A Componential Coding Neural Network based Signal Modelling for Condition Monitoring

Khalid Rabeyee, Yuandong Xu, Aisha Alashter, Fengshou Gu, Andrew D. Ball
University of Huddersfield, HD1 3DH, UK
khalid.rabeyee@hud.ac.uk

Abstract. Many condition monitoring (CM) techniques have been investigated for early fault detection and diagnosis in order to avoid unexpected breakdowns due to machinery failures. However, manual techniques require well-skilled labours which will increase the cost of the monitoring process and may not always be available at the site. One of the most promising approaches is to automate the monitoring process using artificial intelligence (AI) techniques. However, the majority of AI-based techniques have been developed in CM for the post-processing stage, whereas the critical tasks including feature extraction and selection are still manually processed. This study focuses on the extending AI techniques in all phases of CM process by using a **Componential Coding Neural Network** (CCNN) which has been found to have unique properties of being trained through unsupervised learning, capable of dealing with raw data sets, translation invariance and high computational efficiency. These advantages of CCNN make it particularly suitable for automated analysis of the vibration data arisen from typical machine components such as the rolling element bearings which exhibit periodic phenomena with high non-stationarity and strong noise contamination. The CCNN was evaluated using both simulated and experimental data collected from a healthy and two defective tapered roller bearings under different operating conditions. Both of the results showed the capability of CCNN in detecting the initial anomalies of roller element bearings.

Keywords: *artificial intelligence, componential coding network, artificial neural network, tapered roller bearing. Fault detection.*

1 Introduction

The information and communication technologies continue expanding the range of possible applications to the industrial world. Hence, the huge amount of data produced and collected could overflow most conventional monitoring systems, and this has opened a new range of possible challenges in fault diagnosis of rotating machinery. One of the most promising approaches is to automate the process of CM by applying automatic and accurate fault identification techniques[1]. Machinery system consists of several components, and rolling element bearings (REBs) are one of the most critical components commonly used in almost all forms of rotary machines[2]. The presence of faults in rolling element bearings can cause catastrophic failure of machinery systems due to overheating or over speed and thermally induced seizure. Incipient fault modes of rolling bearings are usually those of a scratched crack, improper lubrication and the inclusion of foreign material. In addition, several mechanisms of wear, such as abrasive, adhesive, fretting etc., can occur in REBs [3] which as a result may change the surface texture due to plastic deformation. The clearance will increase with the occurrence of wear. Howard in [4]stated that "Severe wear changes the raceway profile and alters the rolling element profile and diameter, increasing the bearing clearance". Furthermore, it was reported by Nguyen-Schäfer [5] that the internal clearance strongly influences load distribution as well as wear. According to the literature, the effect of internal clearance on REBs' life has received considerable research interest. Several studies have investigated the impact of wear and simulated wear on the condition monitoring of REBs with the assumption of clearances variation such as Goerke et al.[6], Rehab et al.[7], Rabeyee et al.[8] and Rabeyee et al.[9]. The studies claimed that the change in internal clearance

has influenced the diagnostic features. It can be concluded that increasing clearance due to unavoidable wear can affect the vibration signature and hence the diagnostic performances. This makes the CM of REBs more a challenging task particularly, in the case of tapered roller bearings (TRBs). Thus, in this study, the vibration responses of TRBs will be investigated with different levels of internal clearances to evaluate the performance of the proposed method using AI-based technique.

To avoid catastrophic failures, early fault detection of REBs is significantly important and many techniques such as vibration, temperature, and acoustic emission etc. can be used for condition monitoring and fault diagnostics. Vibration analysis has been widely accepted as an effective method and a popular strategy for monitoring rotating machinery due to its high sensitivity to defects [10-15]. A number of research efforts were made to develop accurate fault detection and diagnose technique which may occur to REBs. However, the vast majority are based on signal processing and conventional data analysis techniques as reviewed by Rai et al in.[11]. and by El-Thalji and Jantunen in [12]. Besides the conventional diagnosis techniques, several attempts have been carried out to automate the fault diagnosis of REBs based on AI. Most of the existing AI approaches for fault diagnosis have employed ANN and its variants (e.g., polynomial neural networks (PNN), dynamic wavelet neural networks (DWNN), self-organizing feature maps (SOM)[16], multilayer perceptron (MLP) neural network[17]. ANN has the advantages that it can readily process nonlinear, high-order, and time-varying dynamics. ANN in a data processing system that consists of three types of layers: the input layer, hidden layer, and output layer. Each layer has a number of simple, neuron-like processing elements called “nodes” or “neurons” that interact with each other by using numerically weighted connections[17]. Mainly, the application of ANN approach consists of three main steps namely, training, testing, and implementation. In the training stage of a model, feature extraction and selection are the most important and critical steps. In this paper, the application of unsupervised AI technique to the CM of TRB is investigated and the method is evaluated using both simulated and experimental datasets.

2 Componential Coding Neural Network

Componential coding neural network (CCNN) is an unsupervised neural network and was developed and introduced for multi-dimensional image data processing by [18-20]. It was developed based on the idea that it can adapt the data-model to the data used for training the model, so on average it is able to reconstruct each input vector optimally. One of the most important features of this network is the translation invariance. The network is trained using a gradient descent algorithm. It has the ability to detect relatively immutable component substructures, which arise again and again in the different data patterns of an ensemble. By learning a function$(\tilde{x}) \approx (x)$, CCNN tries to learn how to make the target values $(\tilde{x})$ equal to the input (x) by minimizing the reconstruction error using the equation as:

$$Er = \left(\left|x - \tilde{x}\right|^2\right) \tag{1}$$

In the typical neural networks, the neurone's output is calculated by a signal scalar multiply between the input (x) and weight vector (w) as $\xi = x.w_j$. Conversely, the neurone's output in CCNN is calculated (J) products for each w_j instead of one scalar product. This will form a periodic correlation function as shown in (2).

$$cr(x, w_j) = F^{-1}\left(F(x) \times F(w_j)^*\right) \tag{2}$$

Where x represents input signal, $F()$ is FFT, $F^{-1}()$ is the inverse IFFT, $()^*$ is the complex conjugate and the operator $\times$ for the point-wise multiplication to get a vector of coordinate products.
Then a nonlinear threshold is applied to produce output vector $y_j(x)$ each of N samples i.e. from $n = 1\ to\ N$.
Activation function implemented in the network is described in (3) where the r indicates a non-linear neuron function [20]:

$$r(\beta) = \frac{1}{2}\left\{ s\log\left[1+\exp\left(\frac{\beta-\vartheta}{s}\right)\right]\right\}^2 \tag{3}$$

Where β is the result of the circular correlation between the input patterns and the basis vector, which is denoted by (4).

$$\beta = cr(x, w_j) \tag{4}$$

The parameter ϑ represents the threshold and the S is softness parameter and hence, their initial values have to be set manually in the network design stage. The code $y_j(x)$ is derived by applying the neuron function to cr correlation function as in(5).

$$y_j(x) = r(cr(x, w_j)) \tag{5}$$

The yielded code $y_j(x)$ is the output code of the encoder for the input x [20]. The output code in (5) is used to perform the reconstruction process $\tilde{x}$ of the input (x), by firstly convolve each of $y_j(x)$ with the corresponding as in (6).

$$cv(w_j, y_j(x)) = F^{-1}\left(F(x)\times F(y_j(x))\right) \tag{6}$$

where cv is a circular convolution operator. Using the efficient FFT algorithm, the convolution function can be computed. Then, the weighted summation is used to combine the (J) resulting convolution functions[19, 21].

$$\tilde{x} \equiv \sum_{j=1}^{J} a_j\, cv(\mathbf{w}_j, y_j(x)) \tag{7}$$

where a_j represents the adaptive parameters weight scales; their values are determined by optimization procedure as (8).

$$a_j \equiv \sum_{j'}^{J} (M^{-1})_{jj'} \left\langle x \bullet cv(w_{j'}, y_{j'}(x))\right\rangle_{\{x\in W\}} \tag{8}$$

where the inverse M^{-1} is a square matrix of size $J \times J$ and its elements are calculated by:

$$m_{jj'} \equiv \left\langle cv\left(w_j, y_j(x)\right) \bullet cv\left(w_{j'}, y_{j'}(x)\right)\right\rangle_{\{\mathbf{x}\in W\}} \tag{9}$$

The scalar product includes summation over all the time samples and $cv(w_j, y_j(x))$ is N dimensional vectors in both of the bove equations.

3 Anomaly Detection

To measure the performance in detecting novelty in a new data set, a normalised error is defined as in(10).

$$ADI = \frac{R_m - R_u}{R_u} \quad (10)$$

Where R_m is the reconstruction error of the monitored data, R_u is the reconstruction error obtained during training from unseen data used for the validation stage. ADI is used to represent the ratio of the reconstruction error from the validation stage to the reconstruction error obtained from the monitoring stage. If the outcome of ADI > 0 by 0.1 or higher, this means the monitored data differ from the baseline by more than 10% and the monitored data is considered an anomaly. It can be used as a clear indication of physical change in the system. In addition, the higher amplitude of ADI the larger the fault would be. Therefore, the amplitude of the ADI is adopted as a measure of the fault severity.

4 Test Rig Facilities

A test rig with a simple structure has been developed to minimise the noise influences in the experimental study. It comprises of several components, such as a motor, a shaft, a coupling and bearings, which are described in Figure 1 (a). A tapered roller bearing type (TIMKEN 31308) is used as a test bearing and it was fitted in the SKF housing. Moreover, to measure the vibration data, two piezoelectric accelerometers (CA-YD-104T) were mounted vertically and horizontally on the housing. For precise clearance measurement, a slip metric gauge box set, type Matrix Pitter 8075C was used. The gauge box is used in several applications such as length measurement for the regulation and adjustment of indicating measuring instruments and linear dimensions of industrial components [22]. The fault frequencies of the bearing are theoretically calculated based on the following formulas shown in Table 1 and the geometrical information provided by the manufacturers.

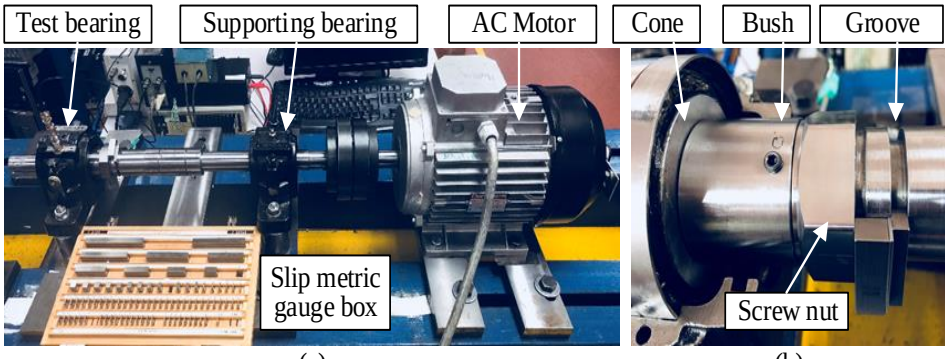

Figure 1. (a) Test rig, and (b) clearance measurement structure

Table 1 Characteristic Fault Frequencies

Fault type	Calculation formula
Ball Pass Frequency of Outer Race (Hz)	$f_{BPO} = \frac{1}{2} z \left(1 - \frac{d}{D} \cos \alpha \right) f_s$
Ball Pass Frequency of Inner Race (Hz)	$f_{BPI} = \frac{1}{2} z \left(1 + \frac{d}{D} \cos \alpha \right) f_s$
Ball Spin Frequency (Hz)	$f_{BS} = \frac{D}{2d} \left[1 - \left(\frac{d}{D} \cos \alpha \right)^2 \right] f_s$
Fundamental Train Frequency (Hz)	$f_c = \frac{1}{2} \left(1 - \frac{d}{D} \cos \alpha \right) f_s$

where Z is the number of rolling elements, d is the rolling element diameter, D is the pitch diameter, α is the contact angle, and f_s is the shaft frequency. A localised defect of a rolling element impacts the outer race and the inner race once a spin and two transients are generated. Thus, $2f_{bs}$ is normally used as the fault frequency to indicate rolling element defects. The theoretical fault frequencies of bearings used in this study are calculated for the outer race, rolling elements, carriage and inner race. Thereafter, at a running speed of 1500 rpm ($f_s = 25Hz$) and a clearance of zero, all characteristic frequencies can be obtained as shown in Table 2

Table 2 Defect frequencies of bearings tested at 1500 rpm

Fault location	Defect frequency (Hz)
Inner race	217.5595
Outer race	157.1041
Roller	133.5958
Cage	10.4736

5 Results and Discussion

5.1 Implementation of the CCNN to Simulated Data

This section examines the detection capability of CCNN using simulation data and presents and discuss the results obtained from applying CCNN to real data collected from the test bearings, so simulating signals were generated and different levels of noise have been added to the signals. The anomaly detection was carried out using CCNN and the results have been discussed and demonstrated.

5.1.1 Simulated Signal

Many attempts have been made to simulate the vibration signal of REBs. The model adopted in this study, as depicted in Equation (11), was developed by Randall et al. [23]. It takes into consideration some of key parameters which characterise the typical bearing vibration signal such as the periodicity nature and the random variation in pulse spacing, moreover, the modulation might occur due to loading distribution effects.

$$x(t) = \sum_{j} A_j y(t - Tj - \tau_j) + n(t) \tag{11}$$

where, A_j is the amplitude of the j_{th} impact, T is the average time between two adjacent pulses and it is derived by $T = 1/f_z$, f_z stands for fault frequency, Tj is the expected j_{th} time of single impact occurrence,τ_j is the random uncertainty around it, $n(t)$is random noise produced by other vibrations in the system. $y(t)$ is the impulse response function, derived in (12), and is simplified as an exponential damping cosinusoidal signal.

$$y(t) = \begin{cases} e^{-\alpha t}\cos(2\pi f_d t) & ;t > 0 \\ 0 & ;otherwise \end{cases} \tag{12}$$

The resonance frequency f_d is set to be [3 kHz, 5 kHz, 8 kHz], with a=0.05 as a damping ratio. A vibration signal with outer race fault spaced with shaft frequency and another vibration signal with roller fault spaced with cage frequency is generated,A_j , can be simplified as

$$A_j = A_1 \cos(2\pi f_k (Tj + \tau_j)) \tag{13}$$

where A_1 stands for the amplitude of the possible modulator, f_k represents the shaft frequency,$Tj + \tau$ stands for the specific time of the j_{th} impact. Whilst, in the case of roller fault, f_k is considered as the cage frequency f_c and $f_r - f_c$. The generated signal for the studied defects is based on theoretical fault frequencies using the calculation formulas and the given parameters in Table 3.

Table 3 Parameters of the generated vibration signal

Parameter	Values for outer race fault	Values for roller fault
Amplitude of pulses(A_1)	0.9	0.9
Shaft frequency(f_k)	25Hz	25Hz
Sampling rate (f_s)	50kHz	50kHz
Fault frequency(f_r)	157.5Hz	133.7Hz

The network parameters shown in Table 4, were set to the same values found in the optimization process.

Table 4 Impact Signal Training Parameters

Parameter	Value
Weight vector number	8
Weight vector size	512
Threshold	0.4
Sigma	0.2
Learning rate	0.05
Batch number	Data length / WV number
Iterations	80

5.1.2 Simulating Data Results

Two aspects were considered with regard to detection performance. The first test of the network was to examine the influence of noise contamination on the network performance in detecting the anomalies. A generated signal is used with different noise levels to train and test the network. The portion of the noise is increased gradually from 14.4dB to -7.31B as seen in Figure 2 (a-d), where the signal shown in (d) was nearly buried in the noise. In first test, same signal structure with different noise levels is used to examine the network sensitivity to the influence of noise. The results as depicted in Figure 3 showed the capability of the network in detecting the hidden periodic components in the signal despite the high level of the random noise components. The ADI did not show any anomalies in the data. The test was conducted by using the same signal structures used during the training stage, each case was tested with the same SNR value used in the training stage. In this test, the network showed that it can discover the hidden patterns even with low SNR values of (-8 dB). The trend of ADI showed an increase with the decline of SNR because the increase of the noise level may change the signal structure, however, the network still able to discover the periodic compnents and thus, the network has not detcted any anomalies.

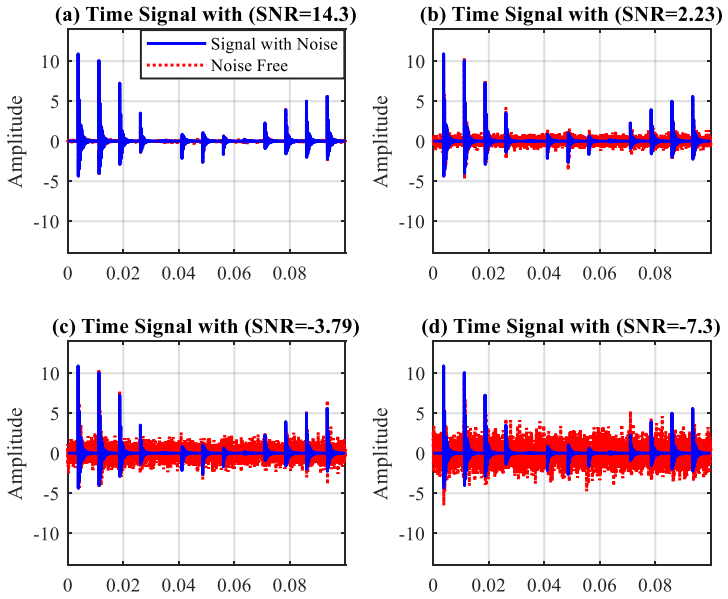

Figure 2 Simulated Signals with Noise used for Training and Testing CC

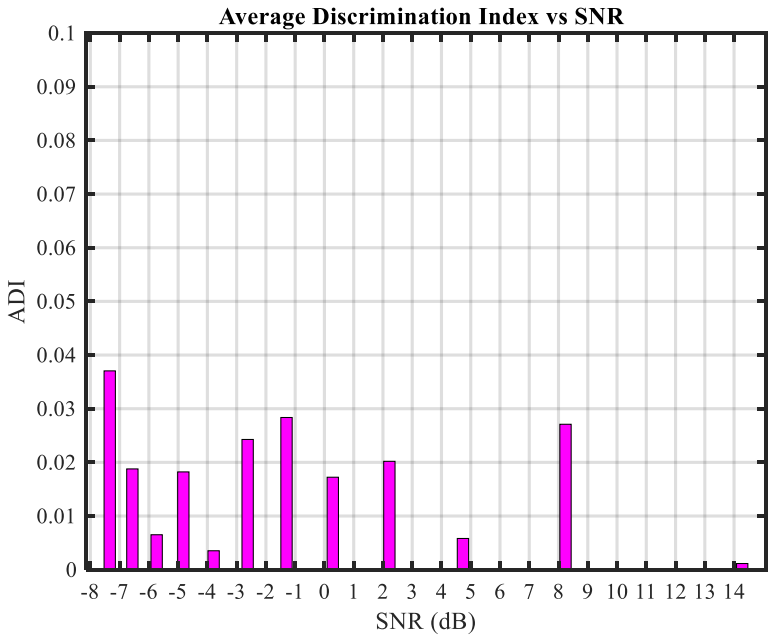

Figure 3 The influmnce of noise level on the network detection capablity

In second test, different signal structures with different noise levels is used to examine the sensitivity of the network in discovering the change in the signal structure in different noise levels. The test of the CCNN is to evaluate the network capability in identifying any changes in the hidden periodic patterns in the signal structure, i.e. the sensitivity to variation of the signal structure. Thus, a generated signals used to evaluate the performance of CCNN that simulate the outer race fault f_r = 157 Hz with different noise levels. The levels of noise used in the training and testing is the same. The results obtained as depicted in Figure 4 showed that CCNN can effectively detect the anomalies until the SNR reached (-6) with ADI of 0.14. In addition, the severity of the anomalies can be estimated in all the simulated cases. Despite the fact that the trend of the ADI amplitude decreases with the declines of the SNR levels, the overall the capability of CCNN in detecting the anomalies is robust and reliable even in the high level of noise contamination.

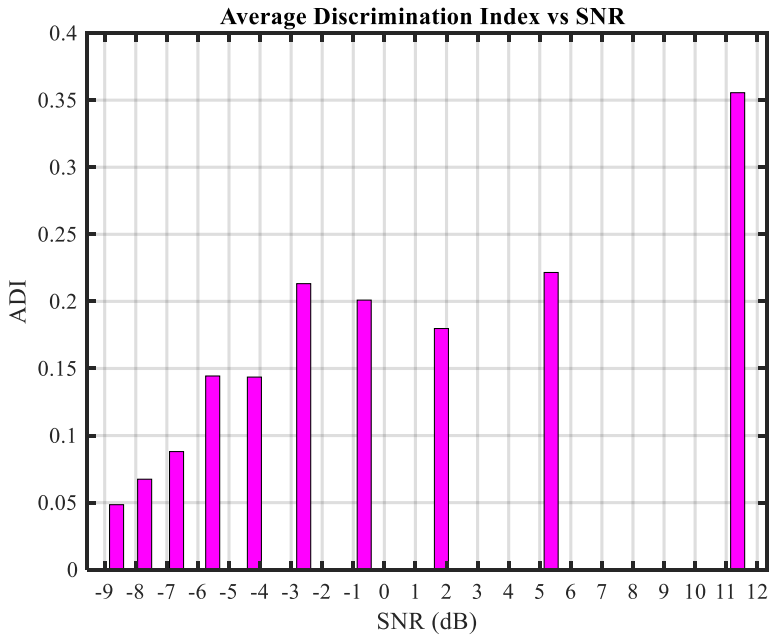

Figure 4 ADI Detection Performance of Simulated Signal with Outer Race Defect

5.2 Implementation of the CCNN to Experimental Data

Five datasets have been collected from tapered roller bearings. The first dataset was collected from a healthy bearing, the second and third datasets were collected from two defective bearings with small and large outer race fault respectively, and the fourth and fifth datasets belong to defective bearings with small and large roller fault respectively.

CCNN was trained using datasets collected during the healthy operating condition and another unseen healthy dataset was used for the validation stage as during the validation stage the network needs to be validated with unseen dataset. The Network configured with weight vectors of 32, more than the weight vectors number used for the simulated signal test as experimental data have more complex data structure. From the waveform of raw signals, each vibration measurements has 150000 data samples. This data length covers more than 750 shaft rotations. The interval between the expected faulty pulses of the outer race fault can be derived as $\Delta t = \frac{1}{f_{BPO}} * f_s$ which gives 320 data samples. For the roller fault $\Delta t = \frac{1}{f_{BS}} * f_s$ which gives 378 data samples, $F_s =$ 50kHz stands for the sampling rate. Therefore, the network was designed with weight vector size of 512 to cover at least one complete fault peak.

5.2.1 Outer Race Data Analysis

Figure 5 shows the baseline and two outer race faulty cases, small and large, for all clearance studied cases. The results showed the performance of CCNN in detecting the anomalies in discriminating accurately the fault severities.

In Figure 6, a comparison was made between the RMS of raw data (a) and the ADI results obtained from applying CCNN to the dataset with small and large outer race faults (b). The results

showed that CCNN can accurately discriminate the studied conditions as seen in (b) whilst, RMS results as seen in (a), showed no trend exists. In addition, CCNN can estimate the severity of the faults as the clearance declines.

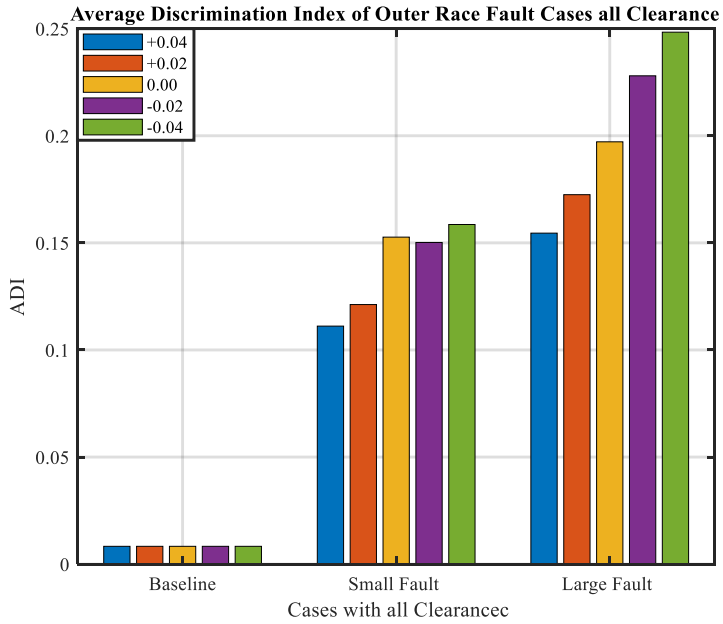

Figure 5 Baseline and Large Outer Race fault Data with Five clearances

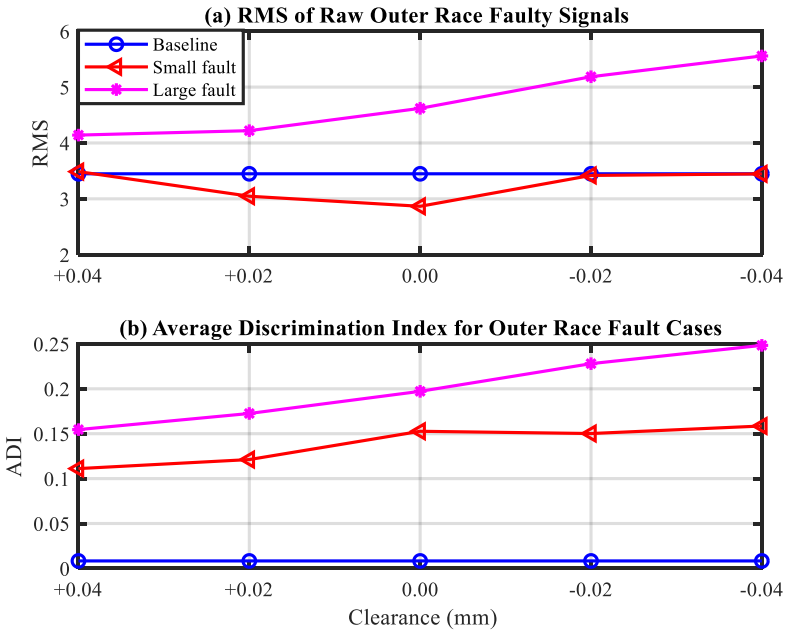

Figure 6 RMS of ADI of baseline and Outer Race Cases with all Clearances

5.2.2 Rolling Element Data Analysis

The same procedure used in the outer race case is adopted in roller fault test. Figure 7 shows the results obtained from the baseline, small and large roller fault conditions. ADI values of the overall detection results show that CCNN produces good detection results and discrimination. The comparison results in Figure 8 showed the capability CCNN in early fault detection task compared to the results of RMS in (a), the severity of the defects can be also estimated due to both the fault size or variation of the internal clearances. The amplitude of the small roller fault has an increasing trend from 0.15 to 0.2 as the preload increases.

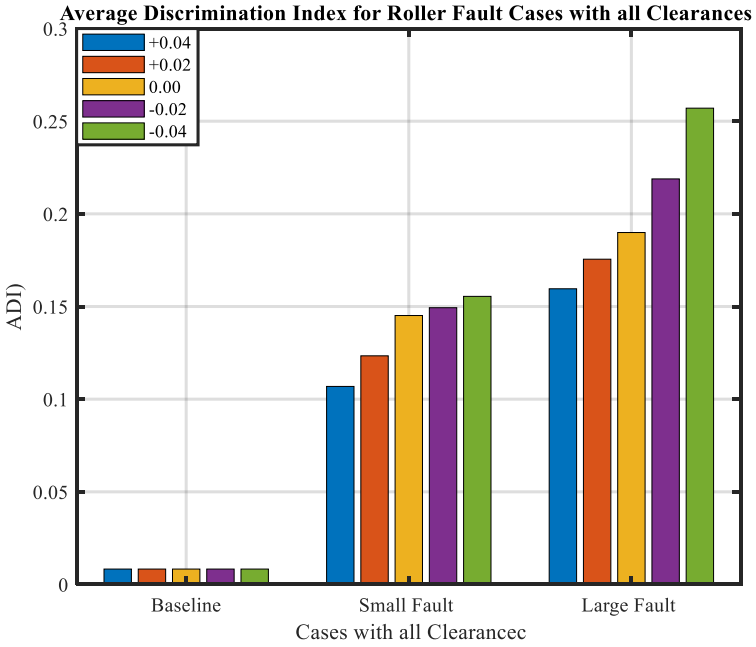

Figure 7 ADI of baseline, Small and Large Roller Fault

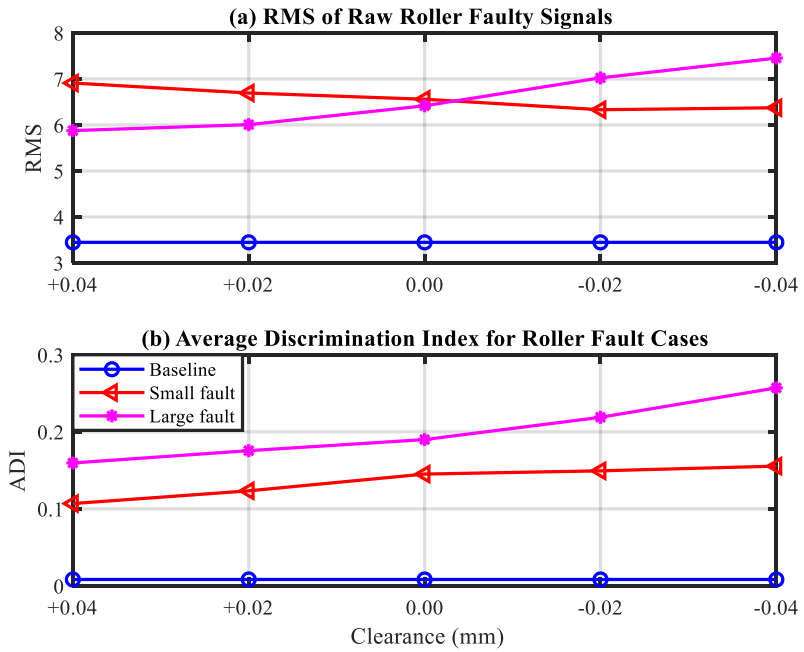

Figure 8 RMS of Raw Signal and ADI for Roller Fault Cases

6 Conclusions

The performance evaluation of CCNN demonstrated that the network has good performance in novelty detection and discrimination when applied to both simulated data and measured data. Both the detection and discrimination are validated through the variation of noise levels and change of the signal structure. The simulation results showed CCNN has a good performance in overall detection and discrimination and moreover, the CCNN is robust to noise even in high noise levels. For experimental studies, the proposed technique perform well in fault detection and severity assessment. The results confirmed that CCNN is an effective method to learn the diagnostic features from periodic impact signals and can detect different levels of the anomalies in the bearings. Due to using unsupervised learning algorithm, CCNN can be used without sny prior-knowledge about the data, and in addition, the proposed method can be applied directly to raw data. Moreover, the CCNN does perform well without the need to record angular position when acquiring the data as it has the feature of translation invariance. Consequently, these merits can improve the efficiency of implementing online condition monitoring.

References

1. Qu, J., Z. Zhang, and T. Gong, *A Novel Intelligent Method for Mechanical Fault Diagnosis Based on Dual-Tree Complex Wavelet Packet Transform and Multiple Classifier Fusion*. Neurocomputing, 2016. **171**: p. 837-853.
2. Tian, X., et al., *A Robust Detector for Rolling Element Bearing Condition Monitoring Based on the Modulation Signal Bispectrum and its Performance Evaluation Against The Kurtogram*. Mechanical Systems and Signal Processing, 2018. **100**: p. 167-187.

3. Halme, J. and P. Andersson, *Rolling Contact Fatigue and Wear Fundamentals for Rolling Bearing Diagnostics-State of the Art.* Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2010. **224**(4): p. 377-393.
4. Howard, I., *A Review of Rolling Element Bearing Vibration'Detection, Diagnosis and Prognosis'*. 1994, DTIC Document.
5. Nguyen-Schäfer, H., *Computational design of rolling bearings*. 2016: Springer.
6. Daniel Goerke, N.M.a.A.G., *Effects of Radial Clearance Changes on Vibration Frequencies in Double-Row Self-Aligning Ball Bearings: An Experimental Study.* Transactions of the Institute of Measurement and Control, 2017. **v1**(10): p. 1-14.
7. Rehab, I., et al., *A Study of the Diagnostic Amplitude of Rolling Bearing under increasing Radial Clearance using Modulation Signal Bispectrum.* 2016.
8. Khalid Rabeyee, X.T., Yuandong Xu, Dong Zhen, Fengshou Gu, Andrew Ball, *diagnosing the Change in the Internal Clearances of Rolling Element Bearings based on Vibration Signatures*, in *Automation & Computing*. 2018: Newcastle upon Tyne, UK.
9. Khalid Rabeyee, Xiaoli Tang, Fengshou Gu, Andrew Ball, *The Effect of the Wear Evolution on Vibration based Fault Detection in Tapered Roller Bearings*, in *Condition Monitoring and Machinery Failure Prevention Technologies 2018*. 2018: Nottingham UK.
10. Shakya, P., M.S. Kulkarni, and A.K. Darpe, *Bearing Diagnosis Based on Mahalanobis–Taguchi–Gram–Schmidt Method.* Journal of Sound and Vibration, 2015. **337**: p. 342-362.
11. Rai, A. and S. Upadhyay, *A Review on Signal Processing Techniques Utilized in the Fault Diagnosis of Rolling Element Bearings.* Tribology International, 2016. **96**: p. 289-306.
12. El-Thalji, I. and E. Jantunen, *A Summary of Fault Modelling and Predictive Health Monitoring of Rolling Element Bearings.* Mechanical Systems and Signal Processing, 2015. **60**: p. 252-272.
13. Doguer, T. and J. Strackeljan. *Vibration Analysis Using Time Domain Methods for the Detection of Small Roller Bearing Defects*. in *8th International Conference on Vibrations in Rotating Machines, Vienna, Austria*. 2009.
14. Al-Bugharbee, H. and I. Trendafilova, *A Fault Diagnosis Methodology for Rolling Element Bearings Based on Advanced Signal Pretreatment and Autoregressive Modelling*. Journal of Sound and Vibration, 2016. **369**: p. 246-265.
15. Chandra, N.H. and A. Sekhar, *Fault Detection in Rotor Bearing Systems Using Time Frequency Techniques.* Mechanical Systems and Signal Processing, 2016. **72**: p. 105-133.
16. Abdusslam, S.A., *Detection and Diagnosis of Rolling Element Bearing Faults Using Time Encoded Signal Processing and Recognition*. 2012, University of Huddersfield.
17. Jiang, R., et al., *A Novel Method of Fault Diagnosis for Rolling Element Bearings Based on the Accumulated Envelope Spectrum Of The Wavelet Packet*. Journal of Vibration and Control, 2015. **21**(8): p. 1580-1593.
18. Webber, C.J., *Self-organisation of Transformation-Invariant Detectors for Constituents of Perceptual Patterns.* Network: Computation in Neural Systems, 1994. **5**(4): p. 471-496.
19. Webber, C.J., *Emergent Componential Coding of A Handwritten Image Database by Neural Self-Organisation.* Network: Computation in Neural Systems, 1998. **9**(4): p. 433-447.
20. Webber, C.J., *Generalisation and Discrimination Emerge From A Self-organising Componential Network: A Speech Example.* Network: Computation in Neural Systems, 1997. **8**(4): p. 425-440.
21. Webber, C., et al., *Componential Coding in the Condition Monitoring of Electrical Machines Part 1: principles and illustrations using simulated typical Faults.* Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science, 2003. **217**(8): p. 883-899.
22. The-Crankshaft Publishing. *Slip Gauges (Metrology)*. 2017 [cited 2017 1 Dec]; Available from: http://what-when-how.com/metrology/slip-gauges-metrology/.
23. Randall, R.B., J. Antoni, and S. Chobsaard, *The Relationship Between Spectral Correlation and Envelope Analysis in the Diagnostics of Bearing Faults and Other Cyclostationary Machine Signals.* Mechanical systems and signal processing, 2001. **15**(5): p. 945-962.