al.*, 2013).

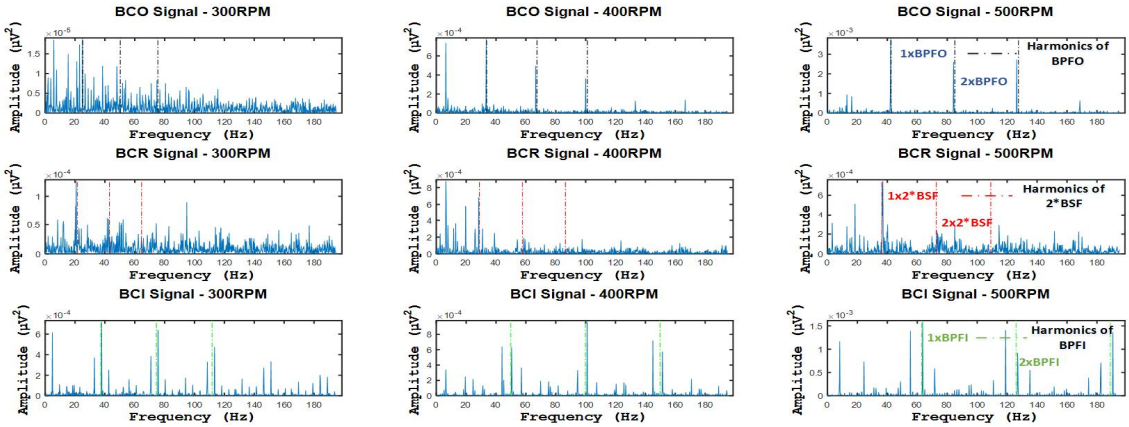


FIGURE 3-4. Envelope power spectra of fault signals under various speed conditions. The bearing defect frequencies increase with increasing shaft speed

$$
\begin{aligned}
BPFI &= \frac{N_{roller}.F_{shaft}}{2}\left(1+\frac{B_d}{P_d}\cos\alpha\right),\\
BPFO &= \frac{N_{roller}.F_{shaft}}{2}\left(1-\frac{B_d}{P_d}\cos\alpha\right),\\
BSF &= \frac{P_d.F_{shaft}}{2.B_d}\left(1-\left(\frac{B_d}{P_d}\cos\alpha\right)^2\right),\\
FTF &= \frac{F_{shaft}}{2}\left(1-\frac{B_d}{P_d}\cos\alpha\right),
\end{aligned}
\tag{3.1}
$$

For a given bearing with known geometry and shaft speed, these characteristic frequencies can be calculated using Eq. (3.1) (Bediaga *et al.*, 2013), where F_{shaft} is the shaft speed in hertz, $N_{rollers}$ is the number of rolling elements in the bearing, α is the contact angle, B_d is the roller diameter, and P_d is the pitch diameter. The magnitude of these characteristic frequencies increases when the shaft speed of a bearing increases, as shown in Fig. 3-4. The characteristic frequencies change with variations in shaft speed, resulting in poor diagnostic performance of envelope analysis-based methods. Moreover, the amplitude of AE signals also increases with increasing shaft speed, as shown in Fig. 3-5; hence, variations in shaft speed lead to inconsistent feature vectors that adversely affect

the diagnostic performance of feature-extraction-based methods.

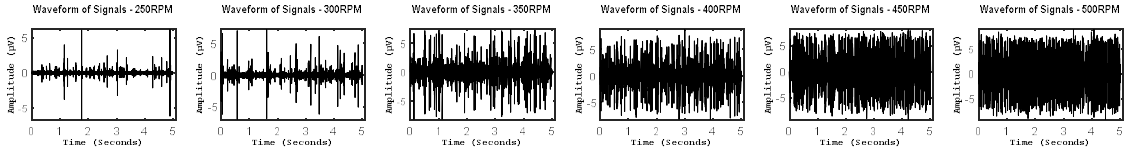


FIGURE 3-5. Time domain waveforms of AE signals for BCO at different shaft speeds. The pulse amplitude of the AE signal increases as the speed of the bearing shaft increases.

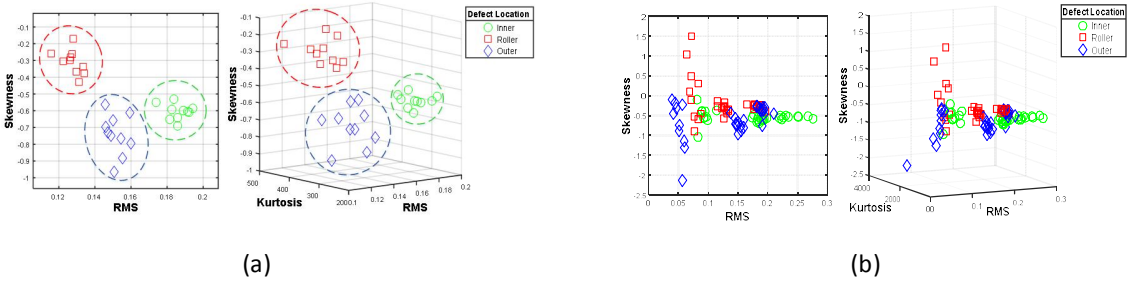


FIGURE 3-6. Feature spaces of Skewness and RMS, and Skewness, RMS and Kurtosis for BCI, BCR, and BCO (a) Under constant speed (b) under variable speed.

This is further illustrated in Fig. 3-6(a) and Fig. 3-6(b), which show the distribution of different features for BCI, BCR, and BCO under constant and variable speed conditions, respectively. As shown in Fig. 3-6(a), for aconstant bearing speed, the features for different bearing defects are easily separable and any good classifier can be used to generate decision boundaries for different fault classes, which can then be used to diagnose various bearing defects. However, when there are variations in the bearing's speed then the features for different bearing defects overlap as shown in Fig. 3-6(b), which degrades the diagnostic performance of the classifier.

In contrast, it can be observed in Fig. 3-7 that the shape and hence the relative energy distribution across the AE spectrum, does not change significantly with changes in shaft speed. Rather, variation in shaft speed only scales the amplitude of the spectrum and slightly translates it along the frequency axis. CNNs are invariant to shifting and translation in their inputs. They can exploit the local structure in the input and hence can be used to automatically learn the features required to distinguish different bearing defects based upon the shapes and therefore the relative energy distribution of their spectra. Hence, the 2-D

energy distribution maps of the spectrum of the AE signal are used as input to the CNNs to detect bearing defects under variable speed conditions.

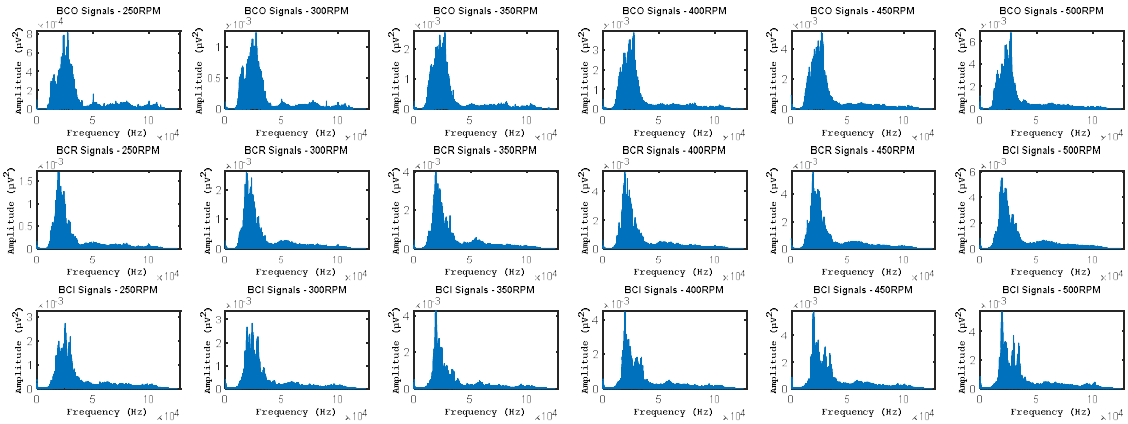


FIGURE 3-7. Single-band power spectrum of AE signals for different bearing defects and at different operating speeds

3.3.2 Energy Distribution Maps of the AE Spectrum

As discussed in Section 3.1 and illustrated in Fig. 3-7, for a given fault the shape and relative energy distribution of the AE spectrum does not change significantly with variation in a bearing's speed. The 1-D AE spectra shown in Fig. 3-7 are therefore converted into 2-D energy distribution maps of practically reasonable dimensions as CNNs are designed to work on two-dimensional data. This conversion process is illustrated in Fig. 3-8.

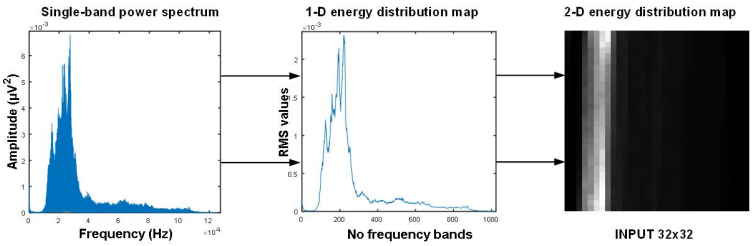


FIGURE 3-8. The process of converting a 1-D AE spectrum into a 2-D energy distribution map.

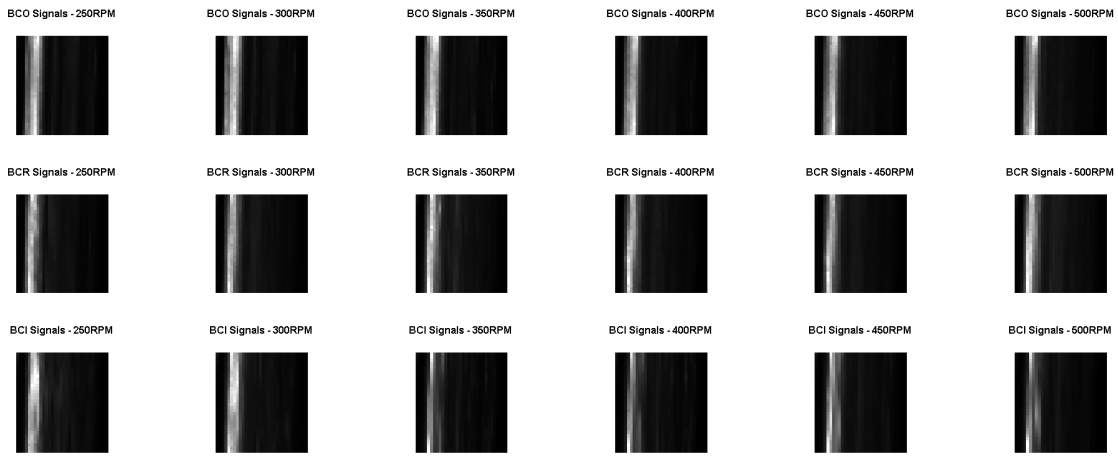


FIGURE 3-9. The 2-D energy distribution maps for BCO, BCR and BCI faults at bearing speeds of 250, 300, 350, 400, 450 and 500 RPM

The AE signals are first multiplied by a Hanning window function, before computing their fast Fourier transform (FFT), to avoid spectral leakage. The spectrum of the AE signal is then split into an appropriate number of frequency bands (i.e., 1024 frequency bands because in this study, the CNNs uses inputs of size 32 x 32 = 1024), and the root mean square (RMS) value of each of these bands is calculated to generate a 1-D energy distribution map. The RMS value gives an approximation of the energy carried by each band. The 1-D energy distribution map preserves the shape of the original AE spectrum. The RMS values in the 1-D map are then stacked to generate a 2-D array. This 2-D array or energy distribution map of the AE signal spectrum is then used as input for the CNNs. Fig. 3-9 shows examples of 2-D energy distribution maps for different types of bearing faults at different speeds.

3.3.3 Overview of CNNs and LeNet-5

CNNs are feedforward artificial neural networks (ANNs) inspired by the mammalian visual cortex, and have been used to great effect for classification tasks where it is difficult to design optimal features such as character recognition, image, and video analysis (LeCun *et al.*, 2010). CNNs combine three architectural concepts, i.e., using local receptive fields instead of fully-connected layers, weight sharing, and spatial or temporal sub-sampling, to achieve a reasonable degree of invariance to shifting, scaling, and distortion in images

(LeCun *et al.*, 1998). Fig. 3-10 shows the architecture of a typical CNNs, the LeNet-5, which was initially proposed for character recognition (LeCun *et al.*, 1998). In this architecture, consecutive layers are not fully connected as in typical ANNs, rather every unit in a particular layer is connected only to a small neighborhood of units or a local receptive field in the preceding layer. The neurons in a particular layer extract simple features from the receptive fields in the previous layers, and this process continues in subsequent layers, which combine those simple features to learn higher-order representations of the input. Any distortion or shift in the inputs alters the position of these simple or salient features. To make the CNNs invariant to these shifts and distortions in the inputs, a weight-sharing mechanism is used. Using this mechanism, units with local receptive fields in separate locations share identical weight vectors, essentially forcing them to perform the same operation and hence extract the same feature from the entire input field. These units thus construct a feature map for the entire input. In fact, using different weight vectors and hence different operations, a convolutional layer constructs several feature maps of the input, where multiple features are extracted from each local receptive field.

The original LeNet-5 architecture is comprised of seven layers excluding the input layer. Layers C1, C3, and C5 are the convolutional layers that learn a hierarchy of features from the raw inputs, whereas layers S2 and S4 are the sub-sampling layers that reduce the spatial resolution of the feature maps learned by the convolutional layers and make the CNNs robust to shifts and distortions in the input. The fully connected layers, F6 and F7, carry out the classification of inputs using the highest order features learned by C5. These layers have weights and biases, which are tuned using training data. The first convolutional layer C1, as shown in Fig. 3-10, has six feature maps and each unit in each of those six feature maps is connected to a receptive field of 5x5 inputs. Since all the units in a feature map use weight sharing, they extract the same feature from various locations throughout the input field. Hence, these six feature maps extract six unique features by sequentially scanning all the local receptive fields in the entire input plane; this operation is equivalent to convolution, and thus these networks are called CNNs.

In the convolutional layers, any shift in the input results in the same amount of shift in

the feature map output. In Fig. 3-10, the second layer, S2, is the sub-sampling layer that reduces the spatial resolution of the feature map, thereby decreasing the precision with which it encodes the position of the features and making it robust to shifts and distortions in the input (LeCun *et al.*, 1998). Corresponding to each of the six feature maps in the previous convolutional layer, the sub-sampling layer, S2, has a feature map with units that are connected to 2x2 neighborhoods or receptive fields in the previous layer. Sub-sampling is performed by computing the average of each 2x2 neighborhood, which is then multiplied by a weight and added to with a bias; the weights and biases are tuned using training data. Compared to the feature maps in the previous layer, the number of rows and columns is halved in each feature map of the sub-sampling layer, as its contiguous units have contiguous but non-overlapping receptive fields in the previous layer.

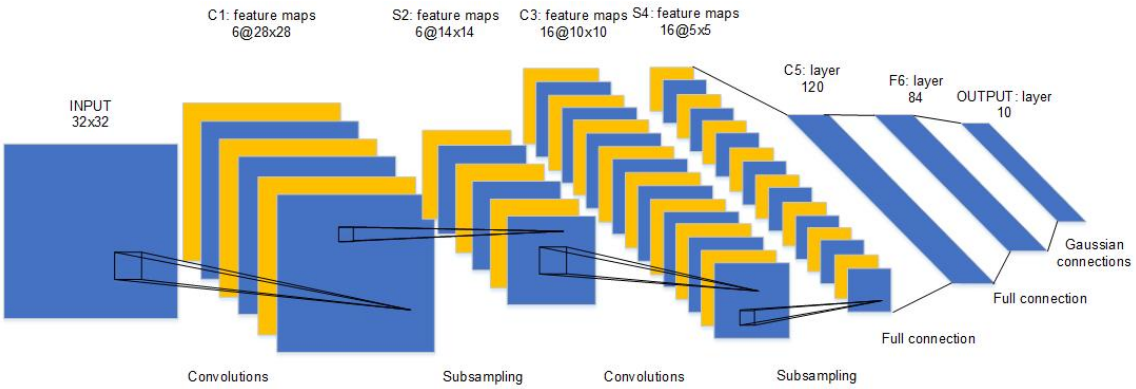


FIGURE 3-10. Architecture of a convolutional neural networks, the LeNet-5

The sub-sampling layer is followed by a convolutional layer, C3, with 16 feature maps, where each unit in every feature map is connected to multiple 5x5 neighborhoods. These neighborhoods belong to a unique subset of the feature maps of S2, and occur at identical locations within those feature maps. As each feature map of C3 is only connected to a small subset of S2 feature maps, the number of connections and hence tunable parameters are kept within reasonable bounds. Furthermore, connections to different inputs constrain the feature maps to learn new features. The output of layer C3 is then sub-sampled by the layer S4, which has sixteen feature maps with 5x5 units each. Layer S4 functions in the same manner as S2 where the units in each of its feature maps are connected to 2x2 neighborhoods in the corresponding feature map in C3. The convolutional layer C5, with

120 feature maps, then learns features from all sixteen feature maps of S4.

The size of each feature map in C5 is 1x1; thus the output of C5 is one dimensional and represents the 120 highest-order features that can be used by the classifier (the fully-connected layers, F6 and F7, with 84 and eight output units, respectively) to distinguish between various bearing defects under variable speed conditions. Fig. 3-11 shows twenty four of these 120 higher order features extracted by the CNNs for different bearing faults at different operating speeds. In the standard LeNet-5 the output layer, F7 has a Euclidean Radial Basis Function (RBF) unit for each class. In this work, the RBF units in the last layer are replaced with a fully connected layer with eight output units corresponding to the eight bearing defects, to simplify the training process (Bouvrie, 2006).

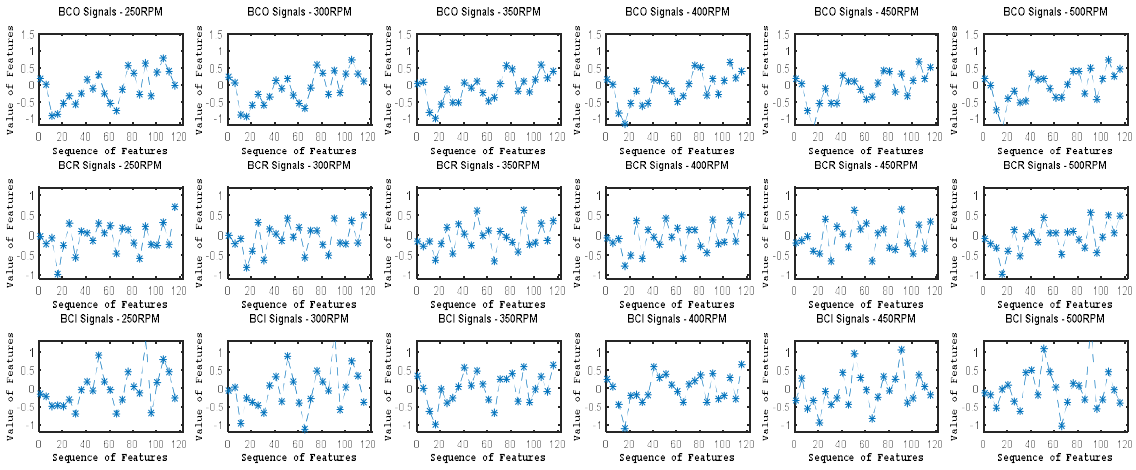


FIGURE 3-11. The final set of higher-order features extracted by the CNNs for different bearing faults under variable operating speeds.

3.3.4 Back-propagation through CNNs

The In this study, the stochastic back-propagation algorithm is used to update the tunable parameters of all units across all layers of the CNN using training data (Bouvrie, 2006; LeCun *et al.*, 2010; LeCun *et al.*, 2012). The error is calculated for each input pattern in the training subset using the squared-error loss function given in Eq. (3.2), which gives the error for the p^{th} pattern in a multiclass problem with c classes.

$$E^p = \frac{1}{2}\sum_{k=1}^{c}\left(d_k^p - o_k^p\right)^2, \tag{3.2}$$

where, d_k^p is the k^{th} component of the target and o_k^p is the value of the k^{th} output, both corresponding to the p^{th} input pattern. The output of each convolution layer is determined by first multiplying the input feature maps with tunable kernels, and then applying an activation function to the sum of this product and a bias. Let ℓ denote the current layers. For the ℓ^{th} convolutional layer in CNNs, the output feature map is calculated as follows:

$$x_j^\ell = f\left(\sum_{i\in M_j} x_i^{\ell-1} * k_{ij}^\ell + b_j^\ell\right), \tag{3.3}$$

where j represents the j^{th} output feature map, i the i^{th} input feature map, M_j the subset of feature maps from the previous layer that is used as input to the current layer, k_{ij} is the learnable convolutional kernel for the j^{th} feature map, b is the additive bias, and f is an activation function. Each output feature map is generated by convolving the inputs with different kernels that are learned during training.

The sub-sampling layers do not learn any new features, instead only reduce the spatial resolution of the input feature maps. While the sub-sampling layers output the same number of feature maps as they receive at the input, the output feature maps are smaller in size than the input. The mathematical form of the sub-sampling operation is given in Eq. (3.4):

$$x_j^\ell = f\left(\beta_j^\ell down(x_j^{\ell-1}) + b_j^\ell\right), \tag{3.4}$$

where β is a tunable coefficient, b is the additive bias, and $down(.)$ denotes the sub-sampling function. The coefficient β and the bias b are tuned using training data. The sub-sampling function calculates the average of distinct blocks of size $n\times n$ in the input feature map; this results in output feature maps that have n times fewer rows and columns.

The last two layers of the CNNs, F6 and F7, are fully connected 1-D layers that diagnose the bearing defects using the high-order features extracted by the CNNs, the

output of layer C5. Layer F7 is the final output layer of the network with P units to distinguish P defect types. The output of this layer is calculated using Eq. (3.5),

$$O^{\ell} = f\left(W^{\ell} x^{\ell-1} + b^{\ell}\right), \tag{3.5}$$

where W and b are the weight and bias vectors, respectively, and $f(.)$ is the activation function. The most commonly used activation functions are the logistic and hyperbolic tangent functions (Bouvrie, 2006).

3.4 Experimental Results

As mentioned in Section 3.3, CNNs automatically learn the optimal features required to distinguish different bearing defects. CNNs view the input data as a hierarchical composition of features, and therefore use multiple layers to successively construct higher-order features. An example of the higher order features extracted by successive convolutional layers of CNNs from the input 2-D energy distribution map is shown in Fig. 3-12. The final set of higher-order features is the output of the convolutional layer, C5, which is a 1-D vector with 120 features that is used by the classifier, i.e., the fully-connected layers, F6 and F7, to distinguish various bearing defects under variable speed conditions. Fig. 3-11 shows examples of these optimal features for outer race, roller, and inner race faults at different speeds. The performance of any diagnostic system depends to a large extent on the features of the measurement signal that are selected for distinguishing different faults. In the previous sections, an effort was made to show the ineffectiveness of traditional envelope analysis and feature extraction based methods for diagnosing bearing defects under variable speed conditions. Moreover, the use of 2-D energy distribution maps was advocated as a more discriminant representation of the AE fault signals, which can be used as input to CNNs and classify bearing defects under variable operating speeds.

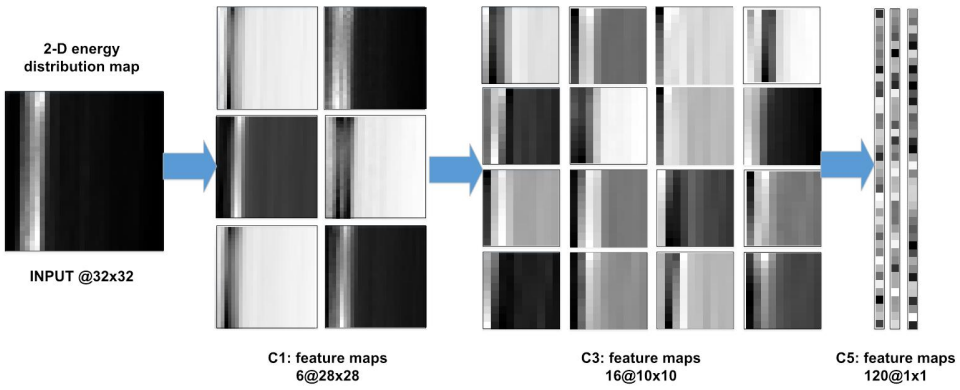
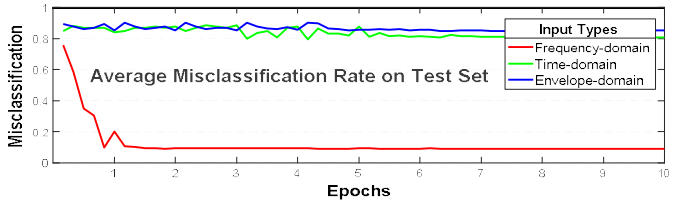


FIGURE 3-12. The representations of input at convolutional layers.

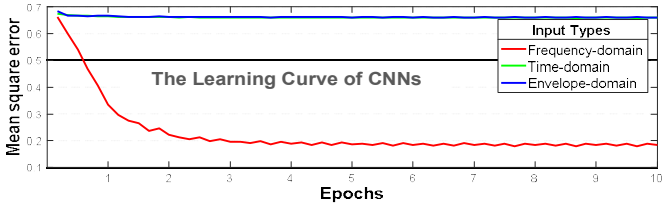
The effectiveness of 2-D energy distribution maps as input to the CNNs is further demonstrated by comparing its performance with two more input types, i.e., the time-domain AE signal and the envelope power spectrum of the AE signal; both of which are first converted into 2-D arrays of size 32 x 32 using the procedure shown in Fig. 3-8. For all the three types of input to the CNNs, the mean square error (MSE) plots for training samples, i.e., the learning curves, and the misclassification rate (MCR) plots for the test samples of dataset 1 are given in Fig. 3-13. The learning curves indicate that the best training results for CNNs are obtained when the 2-D energy distribution maps, which are constructed from the frequency spectrum of the AE signals, are used as input as compared to the other two input types. Similarly, the MCR plots for the test samples indicate that when the CNNs are trained on the proposed 2-D energy distribution maps as input, they yield MCR values closer to zero after just two training epochs. Whereas, in the case of the other two input types, the MCR values are closer to 1 and do not improve with more training epochs.

The effectiveness of the proposed method in diagnosing single and compound bearing defects at variable speeds is validated by comparing its performance with other AE-based methods for bearing fault diagnosis proposed in (Kang *et al.*, 2016), (Tra *et al.*, 2017), and (Kang *et al.*, 2015c). In these methods, various statistical measures of the time and frequency domain AE signals are used as features, along with features extracted from the complex envelope signal (Kang *et al.*, 2016; Tra *et al.*, 2017). Moreover, before feature extraction, the study in (Kang *et al.*, 2015c; Tra *et al.*, 2017) also considers the subband analysis of the AE signals using the discrete wavelet packet transform (DWPT). The

extracted features are then evaluated and selected using feature selection algorithms such as outlier-insensitive hybrid feature selection (OIHFS) (Kang *et al.*, 2016) and the genetic algorithm (Kang *et al.*, 2015c). Finally, the selected features are used to detect bearing defects using classifiers such as K-NN (Kang *et al.*, 2016), SVM (Kang *et al.*, 2015c; Tra *et al.*, 2017) and weighted committee machines (Tra *et al.*, 2017).



(a)



(b)

FIGURE 3-13. (a) Training mean square error, and (b) Testing misclassification rate, for different input types.

TABLE 3-2. A Comparison of the Average Classification Accuracy and Sensitivity for each Bearing Defect under Variable Speed Conditions for the Proposed Method and Existing AE Based Methods

Datasets	Methodologies	Average sensitivity of each fault type								ACA (%)
		BCI	BCO	BCR	BCIO	BCIR	BCOR	BCIOR	BNC	
Dataset 1	(Kang *et al.*, 2016)	19.62	47.40	75.18	17.03	59.62	30.74	10	3.33	32.87
	(Tra *et al.*, 2017)	11.11	13.33	100	100	97.77	97.77	0	0	52.5
	(Kang *et al.*, 2015c)	83.70	100	63.70	66.66	95.92	60	98.88	67.77	79.58
	Proposed method	66.66	97.77	100	96.29	99.25	98.14	99.62	97.40	94.39
Dataset 2	(Kang *et al.*, 2016)	7.03	70	66.66	79.62	5.92	44.81	74.07	62.96	51.38
	(Tra *et al.*, 2017)	100	100	97.77	97.77	100	100	0	0	74.44
	(Kang *et al.*, 2015c)	100	73.33	88.14	70	45.55	66.29	98.51	100	80.23
	Proposed method	100	80.74	75.55	100	84.07	100	100	98.51	92.36

Table 3-2 presents the diagnostic performance of the proposed method and the methodologies proposed in (Kang *et al.*, 2016), (Tra *et al.*, 2017) and (Kang *et al.*, 2015c) for each dataset in terms of the average classification accuracy (ACA), which is defined as follows:

$$ACA = \frac{\sum_{N_{classes}} N_{TP}}{N_{testdata}} \times 100(\%), \tag{3.6}$$

where $N_{testdata}$ is the total number of data points used to test the classification accuracy of the proposed method, $N_{classes}$ is the total number of defect types (eight in this study including seven single and compound bearing defects and a defect-free condition), and N_{TP} is the number of data points in class *i* that are correctly classified as class *i* . In addition, sensitivity, which is a useful metric for evaluating the diagnostic performance of each bearing condition, is also provided in Table 3-2, and is defined as:

$$Sensitivity = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100(\%), \tag{3.7}$$

where N_{FN} is the number of data points in class *i* that are classified incorrectly. The results in Table 3-2 show that the proposed method delivers better diagnostic performance than the methods in (Kang *et al.*, 2016), (Tra *et al.*, 2017) and (Kang *et al.*, 2015c). The average classification accuracy of the proposed method for both datasets is 93.4%, which is significantly better than that of the methods proposed in (Kang *et al.*, 2016), (Tra *et al.*, 2017) and (Kang *et al.*, 2015c), i.e., 42.12%, 63.47%, and 79.9%, respectively.

3.5 Conclusion

Existing AE-based methods for the diagnosis of incipient defects in bearings do not perform well under variable operating speeds, because the diagnostic performance of these methods depends upon either the precise position of bearing defect frequencies in the envelope power spectrum or on the quality and consistency of the features extracted from the AE signals. Variation in the shaft speed of a bearing changes the bearing defect frequencies and affects the quality and consistency of the AE signal features, which are

typically used to create discriminative models for diagnosing different bearing defects. However, for a particular bearing defect, these speed variations do not alter the shape and the relative distribution of energies across the spectrum of the AE signal. The shape of the frequency spectrum of the AE signals is primarily defined by the defect frequencies and their harmonics. The relative position of BDFs and their harmonics does not change even if there are changes in the BDFs; similarly, the relative distribution of energies between these frequencies does not change significantly. Hence, in this study we use the energy distribution maps of the AE signal spectra as inputs to CNNs to automatically extract higher-order discriminative features and classify various bearing defects under variable operating speeds. The proposed method uses CNNs, as they focus on the topology of the inputs and are immune to any translation or local distortions in the inputs. The proposed method does not extract or select any features from the input. The CNNs can automatically learn the optimal features to diagnose both single and combined bearing defects. The proposed method is also compared with three state-of-the-art AE-based methods that use traditional approaches for the diagnosis of incipient bearing defects, and yields an average classification accuracy of 93.4% for all the datasets used in this study, which is significantly better than the existing methods. These results lead us to conclude that energy distribution maps of the AE spectrum and CNNs are effective in diagnosing both single and compound bearing defects under variable operating speeds. This study considered 100% variation in the operating speeds of bearings from 250 to 500 revolutions per minute.

3.6 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. 20162220100050, No. 20161120100350, and No. 20172510102130). It was also funded in part by The Leading Human Resource Training Program of the Regional Neo industry through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (NRF-2016H1D5A1910564) and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A3B03931927).

4 Bearing Fault Diagnosis under Variable Speed using Convolutional Neural Networks and the Stochastic Diagonal Levenberg-Marquardt Algorithm

This chapter represents a novel method in previous chapter for diagnosing incipient bearing defects under variable operating speeds using convolutional neural networks (CNNs) trained via the stochastic diagonal Levenberg-Marquardt (S-DLM) algorithm. The CNNs utilize the spectral energy maps (SEMs) of the acoustic emission (AE) signals as inputs and automatically learn the optimal features, which yield the best discriminative models for diagnosing incipient bearing defects under variable operating speeds. This hypothesis is confirmed by the experimental results, where CNNs trained using the S-DLM algorithm yield significantly better diagnostic performance under variable operating speeds compared to existing methods. In this study, the performance of different training algorithms is also evaluated to select the best training algorithm for the CNNs. The proposed method is used to diagnose both single and compound defects at six different operating speeds.

4.1 Introduction

This study uses the LeNet-5 architecture to automatically learn unique representations from the input SEMs for classifying various single and compound bearing defects. The performance of LeNet-5 is improved by finding the best training algorithm. A popular choice for training CNNs is the first order gradient descent (GD) algorithm, but it converges rather slowly and in the case of a non-quadratic error surface, the GD algorithm can get stuck in local minima (Osborne, 1992). The convergence speed of the GD is usually enhanced by using the Gauss-Newton algorithm, which approximates the error function using an appropriate quadratic function to determine the optimal step size in each direction

(Osborne, 1992). However, the selection of an inappropriate quadratic function can cause the Gauss-Newton algorithm to diverge. The Levenberg-Marquardt algorithm is a robust training algorithm that combines the Gauss-Newton and the steepest descent methods to exploit the speed advantage of the former and the stability of the later. It yields better convergence results, even for complex non-quadratic error functions.

However, the Levenberg-Marquardt algorithm requires calculation of the Hessian matrix and its inverse, which is computationally very expensive, especially for large problems (LeCun *et al.*, 2012). Therefore, we use the stochastic diagonal Levenberg-Marquardt (S-DLM) algorithm to train the CNNs, as it simplifies the computation of the Hessian and its inverse as well as speeds up the training process. The S-DLM uses the diagonal terms of the Hessian matrix to compute the individual learning rates of all parameters of the CNNs before scanning through the training set (LeCun *et al.*, 1998). Our empirical results indicate that CNNs converge quickly when trained using the S-DLM algorithm, thereby enhancing the performance of the proposed fault diagnosis system. The main contributions of this chapter include the following.

1) This work also investigates various training algorithms for CNNs and proposes the use of S-DLM algorithm for training the CNNs as it results in faster convergence and a better diagnostic performance.

The rest of this chapter is organized as follows: Section 4.2 presents the proposed method, CNNs, the LeNet-5 architecture, and stochastic diagonal Levenberg-Marquardt algorithm. Section 4.3 provides details of the experiments carried out to validate the proposed method, and Section 4.4 finally concludes this study.

4.2 The Proposed Method for Diagnosing Bearing Defects under Variable Speeds using CNNs and the Stochastic Diagonal Levenberg-Marquardt Algorithm

The proposed method for bearing fault diagnosis under variable operating speeds is illustrated in Fig. 4-1. It uses CNNs because they are very good at automatically learning

distinctive features from their inputs that result in the best classification performance. The proposed method uses a representation learning technique, such as CNNs, because it is difficult to manually design features of the AE signal that would be effective in diagnosing both simple and compound bearing defects under variable operating speeds. This point is further illustrated in Section 4.3. CNNs are very good at learning distinctive representations of their inputs, nevertheless, they come with certain caveats; first, the input to the CNN must have strong spatial correlation, and second, it must have reasonable size or dimensions, i.e., input with large dimensions would require more training data and prolonged training times that would restrict their practical use. A five second AE signal that is sampled at 250 KHz has 1.25 x 106 samples, and hence cannot be used as input to the CNN. Therefore, just as important it is to use CNNs, an even more important concern is to decide the appropriate type and size of input for the CNN that can be effective in diagnosing bearing defects under variable operating speeds. We propose to use 2-D spectral energy maps or SEMs as inputs to the CNN. The intuition behind using SEMs as inputs to the CNNs is that for a given fault type, the shape of the AE power spectrum and SEM doesn't change much with changes in the rotational speed of the bearing, and hence SEMs can be used as inputs to the CNNs to diagnose bearing defects when there are variations in rotational speed. This is further illustrated in Section 4.3.

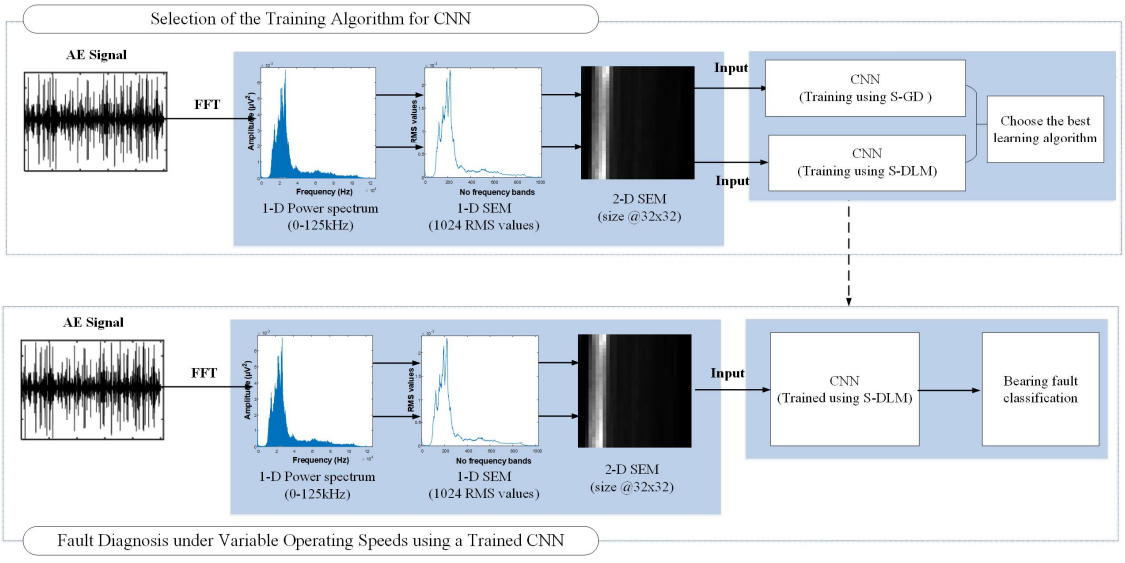


FIGURE 4-1. The proposed method for the diagnosis of single and compound bearing defects usin g acoustic emission signals under variable operating speeds.

In the proposed method, the raw AE signals are first converted into 2-D SEMs. First, the fast Fourier transform (FFT) of the raw AE signals is computed to get their frequency content. As discussed earlier, for a given fault type, the shape of the AE spectrum is not significantly altered by changes in the bearing's rotational speed. The AE spectrum has 1.25 x 105 frequency components and it is also not a suitable input to the CNN. Therefore, the AE spectrum is divided into 210 frequency bands and for each of these bands the root mean square (RMS) frequency is calculated, resulting in a 1-D SEM with RMS values for 1024 frequency bands. These RMS values approximate the energy carried by each band. The RMS values in the 1-D SEM are stacked on top of each other to create a 2-D SEM of size 32x32. The 2-D SEM has reasonable dimensions and good spatial correlation making it a suitable input for the CNN. These 2-D SEMs are then used to train the CNNs using different training algorithms as discussed in Section 4.3. The CNNs trained with the best training algorithm, i.e., S-DLM are then used for diagnosing bearing defects under variable operating speeds.

CNNs are trained using learning algorithms such as the Levenberg-Marquardt algorithm, which determines the optimal network parameters, i.e., weights and biases. These algorithms are locally adaptive, as they determine the learning rates for each weight and bias in the network by considering both the gradient and curvature of the error function (LeCun *et al.*, 2012). The weight update for a second order method such as Newton's method is calculated as follows (Haykin *et al.*, 2009):

$$\Delta w = \eta_g \left(\frac{\partial^2 E}{\partial w^2} \right)^{-1} \frac{\partial E}{\partial w} = \eta_g H(w)^{-1} \frac{\partial E}{\partial w}, \tag{4.1}$$

where $\eta_g \in (0,1)$ is the global learning rate and $(\partial^2 E / \partial w^2)^{-1}$ is the inverse of Hessian matrix, which is used to compute the individual or local learning rates. The computation of the inverse of the Hessian matrix is impractical for large neural networks since it requires $O(N^3)$ operations for each update to N network parameters (LeCun *et al.*, 2012).

Therefore, the stochastic diagonal Levenberg-Marquardt method (S-DLM) is proposed to update the network parameters. It approximates the Hessian matrix only by its diagonal

terms and drops the off-diagonal elements. The S-DLM requires $O(N)$ operations for each update to the N network parameters, making it suitable for large networks. The individual learning rate for a given weight, w_{ij} , is determined as follows (Becker and Le Cun, 1988):

$$\eta_{ij} = \frac{\eta_{\mathrm{g}}}{\left(\frac{\partial^2 E}{\partial w_{ij}^{\ 2}}\right) + \mu} \tag{4.2}$$

The adjustment $\Delta w_{ij}(n)$ to the weight w_{ij} is calculated using Eq. (4.3).

$$\Delta w_{ij} = \frac{\eta_{\mathrm{g}}}{\left(\frac{\partial^2 E}{\partial w_{ij}^2}\right) + \mu}\left(\frac{\partial E^p}{\partial w_{ij}}\right), \tag{4.3}$$

here, $h_{kk} = (\partial^2 E / \partial w_{ij}^2)$ estimates the second order derivative of the k-th diagonal element of the Hessian matrix with respect to the weight w_{ij} . The parameter μ is used to restrict the step size from becoming too large when the second order derivative becomes too small. The instantaneous gradient of the synaptic weight $(\partial E^p / \partial w_{ij})$ can be efficiently calculated through backpropagation using the stochastic gradient descent method, which is explained in detail in (Bouvrie, 2006). The running estimate of the second order derivative $\partial^2 E / \partial w_{ij}^2$ over the training samples is computed as follows:

$$\left(\frac{\partial^2 E}{\partial w_{ij}^2}\right)_{new} = (1-\gamma)\left(\frac{\partial^2 E}{\partial w_{ij}^2}\right)_{old} + \gamma\left(\frac{\partial^2 E^p}{\partial w_{ij}^2}\right), \tag{4.4}$$

where γ is a constant that determines the amount of memory being used. The computational cost of the second order derivatives given in Eq. (4.4) is almost the same as that of the gradient in a standard back-propagation pass, except that the weighted sums use the square of the weights (Becker and Le Cun, 1988; LeCun, 1989). The instantaneous second order derivatives of the diagonal elements relative to the weights in the CNNs can be calculated by back-propagating the diagonal Hessian (Becker and Le Cun, 1988; LeCun,

1989; LeCun *et al.*, 2012).

4.1.1.1 Fully Connected Layers

The S-DLM algorithm requires that the error be calculated for every input pattern using the squared-error loss function. For the p^{th} pattern, in a multiclass problem with c classes, the error is given by Eq. (4.5).

$$E^p = \frac{1}{2}\sum_{k=1}^{c}\left(d_k^p - o_k^p\right)^2, \tag{4.5}$$

where d_k^p is the k^{th} component of the target which corresponds to the p^{th} pattern and o_k^p is the value of the k^{th} output for the p^{th} input pattern. Let ℓ and L denote the current and output layers, respectively. Layer ℓ calculates its output by applying an activation function f to the sum of the dot product of the weight and input vectors, and the bias vector, as given in Eq. (4.6), where W and b are the weight and bias vectors, respectively, and x is the input vector.

$$x^{\ell} = f(y^{\ell}), \qquad with \quad y^{\ell} = W^{\ell}x^{\ell-1} + b^{\ell} \tag{4.6}$$

Using the rules for the back-propagation of the diagonal Hessian in neural networks (LeCun *et al.*, 2012), the square of the local gradient is calculated as follows:

$$(\delta^{\ell})^2 = \frac{\partial^2 E^p}{\partial (y^{\ell})^2} = ((W^{\ell+1})^T)^2(\delta^{\ell+1})^2 \circ (f'(y^{\ell}))^2, \tag{4.7}$$

where "$\circ$" denotes element-wise multiplication. For the output layer neurons, the square of the local gradient will take a slightly different form as follows:

$$((\delta^L))^2 = (f'(y^L))^2 \tag{4.8}$$

Finally, the delta rule is used to calculate the squares of the weight and bias gradients for a given neuron in the fully connected layer of the CNNs using Eq. (4.9) and Eq. (4.10), respectively.

$$\frac{\partial^2 E^p}{\partial (W^\ell)^2} = (x^{\ell-1})^2((\delta^\ell)^T)^2 \tag{4.9}$$

$$\frac{\partial^2 E^p}{\partial (b^\ell)^2} = \frac{\partial^2 E^p}{\partial (y^\ell)^2}\left(\frac{\partial y^\ell}{\partial b^\ell}\right)^2 = (\delta^\ell)^2 \tag{4.10}$$

4.1.1.2 Convolution Layers

Similarly, the square of the local gradient can be calculated for the convolution layers. Let j denote the current feature map. Then, for the convolution layer ℓ, the square of the local gradient is given by Eq. (4.11) (Bouvrie, 2006).

$$(\delta_j^\ell)^2 = (\beta_j^{\ell+1})^2((f'(y_j^\ell))^2 \circ up(\delta_j^{\ell+1})^2), \tag{4.11}$$

Here, $up(.)$ represents the up-sampling operation, which copies each input *n* times in the vertical and horizontal directions of the output. The up-sampling factor *n* must be equal to the subsampling factor of the subsampling layer. Using the square of the local gradient for each feature map, the square of the bias gradient can be computed by summing the local gradient over all the entries, as given by Eq. (4.12). The square of the kernel gradients can be determined using the recurrence relation in Eq. (4.13).

$$\frac{\partial^2 E^p}{\partial (b_j^\ell)^2} = \sum_{u,v} (\delta_j^\ell)^2{}_{uv} \quad , \tag{4.12}$$

$$\frac{\partial^2 E^p}{\partial (k_{ij}^\ell)^2} = \sum_{u,v} (\delta_j^\ell)^2{}_{uv} (p_i^{\ell-1})^2{}_{uv} \quad , \tag{4.13}$$

Here, $(p_i^{\ell-1})_{uv}$ is the patch in $x_i^{\ell-1}$, which is used to compute the element at position (u,v) in the output feature map x_j^ℓ through convolution with the kernel k_{ij}^ℓ.

4.1.1.3 Sub-sampling Layers

Similarly, after finding the square of the local gradient $(\delta_j^\ell)^2$ for neurons in the sub-sampling layer, the gradient of the bias *b* and weight β can be obtained using Eq. (4.14) and Eq. (4.15), respectively.

$$\frac{\partial^2 E^p}{\partial (b_j^\ell)^2} = \sum_{u,v} (\delta_j^\ell)^2{}_{uv} \tag{4.14}$$

$$\frac{\partial E^p}{\partial (\beta_j^\ell)^2} = \sum_{u,v} ((\delta_j^\ell)^2 \circ down(x_j^{\ell-1})^2{}_{uv}) \tag{4.15}$$

In Eq. (4.15), *down*(.) represents the sub-sampling operation. Since the second order properties of the error function change rather slowly and are mostly determined by the structure of the network and not the statistical nature of the training data (LeCun *et al.*, 1998), in practice, the second order derivative in Eq. (4.4) needs to run only on a small random subset of the training data before each pass of the learning algorithm (LeCun *et al.*, 1998).

4.3 Experimental Results and Discussion

4.3.1 Configuration of the Fault Signatures' Pool

The performance of the proposed method is confirmed using two experiments. The first experiment verifies the effectiveness of the stochastic diagonal Levenberg-Marquardt (S-DLM) algorithm by comparing it with the stochastic gradient descent (S-GD) algorithm. The second experiment compares the performance of the proposed method with traditional AE-based methods for bearing fault diagnosis under variable operating speeds.

TABLE 4-1. Description of the datasets used to evaluate the proposed method in experiment 1

fs=250 kHz		Operating speed (RPM)[1]	Crack size		
			Length (mm)	Width (mm)	Depth (mm)
Dataset 1	Training set	300	3	0.35	0.30
	Testing set	300		0.35	0.30
Dataset 2	Training set	400	3	0.35	0.30
	Testing set	400		0.35	0.30
Dataset 3	Training set	500	3	0.35	0.30
	Testing set	500		0.35	0.30

[1]*For each bearing condition, i.e., a defect-free bearing and 7 defective bearings, 90 five-second AE signals were recorded at each RPM.*

The first experiment uses three datasets. Each dataset contains AE signals for a defect-free bearing and bearings seeded with seven types of localized defects, as shown in Fig. 4-2. The dimensions of the defects and the operating speeds at which the AE signals are

recorded are given in Table 4-1. For each operating speed, 90 AE signals with a duration of five seconds each are recorded for each fault type. Hence, each dataset contains a total of 720 AE signals. Each dataset is divided into training and testing subsets containing half of the total, i.e., 360 AE signals each.

TABLE 4-2. Description of the datasets used to evaluate the proposed method in Experiment 2

fs=250 kHz		Operating speed (RPM)[1]	Crack size		
			Length (mm)	Width (mm)	Depth (mm)
Dataset 4	Training set	300, 400, 500	3	0.35	0.30
	Testing set	250, 350, 450		0.35	0.30
Dataset 5	Training set	300, 400, 500	12	0.49	0.50
	Testing set	250, 350, 450		0.49	0.50

[1]*For each bearing condition, i.e., a defect-free bearing and 7 defective bearings, 90 five-second AE data samples were obtained at each RPM*

Two datasets are used in the second experiment to verify the diagnostic performance of the proposed method under variable operating speeds. Both datasets contain AE signals for bearings seeded with the seven types of faults shown in Fig. 4-2 and a defect free bearing, i.e., a total of eight different bearing conditions. The datasets consider crack sizes of 3 mm and 12 mm. Hence, each dataset has signals for eight bearing conditions, which are recorded at six operating speeds of 250, 300, 350, 400, 450, and 500 revolutions per minute (RPM). For each dataset, the AE data is divided into training and testing subsets, as shown in Table 4-2. The training subset includes AE signals acquired at speeds of 300, 400, and 500 RPM, while the testing subset includes the AE signals recorded at speeds of 250, 350, and 450 RPM. For each operating speed, 90 AE signals for each bearing condition are recorded for a duration of five seconds each. Hence, each dataset contains $N_{RPM} \times N_{Classes} \times N_{Signals}$ or 4,320 AE signals, where N_{RPM} is the number of operating speeds for which the AE signals are recorded ($N_{RPM} = 6$), $N_{Classes}$ is the total number of defect types or bearing conditions ($N_{Classes} = 8$), and $N_{Signals}$ is the total number of AE signals recorded for each bearing condition at each shaft speed ($N_{Signals} = 90$).

The CNNs were designed to work on image data, which is intrinsically 2-D. On the other hand, we diagnose bearing defects using AE signals, which are 1-D in nature and are captured at very high sampling rates. Even if we stack the 1-D AE signals to create a 2-D

input for the CNNs, mining the raw AE signal for distinctive features would require very large CNNs. Hence, as discussed earlier the raw AE signals are converted into an equivalent 2-D representation, which shows the distribution of energies in various frequency bands of the AE signal, i.e., the spectral energy map or SEM, which is then used as the input to the CNNs. The SEM is generated by first multiplying the AE signal by a Hanning window function and then computing its FFT, which is shown in Fig. 4-2.

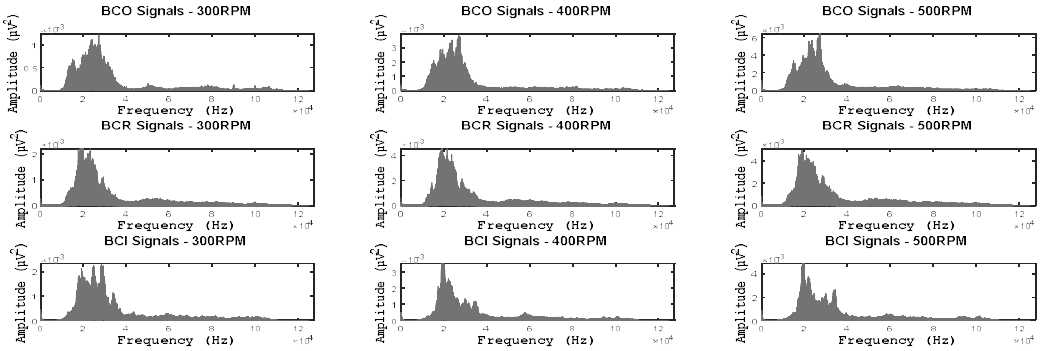


FIGURE 4-2. The Single-band power spectra of fault signals at different operating speeds.

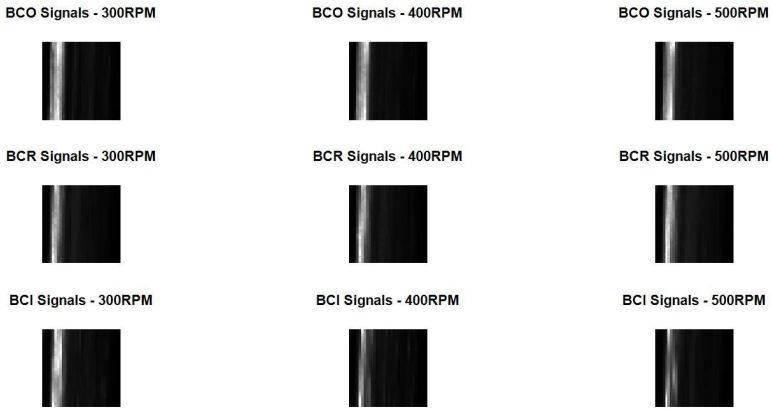


FIGURE 4-3. The spectral energy maps or SEMs for three different fault types at three different operating speeds. These SEMs are used as input by the CNNs.

The spectrum of the AE signal is then split into an appropriate number of frequency bands and the root mean square (RMS) value of each of these bands is calculated. The RMS value gives an approximation of the energy carried by each band (Wang *et al.*, 2015). These RMS values are arranged in the form of a 2-D array, which shows the distribution of energies across the entire spectrum of the AE signal. This 2-D array or SEM of the AE signal, examples of which are shown in Fig. 4-3, is then used as the input for the CNNs.

4.3.2 Efficacy of the Stochastic Diagonal Levenberg-Marquardt Algorithm

As discussed in Section 4.3.2, the second order derivative in Eq. (4.4) can be estimated without scanning the entire training set. Rather, it can be approximated using only a small subset or mini-batch of the training data before each iteration of the learning algorithm. Therefore, the second order derivative in Eq. (4.4) can be estimated as follows:

$$\frac{\partial^2 E}{\partial w_{ij}^2} = \frac{1}{P}\sum_{p=1}^{P}\frac{\partial^2 E^p}{\partial w_{ij}^2}, \tag{4.16}$$

where P is the number of samples used to estimate the second order diagonal derivative. The approximation of the second order derivative in Eq. (4.16) is not affected by the choice of the mini-batch (LeCun *et al.*, 1998). Due to this property, such a training algorithm is labeled as the Mini-batch diagonal Levenberg-Marquardt algorithm (M-DLM).

This study compares the performances of three training algorithms for CNNs under the same conditions: the stochastic diagonal Levenberg-Marquardt algorithm (S-DLM), the stochastic gradient descent algorithm (S-GD), and the mini-batch diagonal Levenberg-Marquardt (M-DLM) algorithm. The performance of the algorithms is determined in terms of the average training time of a single epoch, convergence, and classification accuracy for the datasets listed in Table 4-1.

4.1.1.4 Average Training Time of a Single Training Epoch

In comparison to S-GD, the S-DLM and M-DLM algorithms are not expected to decrease the training time as they must estimate the Hessian, which is time consuming. The results in Fig. 4-4 affirm this notion as the S-GD requires the least amount of time to train the CNNs. The S-DLM (equivalent to M-DLM with a mini-batch size of 1) is the slowest training algorithm as it requires scanning through the entire training set before each learning iteration. The M-DLM requires more time compared to S-GD, but less than S-DLM since it uses only a small number of training samples and also omits the memory constant, further simplifying its computation. The training time of the M-DLM algorithm can be further reduced by either decreasing the mini-batch size or the frequency of re-

estimating the second order derivative.

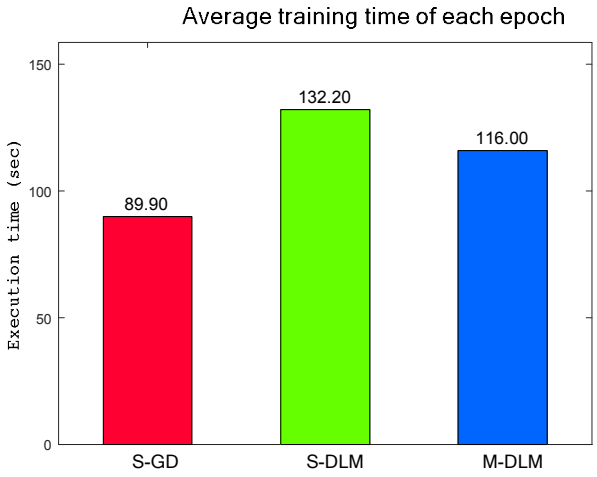


FIGURE 4-4. The average execution time of a single training epoch for various learning algorithms.

In this test, the M-DLM algorithm uses a mini-batch of 50 training samples and re-estimates the second order derivative after each epoch. The results are generated using Matlab on a general purpose computing platform with an overclocked 3.3 GHz Intel Core i5-2500 CPU.

4.1.1.5 Convergence of the Learning Algorithms

The learning performance of these algorithms is measured primarily by examining the convergence speed of these algorithms. Fig. 4-5 shows both the mean square error (MSE) plots for the training samples and the misclassification rate (MCR) plots for the testing samples of the datasets given in Table 4-1. The learning curves, i.e., MSE plots, indicate that the S-DLM and M-DLM algorithms train the CNNs better than the S-GD, as these algorithms utilize the second order curvature information of the error function. The S-DLM and M-DLM algorithms enable the CNNs to reach the global minimum in just 2 epochs. In contrast, the S-GD yields relatively poor convergence performance. The learning performance of these algorithms is also examined through their MCR on the testing subset. The MCR plots in Fig. 4-5 clearly indicate that the S-DLM and M-DLM algorithms enable the CNNs to learn the discriminative model for the training data better than the S-GD and thus, yield better MCR values for the testing subset, which are close to zero after only two training epochs.

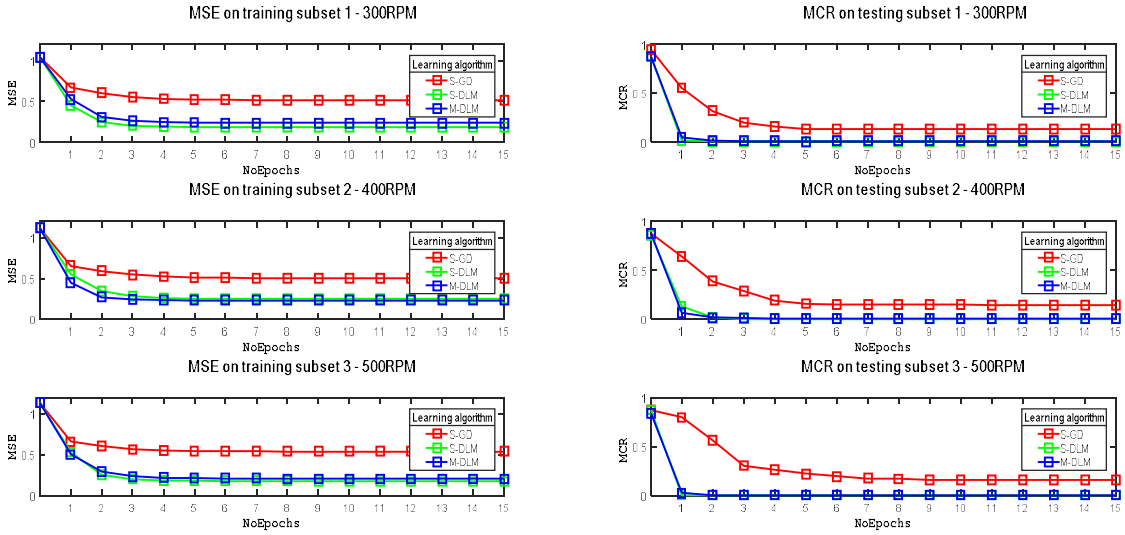


FIGURE 4-5. Training mean square error and testing misclassification error for dataset 1 using S-GD, S-DLM, M-DLM learning algorithms.

4.1.1.6 Classification Accuracy of Different CNNs

The effectiveness of the proposed training scheme is also validated by comparing the average classification accuracy (ACA), as defined in Eq. (4.17), of the different CNNs that are trained using different training algorithms on the datasets shown in Table 4-1. The ACA values for the CNNs trained using the different training algorithms are given in Table 4-3.

$$ACA = \frac{\sum_{N_{classes}} N_{TP}}{N_{testdata}} \times 100(\%) , \tag{4.17}$$

Here, $N_{testdata}$ is the total number of samples used to test the classification accuracy of the proposed method, $N_{classes}$ is the total number of defect types, and N_{TP} is the number of data points in class i, which are correctly classified as class i. In addition, the sensitivity, which is a useful metric for evaluating the diagnostic performance of the proposed method in relation to each bearing condition, is defined as follows:

$$Sensitivity = \frac{N_{TP}}{N_{TP} + N_{FN}} \times 100(\%) , \tag{4.18}$$

where N_{FN} is the number of data points in class i, which are classified incorrectly. The results shown in Table 4-3 demonstrate that the S-DLM and M-DLM algorithms are more powerful than the traditional S-GD algorithm for training CNNs, and can yield better diagnostic performance.

TABLE 4-3. The average classification accuracy and sensitivity for each bearing defect type under variable speed conditions using different training algorithms for the CNNs

Datasets	Learning algorithm	Average sensitivity of each fault type								ACA (%)
		BCI	BCO	BCR	BCIO	BCIR	BCOR	BCIOR	BNC	
Dataset 1	S-GD	100	91.11	100	51.11	60	91.11	100	100	86.66
	S-DLM	100	95.55	100	100	100	100	100	100	99.44
	M-DLM	100	93.33	100	100	100	100	100	97.77	98.88
Dataset 2	S-GD	95.55	100	100	100	86.66	4.44	97.77	100	85.55
	S-DLM	100	100	100	100	100	97.77	100	100	99.72
	M-DLM	97.77	100	100	100	100	97.77	100	100	99.44
Dataset 3	S-GD	100	6.66	100	100	100	75.55	100	91.11	84.16
	S-DLM	100	100	100	100	100	100	100	100	100
	M-DLM	97.77	100	100	100	100	100	100	100	99.72

4.3.3 Performance Evaluation of the Proposed Method for Bearing Fault Diagnosis under Variable Operating Speeds

The key difference between CNNs-based techniques and the traditional analysis-based methods is the manner of extracting features. The diagnostic performance of conventional methods depends on the extraction and selection of the best feature set, whereas CNNs can automatically learn the optimal features from the original signals. Traditional methods have mostly been tested under constant operating speeds because variation of the speed results in unstable and inconsistent features, which significantly degrade their diagnostic performance. The proposed method, however, uses the SEMs and CNNs to diagnose bearing defects under variable operating speeds and therefore, can effectively diagnose the eight types of bearing conditions considered in this study. The proposed method is compared to AE-based methods used for bearing fault diagnosis (Kang *et al.*, 2016; Tra *et al.*, 2017). In these methods, various statistical measures of the time and frequency domain AE signals are used as features (Kang *et al.*, 2016; Tra *et al.*, 2017), along with statistical quantities calculated through complex envelope analysis (Kang *et al.*, 2016). Moreover, before feature extraction, the authors in (Tra *et al.*, 2017) also consider the subband analysis of the AE signals using the discrete wavelet packet transform (DWPT). The extracted features are then evaluated and selected using feature selection algorithms such as outlier-insensitive hybrid feature selection (OIHFS) (Kang *et al.*, 2016). Finally, the selected features are used to detect bearing defects using classifiers such as K-NN (Kang *et al.*, 2016) and SVM (Tra *et al.*, 2017).

Table 4-4 presents the diagnostic performance of the proposed method and the techniques previously proposed in (Kang *et al.*, 2016; Tra *et al.*, 2017) for the same datasets as described in Table 4-2. The diagnostic performance of all the methods is measured in terms of the ACA and sensitivity. It is clear from the results given in Table 4-4 that the proposed method delivers better diagnostic performance than the previous methods (Kang *et al.*, 2016; Tra *et al.*, 2017). The average classification accuracy of the proposed method is 98.47%, which shows significant improvement when compared to the methods proposed previously (Kang *et al.*, 2016; Tra *et al.*, 2017), i.e., by 47% and 23%, respectively.

TABLE 4-4. The average classification accuracy (ACA) and sensitivity for each bearing defect type under variable speed conditions using the proposed method and two state-of-the-art methods.

Datasets	Methodologies	Average sensitivity for each fault type								ACA (%)
		BCI	BCO	BCR	BCIO	BCIR	BCOR	BCIOR	BNC	
Dataset 4	(Kang *et al.*, 2016)	19.62	47.40	75.18	17.03	59.62	30.74	10	3.33	32.87
	(Tra *et al.*, 2017)	11.11	13.33	100	100	97.77	97.77	0	0	52.5
	Proposed	66.66	100	100	100	89.25	99.25	99.25	99.62	94.25
Dataset 5	(Kang *et al.*, 2016)	7.03	70	66.66	79.62	5.92	44.81	74.07	62.96	51.38
	(Tra *et al.*, 2017)	100	100	97.77	97.77	100	100	0	0	74.44
	Proposed	100	100	91.85	98.14	99.25	99.25	100	99.25	98.47

4.4 Conclusions

The diagnostic performance of traditional methods for bearing fault diagnosis relies on the quality of the features used to construct discriminative fault models, which can be a huge limitation especially when it is difficult to design such features or the features become inconsistent with changes in the operating conditions. Under variable operating speeds, feature analysis-based methods yield poor diagnostic performance, as changes of speed result in inconsistent features. The proposed method used CNNs to automatically learn the optimal features, which would yield the best diagnostic performance. Moreover, the CNNs used SEMs of the AE signals as inputs, which are two-dimensional maps of the distribution of energy across different bands of the AE spectrum. It is hypothesized that the variation of a bearing's speed would not alter the overall shape of the AE spectrum, rather it may only scale and translate it. Thus, at different speeds, the same defect would yield SEMs

that are scaled and shifted versions of each other. Since, the classification performance of CNNs is not affected by distortion in their inputs, it makes them ideal for using SEMs as inputs to diagnose different bearing defects. The experimental results confirmed our hypothesis, as the CNNs trained using the S-DLM algorithm yielded significantly better diagnostic performance under variable operating speeds compared to existing methods. The proposed method can diagnose both single and compound bearing defects with 47.09% and 23.03% more accuracy compared to the two state-of-the-art methods. The S-DLM algorithm was used to train the CNNs after evaluating three different training algorithms in terms of speed and accuracy.

4.5 Acknowledgments

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. 20162220100050, No. 20161120100350, and No. 20172510102130). It was also funded in part by The Leading Human Resource Training Program of the Regional Neo industry through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and future Planning (NRF-2016H1D5A1910564) and in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2016R1D1A3B03931927).

5 Conclusion and Future Work

5.1 Summary of the research

The accomplished research work focuses on performing fault diagnosis of rolling element bearings. Which is an important component of induction motors. Defects in bearings causes failures in rotary machines most often. Thats why condition monitoring and fault diagnosis of rolling element bearings are high importance. Incipient bearing defects are primarily diagnosed by searching for peaks at bearing defect frequencies (BDFs) in the envelope power spectrum of acoustic emission (AE) signals, or by developing discriminative models for features extracted from the AE signals through time and frequency domain analysis. These methods perform poorly under variable speed conditions, as variation in a bearing's speed can drastically change BDFs, and result in bad discriminative models for the extracted features. Hence, in this thesis, we propose new methods to solve the problem of fault diagnosis under variable speed condition as follows. Chapter 2 in the thesis reports a proposed method that decomposes the raw AE signal into different subband signals using the discrete wavelet packet transform (DWPT), which is more effective for the multiresolution analysis of non-stationary signals. A feature vector is extracted from each subband signal, resulting in multiple feature vectors from the raw AE signal. The feature vectors extracted from different subbands of the original AE signal result in models of the fault data that are discriminatory, even under variations in the rotational speed. The weighted committee machine (WCM), which is an ensemble of support vector machines (SVMs) and artificial neural networks (ANNs), is employed to build discriminatory models of the feature vectors for different bearing defects. The proposed method also improves the generalization performance of the neural networks to enhance their classification accuracy, particularly with limited training data. In the chapter 3, we proposed a novel method for diagnosing bearing defects under variable speed conditions. The proposed method does not require any specialized hardware, rather it employs convolutional neural networks to mine the energy distribution maps for discriminative features, which can be used to diagnose bearing defects under variable operating speeds. This work proposes the use of energy distribution maps of the AE signal

spectrum as inputs to the CNNs for the diagnosis of bearing defects under variable operating speeds. The use of energy distribution maps as inputs to the CNNs is justified by a detailed analysis of the effects of speed on the AE signals, and a comparison of different types of inputs used to train the CNNs. Experimental results demonstrate that the proposed method yields better diagnostic performance in comparison to state-of-the-art AE-based methods. Besides, we also investigated various training algorithms for CNNs and proposed the use of S-DLM algorithm for training the CNNs. The S-DLM uses the diagonal terms of the Hessian matrix to compute the individual learning rates of all parameters of the CNNs before scanning through the training set. Our empirical results indicate that CNNs converge quickly when trained using the S-DLM algorithm, thereby enhancing the performance of the proposed fault diagnosis system. Details of the method are presented in Chapter 4.

5.2 Future Directions

Although fault diagnosis performance of the proposed algorithms is good to compare to previous work. The algorithms solved some of the significant issues in this domain. But still, there is room for improvement in it. There are many aspects where the diagnosis methods can be improved, i.e., using noise isolation techniques to separate useful signals in the noised signal acquired from sensors to enhance the quality of the analyzed signal. Hence, the extracted characteristics of the signal will be better, thus improving the system's diagnostic capability. Also, we'd like to focus on feature-selective techniques to select the best features for fault diagnosis in addition to the currently used methods such as genetic algorithm, hybrid genetic algorithm, HillClimbing... The classifier is one of the most crucial parts of the fault diagnosis system. In this thesis, we introduce two classification techniques based on machine learning as weighted committee machine and CNNs. However, there are still many other techniques that we can also research and exploit, so in future, we will continue research and develop different classification techniques to improve the system's diagnostic capabilities.

Another critical point here is that the techniques I studied and recommended not only apply to diagnosis faults of the bearing but also can be used to diagnose defects in much other industrial equipment. Next time, we will start to research and apply these fault dectect algorithms in essential and sensitive industrial equipment such as vessel pressure, boiler

tube...If the application is successful, it will bring tremendous economic benefits as well as regard production safety.

Bibliography

Abu-Mahfouz, I. (**2005**). "A comparative study of three artificial neural networks for the detection and classification of gear faults," International Journal of General Systems **34**, 261-277.

Avidan, S. (**2004**). "Support vector tracking," IEEE transactions on pattern analysis and machine intelligence **26**, 1064-1072.

Becker, S., and Le Cun, Y. (**1988**). "Improving the convergence of back-propagation learning with second order methods," in *Proceedings of the 1988 connectionist models summer school*, pp. 29-37.

Bediaga, I., Mendizabal, X., Arnaiz, A., and Munoa, J. (**2013**). "Ball bearing damage detection using traditional signal processing algorithms," IEEE Instrumentation & Measurement Magazine **16**, 20-25.

Berriri, H., Naouar, M. W., and Slama-Belkhodja, I. (**2012**). "Easy and fast sensor fault detection and isolation algorithm for electrical drives," IEEE transactions on power electronics **27**, 490-499.

Bouvrie, J. (**2006**). "Notes on convolutional neural networks."

Cabal-Yepez, E., Garcia-Ramirez, A. G., Romero-Troncoso, R. J., Garcia-Perez, A., and Osornio-Rios, R. A. (**2013**). "Reconfigurable monitoring system for time-frequency analysis on industrial equipment through STFT and DWT," IEEE Transactions on Industrial Informatics **9**, 760-771.

Caesarendra, W., Kosasih, P. B., Tieu, A. K., Moodie, C. A. S., and Choi, B.-K. (**2013**). "Condition monitoring of naturally damaged slow speed slewing bearing based on ensemble empirical mode decomposition," Journal of mechanical science and technology **27**, 2253-2262.

Elbouchikhi, E., Choqueuse, V., Auger, F., and Benbouzid, M. E. H. (**2017**). "Motor Current Signal Analysis Based on a Matched Subspace Detector," IEEE Transactions on Instrumentation and Measurement.

Frosini, L., and Bassi, E. (**2010**). "Stator current and motor efficiency as indicators for different types of bearing faults in induction motors," IEEE Transactions on Industrial electronics **57**, 244-251.

Fyfe, K., and Munck, E. (**1997**). "Analysis of computed order tracking," Mechanical Systems and Signal Processing **11**, 187-205.

Hagiwara, K. (**2002**). "Regularization learning, early stopping and biased estimator," Neurocomputing **48**, 937-955.

Haykin, S. S., Haykin, S. S., Haykin, S. S., and Haykin, S. S. (**2009**). *Neural networks and learning machines* (Pearson Upper Saddle River, NJ, USA:).

Immovilli, F., Bianchini, C., Cocconcelli, M., Bellini, A., and Rubini, R. (**2013**). "Bearing fault model for induction motor with externally induced vibration," IEEE Transactions on Industrial Electronics **60**, 3408-3418.

ISO/TC 108, S. (**2007**). "ISO 22096:2007(E)," Condition monitoring and diagnosis of machines–Acoustic Emission.

Jin, X., Zhao, M., Chow, T. W., and Pecht, M. (**2014**). "Motor bearing fault diagnosis using

trace ratio linear discriminant analysis," IEEE Transactions on Industrial Electronics **61**, 2441-2451.

Kang, M., Islam, M. R., Kim, J., Kim, J.-M., and Pecht, M. (**2016**). "A hybrid feature selection scheme for reducing diagnostic performance deterioration caused by outliers in data-driven diagnostics," IEEE Transactions on Industrial Electronics **63**, 3299-3310.

Kang, M., Kim, J., Choi, B.-K., and Kim, J.-M. (**2015a**). "Envelope analysis with a genetic algorithm-based adaptive filter bank for bearing fault detection," The Journal of the Acoustical Society of America **138**, EL65-EL70.

Kang, M., Kim, J., and Kim, J.-M. (**2015b**). "High-performance and energy-efficient fault diagnosis using effective envelope analysis and denoising on a general-purpose graphics processing unit," IEEE Transactions on Power Electronics **30**, 2763-2776.

Kang, M., Kim, J., Kim, J.-M., Tan, A. C., Kim, E. Y., and Choi, B.-K. (**2015c**). "Reliable fault diagnosis for low-speed bearings using individually trained support vector machines with kernel discriminative feature analysis," IEEE Transactions on Power Electronics **30**, 2786-2797.

Kang, M., Kim, J., Wills, L. M., and Kim, J.-M. (**2015d**). "Time-varying and multiresolution envelope analysis and discriminative feature analysis for bearing fault diagnosis," IEEE Transactions on Industrial Electronics **62**, 7749-7761.

Khan, S. A., and Kim, J.-M. (**2016**). "Rotational speed invariant fault diagnosis in bearings using vibration signal imaging and local binary patterns," The Journal of the Acoustical Society of America **139**, EL100-EL104.

Lacey, S. (**2008**). "An overview of bearing vibration analysis," maintenance & asset management **23**, 32-42.

Lau, E. C., and Ngan, H. (**2010**). "Detection of motor bearing outer raceway defect by wavelet packet transformed motor current signature analysis," IEEE Transactions on Instrumentation and Measurement **59**, 2683-2690.

LeCun, Y. (**1989**). "Generalization and network design strategies," Connectionism in perspective, 143-155.

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

LeCun, Y., Cortes, C., and Burges, C. J. (**2010**). "Mnist handwritten digit database. AT&T Labs."

LeCun, Y. A., Bottou, L., Orr, G. B., and Müller, K.-R. (**2012**). "Efficient backprop," in *Neural networks: Tricks of the trade* (Springer), pp. 9-48.

Liu, B., and Pan, H. (**2010**). "Fault diagnosis of bearing using wavelet packet transform and PSO-DV based neural network," in *Natural Computation (ICNC), 2010 Sixth International Conference on* (IEEE), pp. 1238-1242.

Liu, H. (**2010**). "On the levenberg-marquardt training method for feed-forward neural networks," in *Natural Computation (ICNC), 2010 Sixth International Conference on* (IEEE), pp. 456-460.

Lu, S., Guo, J., He, Q., Liu, F., Liu, Y., and Zhao, J. (**2016**). "A novel contactless angular resampling method for motor bearing fault diagnosis under variable speed," IEEE Transactions on Instrumentation and Measurement **65**, 2538-2550.

Lu, S., He, Q., Hu, F., and Kong, F. (**2014**). "Sequential multiscale noise tuning stochastic resonance for train bearing fault diagnosis in an embedded system," IEEE

Transactions on Instrumentation and Measurement **63**, 106-116.

Ma, H., Xiong, Y., Fang, H., and Gu, L. (**2015**). "Fault diagnosis of bearing based on fuzzy support vector machine," in *Prognostics and System Health Management Conference (PHM), 2015* (IEEE), pp. 1-4.

Nguyen, P., Kang, M., Kim, J.-M., Ahn, B.-H., Ha, J.-M., and Choi, B.-K. (**2015**). "Robust condition monitoring of rolling element bearings using de-noising and envelope analysis with signal decomposition techniques," Expert Systems with Applications **42**, 9024-9032.

Niknam, S. A., Songmene, V., and Au, Y. J. (**2013**). "The use of acoustic emission information to distinguish between dry and lubricated rolling element bearings in low-speed rotating machines," The International Journal of Advanced Manufacturing Technology **69**, 2679-2689.

Osborne, M. R. (**1992**). "Fisher's method of scoring," International Statistical Review/Revue Internationale de Statistique, 99-117.

Randall, R. B., and Antoni, J. (**2011**). "Rolling element bearing diagnostics—a tutorial," Mechanical systems and signal processing **25**, 485-520.

Rauber, T. W., de Assis Boldt, F., and Varejão, F. M. (**2015**). "Heterogeneous feature models and feature selection applied to bearing fault diagnosis," IEEE Transactions on Industrial Electronics **62**, 637-646.

Seshadrinath, J., Singh, B., and Panigrahi, B. K. (**2014**). "Investigation of vibration signatures for multiple fault diagnosis in variable frequency drives using complex wavelets," IEEE Transactions on Power Electronics **29**, 936-945.

Tandon, N., and Choudhury, A. (**1999**). "A review of vibration and acoustic measurement methods for the detection of defects in rolling element bearings," Tribology international **32**, 469-480.

Thorsen, O. V., and Dalva, M. (**1999**). "Failure identification and analysis for high-voltage induction motors in the petrochemical industry," IEEE Transactions on Industry Applications **35**, 810-818.

Tra, V., Kim, J., Khan, S. A., and Kim, J.-M. (**2017**). "Incipient fault diagnosis in bearings under variable speed conditions using multiresolution analysis and a weighted committee machine," The Journal of the Acoustical Society of America **142**, EL35-EL41.

Wang, J., Gao, R. X., and Yan, R. (**2014a**). "Multi-scale enveloping order spectrogram for rotating machine health diagnosis," Mechanical Systems and Signal Processing **46**, 28-44.

Wang, T., Liang, M., Li, J., and Cheng, W. (**2014b**). "Rolling element bearing fault diagnosis via fault characteristic order (FCO) analysis," Mechanical Systems and Signal Processing **45**, 139-153.

Wang, Y., Xu, G., Zhang, Q., Liu, D., and Jiang, K. (**2015**). "Rotating speed isolation and its application to rolling element bearing fault diagnosis under large speed variation conditions," Journal of Sound and Vibration **348**, 381-396.

Yang, Z., Peng, T., Li, J., Yang, H., and Jiang, H. (**2009**). "Fault Diagnosis of Rolling Bearing Based on Wavelet Packet Transform and Support Vector Machine," in *Measuring Technology and Mechatronics Automation, 2009. ICMTMA'09. International Conference on* (IEEE), pp. 650-653.