Songklanakarin J. Sci. Technol.
43 (5), 1482-1490, Sep. - Oct. 2021

http://www.sjst.psu.ac.th

Original Article

Automating plastic bearing fault diagnosis using continuous wavelet transform and neural networks

Saipul Azmi Mohd Hashim[1*], N. Abdul Razak[2], and N. S. Sholahuddin[1]

[1] *Manufacturing Technology Unit, Community College, Kepala Batas, Penang, 13200 Malaysia*

[2] *Penang Matriculation College, Kepala Batas, Penang, 13200 Malaysia*

Received: 30 April 2019; Revised: 1 December 2020; Accepted: 15 December 2020

Abstract

Faulty plastic bearing is an initial alarm for bearing failure that can cause massive losses in the production line. The losses include restarting of production, producing of defective products, and even human casualty. This paper therefore aims to propose an automated plastic bearing fault detection and classification system. The system begins by transforming the bearing vibration signals into coefficients with continuous wavelet transform. The coefficients are then filtered by coefficients reduction and smoothing thereafter. Then, the filtered coefficients are classified by two ANN classifiers i.e. feed forward backpropagation (FFB) and recurrent neural network (RNN). The performance of both classifiers are finally measured and compared. The best overall performance is 90% detection rate by FFB. This system prevents bearing failure by giving an early alarm for faults detection and making the corrective action easier. Also, the single stage data processing and single signal type increase the data processing efficiency.

Keywords: bearing fault diagnosis, continuous wavelet transform, exponential moving average, neural network, principle component analysis

1. Introduction

Plastic bearing application is a common requirement in light weight products production line. Due to its lack of strength (Ning *et al*., 2015), it supports light weight products and is applied widely in electronic components, food processing, and medicine production industries. The bearing becomes industries preference because of its behaviors such as non-corrosion, light weight, non-magnetic, lubrication-free, good friction, and easily machined (Koike *et al*., 2011), which lessen product constrains in production line application. However, the major cause of breakdown in rotating machinery is bearing failure (Li *et al*., 1999). The failure may result in huge economy and human losses, e.g. restarting of production and producing defective products. Production line that cannot tolerate with sudden breakdown (Smadi & Kamrani, 2011) e.g. cable production and wire-cut, needs to be restarted with a new preparatory production run.

A major source of conveyor belt malfunction is due to bearing failure (Lodewijks, 2004). Unstable conveyor belt due to bearing faults may result in vibration that leads to defective electronic products. A production line that holds hazardous chemical or products at high points may fall on and endanger production workers due to the breakdown. Hence, early detection and classifying of the bearing faults system are needed to prevent bearing failure in production line.

However, standard bearing monitoring in detection and classifying the faults are proven ineffective (Yang & Court, 2013). So an effective, consistent, and reliable system is required. The system must consider the behavior of the bearing critically. Plastic bearing behavior is more loosely assembled and easier to deform i.e. insufficient accuracy (Riahi & Stadlmayr, 2012). This complex behavior widens fault vibration beat characters and weakens fault beat signals. Consequently, this issue increases the difficulty in bearing faults traceability (He, Li, & Zu, 2012). To date, plastic

*Corresponding author
Email address: saipulneural@gmail.com

bearing faults diagnosis' study attracted less researchers' attention. Only a study of plastic bearing faults diagnosis (He *et al*., 2012) is recorded. Thus, plastic bearing fault diagnosis is the prime focus of this study.

Between plastic and steel bearing faults diagnosis, the latter attracted extensive researchers' attention. In previous researches, the initial main stage was extracting vibration signal features. Researchers extracted signal features by empirical-mode decomposition (Ali *et al*., 2015) and statistical analysis (Immovilli, Bellini, & Rubini, 2009; Immovilli, Bianchini, Cocconcelli, Bellini, & Rubini, 2012; Liang & Bozchalooi, 2010; Zarei, Tajeddin, & Karimi, 2014). Packet wavelet used in this study line began by Eren and Devaney (2004); this then was followed by Peter & Wang (2013); Wu & Liu (2009); Pan, Chen, and Guo (2009), and Pan, Chen, and Li (2010). Continuous wavelet transform (CWT) was also used in similar studies (Immovilli, Rubini, & Tassoni, 2010; Konar & Cattopadhyay, 2011; Rafiee, Rafiee, & Tse, 2010).

The second main stage is defining the extracted features into their fault classes by classification methods. The methods include statistical model (Immovilli *et al*., 2010; Lau & Ngan, 2010; Peter & Wang, 2013; Rafiee *et al*., 2010), neural network (Ali *et al*., 2015; Wu & Liu, 2009; Zarei *et al*., 2014), support vector machine (Konar & Chattopadhyay, 2011; Shen *et al*., 2013), fuzzy c-means (Pan *et al*., 2010); vibration based envelope analysis (Peter & Wang, 2013), adaptive neuro-fuzzy (Zhang *et al*., 2010), and k-nearest neighbor (Huang, Hu, & Yang, 2011).

In this paper, an automated plastic bearing faults diagnosis method shown in Figure 1 is proposed. This paragraph lists the different use of methods than the paper where the data is tapped (He *et al*., 2012). The obvious difference is, this paper uses only time-domain bearing fault signals. This signal is transformed by other approach, i.e. CWT, into wavelet coefficients. Also, the classification method is simplified into a single stage. These coefficients are classified by neural network (ANN) classifiers – FFB and RNN. In reducing coefficient dimensions and noises, two filters are applied before the classification stage. The filters are principle component analysis (PCA) and exponential moving average (EMA).

The subsequent sections of this paper is presented as follows. Section 2 introduces CWT and justifies the need for CWT. Sections 3 and 4 highlight filters for dimension and noise reduction. Section 5 describes the ANN character selection and operation. Result, discussion, and conclusion can be found in Section 6 and 7.

2. Continuous Wavelet Transform

Plastic bearing vibration signals require signal transformation into coefficients in order to analyze the signals. Conventionally, the Fourier transform (FT) (Eren & Devaney, 2004) is applied to transform the signal into coefficients. But FT is inefficient in analyzing spike shape vibration signal (He *et al*., 2012). It is also not a time-dependent analysis (Sadowsky, 1994). As the used signals are in spike shape and this study requires a time indication of faults detection, thus other analysis is needed. According to Graps (1995), wavelet transform (WT) analysis is good in transforming the spike shape signal. The analysis also is time-dependent (Wu & Liu, 2009). Due to the used signal in continuous-time and the transformed signals, i.e., coefficients are required in the continuous periodic scale – CWT is suitable for these needs.

$$CT(a,b) = \int_{-\infty}^{\infty} x(t)\psi^*_{(a,b)}(t)dt \tag{1}$$

$$\psi^*_{(a,b)}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \tag{2}$$

$$a, b \in R, a \neq 0$$

The CWT function (Konar & Cattopadhyay, 2011) is based on time $x(t)$ and complex conjugate operation (*) as in equation (1). It is the summation of all signal times multiplied by scale. The scale in continuous function ($\psi^*_{(a,b)}(t)$) with real number (R), a and b are used to adjust wavelet scale and position (2). The output is wavelet coefficients in multiple dimensions. Basically, the output is in classes and each class has subclasses, i.e. components. Each component contains a set of coefficients. The use of CWT is set to the highest class scale i.e. 32 classes. This is to cover the widest possible coefficients generation. Based on the scale, each class has five components and each component has 139 coefficients. This also means the total number of components are 160 (32 × 5 = 160) and the total number of coefficients are 22,240 (32 × 5 × 139 = 22,240). This high-dimensional data i.e. coefficients are hard to represent a system. Thus, coefficient's dimensional reduction, i.e. classes are applied to simplify the data.

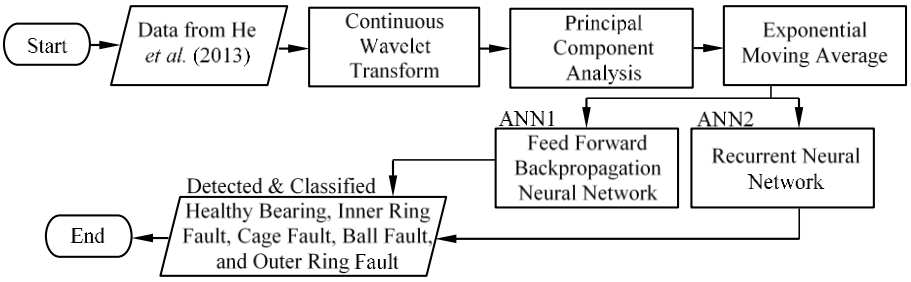

Figure 1. Flowchart of the methodology

3. Principal Component Analysis

PCA is used to reduce high-dimensional data (Nie *et al.*, 2011) that is generated by CWT. The analysis eliminates the redundancy of data (Jolliffe, 2011) i.e. coefficients' classes. The redundancy is based on correlation distance or variance (Shlens, 2014) among coefficients' classes. The low variance classes that represent high co01efficient patterns redundancy are eliminated, while classes with high variance are retained (Ringnér, 2008) for further study. Based on the analysis, this study selected classes with meaningful information systematically represented by high variance classes.

$$X = \underbrace{\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}}_{m\ dimensions} \Big\} n\ samples \tag{3}$$

$$\frac{1}{n} X^T X = C_X \tag{4}$$

The assumption is that some dimensions are correlated in patterns (Shlens, 2014). Thus, PCA deduces these patterns. Each coefficient has two covariance matrix (X) dimensions, as shown in (3). PCA is Eigen-decomposed of X multiplied by the transpose of $X(X^T)$. This resulted C_X as in (4), which contained an eigenvector (W) and eigenvalues (λ) set. These eigenvalues are coefficients' classes variance value in percentage. The variance value of the studied classes will be the base to pick cut-off variance value for classes selection (Demuth & Beale, 2009). Because there are classes with eigenvalues that reached above 90%, this variance value (above 90%) is set for this study. Based on this set, only five are selected out of the total 32 classes.

4. Exponential Moving Average

EMA (μ_n) is effective for pattern indicator (Klinker, 2011). Also, EMA corrects and removes noises in signal processing and data analysis (Grenbenkov & Serror, 2014 Awheda & Schwartz, 2014). Nakano, Takahashi, and Takahashi (2017) applied EMA to simplify signal for classification operation. In this study, EMA extracts important information within the components that contained 139 coefficients. It smoothens short series fluctuation and highlights longer series of coefficient trends. Here, EMA acts as a low pass filter on these trends and weakens high vibration readings that impart noises in the trends.

$$\mu_n = x_n\left(\frac{S}{1+n}\right) + \mu_{n-1}\left(1 - \frac{S}{1+n}\right) \tag{5}$$

New EMA calculated as in equation (5), from new sample of coefficient (X_n) and reused previous EMA (μ_{n-1}). As in simple moving average formulation, smoothing (S) that is set as two was used. The setting is to average two points of value – initial value at the beginning point and accumulative value at the end point. The number of distance in EMA (n) must consider the most effective distance for the data usage. Hence, trials are made between n = 2 up to n = 80. Here n = 70 is the most effective for use by the ANN classifiers. The n determines the number of processed coefficients. n = 70 and initial coefficients of 139, resulted the processed coefficients of 69. PCA in the last stage reduces data into five classes and each class has five elements, thus 25 EMA operations are needed.

5. Artificial Neural Network

The proposed system must be designed to meet the selected classifiers' requirements i.e. ANN. ANN classifiers perform well on any analytical operation that deduces a function (Ali *et al.*, 2015) from the used data or coefficients. Thus, ANN works analytically well on linear or non-linear or even mixed-mode relationship system. By ANN the system analytical model design is not required (Zahran *et al.*, 2013; Zain, Haron, & Sharif, 2010) which is impractical for this high-complexity data. Also, ANN is excellent with multi-input and data fusion (Kumar, Natarajan, & Ananthan, 2012; Wu & Liu 2009), which are expected to deal well with the study coefficients. Furthermore, ANN is a non-parametric approach (Belkacem, Bouafia, & Chabani, 2017) which is independent to estimator parameters – fastens the classifiers modelling process. All of these features and previous studies in faults diagnosis (Ali *et al.*, 2015; Kumar *et al.*, 2012; Wu & Liu, 2009) make ANN classifiers' requirements suitable for this study. The model is designed detecting and classifying inputs into plastic bearing faults, namely healthy bearing, outer-ring fault, inner-ring fault, ball fault, and cage fault. Based on PCA and EMA outputs, there are five faults i.e. five classes (each class has five components and each component contained 69 coefficients). The total of 1725 (5 × 5 × 69 = 1,725) coefficients will be used to classify the two ANN classifier models i.e. FFB and RNN in this study.

For ANN models' prediction generalization purpose; composition of learning and testing dataset that set the generalization (Kant & Sangwan, 2015; Zain *et al.*, 2010;) must be prepared. The datasets also must be organized so that the information pattern exists in the learning dataset must be existed in testing dataset (Ioannis & Dimitrious, 2010). For a guideline, learning and testing dataset are set to 70%:30% composition (Nahato, Harichandran, & Arputharaj, 2015; Zahran, *et al.*, 2013). Hence, the 1,725 coefficients are set by this composition in learning and testing datasets through 25 component turns.

$$\text{Detection rate} = \frac{\text{Number of testing dataset with correctly detected}}{\text{Total number of testing dataset}} \text{ X } 100\% \tag{6}$$

The overall classifier performances are evaluated by detection rate (Kumar *et al.*, 2012) as per equation (6). The performance of various function-combinations and architectures set to the models are based on this rate. Although the equation is directly unrelated to the use function-combinations performance, the best classifiers overall performance pin-pointed the best function-combination. Moreover, this also pin-pointed the best classifier architecture. Thus, discussion of the two ANN classifiers, i.e. feed forward backpropagation (FFB) and recurrent neural network (RNN)

can be specifically compared. In applying full-factorial experiment, the aforementioned parameters must be fully tested and discussed in the following sections.

5.1 Feed forward backpropagation network

FFB neural networks demonstrated excellent detection rate which is above 90% by Belkacem *et al*. (2017); Kant and Sangwan (2015); and Kumar *et al*. (2012). FFB algorithm learns by evaluating output difference between generated and actual output. The difference propagates backward to input layer to modify the network's weights and repeats until the smallest output variation is obtained (Zain *et al*., 2010). The best weights are reused to predict testing dataset.

The selected four FFB network functions i.e. training, learning, performance, and transfer functions are shown in Table 1. In the learning process stage, neuron weights are updated by the learning function. In order to transfer outputs from one neuron to the following layer and neuron, transfer function is required. Then the final output is evaluated by performance function. All of these processes are repetitive tasks until the smallest variation between generated and actual output is obtained. Based on this construction, learning dataset is used to simulate and generate the needed weights and the testing dataset is used to test the FFB classifier with these weights.

For the models prediction generalization purpose; composition of learning and testing dataset that sets the generalization (Kant & Sangwan, 2015; Zain *et al*., 2010), must be prepared. Researchers (Nahato *et al*., 2015; Zahran *et al*., 2013) suggest the ratio of learning and testing dataset for using ANN classifier is 70%:30% accordingly. Besides, testing dataset with sufficient information as in learning dataset is compulsory (Ioannis & Dimitrious, 2010; Kumar *et al*., 2012). Thus, the existence of logical association between the datasets ensure reasonable prediction result. In order to follow this guideline, every element with 69 coefficients is divided into 48 coefficients (70%) for learning dataset and 21 coefficients (30%) for testing dataset. In addition, each dataset division is processed to ensure information in learning dataset exists in the testing dataset.

In this paper, FFB network's architecture is applied with a single hidden layer. Based on random tests of FFB classifier model simulation, seven to ten neurons in hidden layer are the most effective in generating good prediction rate. Thus, a model with seven to ten neurons in single hidden layer architecture is used to for the FFB simulation, represented by the symbols of 5–7–1, 5–8–1, 5–9–1, 5–10–1, and 5–7–1 architecture pictured in Figure 2. Besides, the network's neurons are set to five inputs and a single output. The five inputs are represented by five bearing faults and the single output represents the classified bearing fault.

Data is generated in five class of faults with each class has five components and each component has five 69 coefficients. Due to the structure of the data and to fit this data into ANN model, ANN simulation iteration is done by batches. Each batch is equivalent to component (containing all the 69 coefficients). By this, completing an epoch requires five iterations. Initially, epoch numbers are set to 3,000. However, this study takes an early stopping technique, hence termination will be based on mean square error (MSE) rate. The software used in this study is neural network toolbox in MATLAB 2015. Here, the monitoring of MSE and iteration status is done through neural network training menu of the software. Based on MSE rate monitoring the converge starts at 940 epochs for the best FFB model classifier. After 940 epochs, MSE starts increasing and this indicates pattern replication overfitting.

This study covers all the combination of possible FFB ANN aforementioned functions and architectures. The probability for four training functions are 24 ($4 \times 3 \times 2 \times 1 = 24$), two functions for each of the rest function is 2 ($2 \times 1 = 2$). For four architectures setup the probability is 24 ($4 \times 3 \times 2 \times 1 = 24$). Thus, for full factorial experiment exercise the total number of combinations are 4,608 ($24 \times 2 \times 2 \times 2 \times 24 = 4{,}608$). These combination numbers represent the required test numbers.

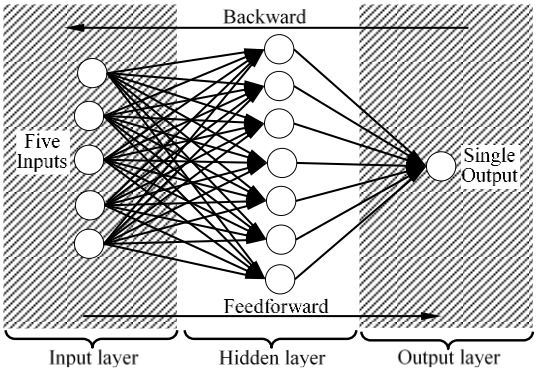

Figure 2. Feed forward backpropagation network with 5-7-1 architecture

Table 1. Applied functions with symbols

Training function	Learning function	Performance function	Transfer function
Levenberg-Marquardt Backpropagation (LMF)	Gradient Descent with Momentum Weight (GMB)	Mean Squared Normalized Error Performance (MEP)	Hyperbolic Tangent Sigmoid (HTS)
Random order Incremental Training with Learning Functions (RLF)	Bias Learning and Gradient Descent Weight and Bias Learning (GWB)	Mean Squared Normalized Error Performance Function with Regularization Implemented (MRI)	Linear Transfer (LTR)
Scaled Conjugate Gradient Backpropagation (SCB)			
Resilient Backpropagation (RBP)			

5.2 Recurrent neural network

RNN network model is tested to compare it with FFB network model performance. Similar to FFB, RNN also has a feedback loop (Du, Wang, & Wang, 2015). However, in each layer, RNN has a recurrent connection and network delay which has infinite dynamic response to time series input data (Anbazhagan, & Kumarappan, 2012). Compared to FFB, RNN has several advantages which include better accuracy, faster training, and lesser training data required (Negnevitsky, 2011). RNN has better accuracy because it runs classification operation in network's input and internal (Guo *et al.*, 2017). So, when two classification stages are applied, a more detailed classification can be obtained. In contrast, FFB only operates at network's input that has lower accuracy.

$$h_T = H(W_{xh}x_T + W_{hh}h_{T-1} + b_h) \quad (7)$$

$$y_T = O(W_{ho}h_T + b_o) \quad (8)$$

RNN network models a system based on temporal sequence (T). For the RNN network, let's input x in sequence $x = (x_0, \ldots, x_{T-1})$ with recurrent hidden state h where $h = (h_0, \ldots, h_{T-1})$. RNN output of $y = (y_0, \ldots, y_{T-1})$, is based on two biases i.e. b_0 and b_h, also two transfer functions applied on RNN model at hidden $H(\cdot)$ and output $O(\cdot)$ layers. Three required connection weights include weight between input layer to hidden layer (W_{xh}), the weight within hidden layer (W_{hh}), and the weight between hidden layer to output layer (W_{h0}). Based on these parameters, recurrent hidden state at specified temporal sequence (H_T) in equation (7) is used to generate output of RNN at specified temporal sequence (Y_T) as in equation (8).

For fair comparison, RNN setting parameters are set similar to FFB network. Network functions, architectures, and composition of training and testing datasets are the similar setting parameters which are applied to both network models. For full factorial experiment similar 4,608 tests are required. Based on MSE monitoring, the best RNN model classifier starts to converge at 1,080 epochs. MSE increases and indicates the condition of overfitting after 1,080 epochs.

6. Results and Discussion

In this study, five signals of bearing faults are monitored. The fault classes are healthy bearing, outer ring fault, inner ring fault, ball fault, and cage fault. Through CWT analysis, the fault classes are composed by 160 components, and each component has 139 coefficients. The high dimensions of data use PCA and EMA to reduce the components and amount of coefficients. By PCA the 160 components are decreased into five components. EMA reduces the coefficient numbers from 139 to 69. The remaining components and coefficients are shown in Figure 3(a) to (e).

ANN model performance is tested with full factorial experiment to demonstrate the best overall classifiers effectiveness in plastic bearing faults classification, the best used functions, and the best architecture setups. The top ten highest detection rate test results are tabled by the classifier networks, Table 2 for FFB and Table 3 for RNN. These tables

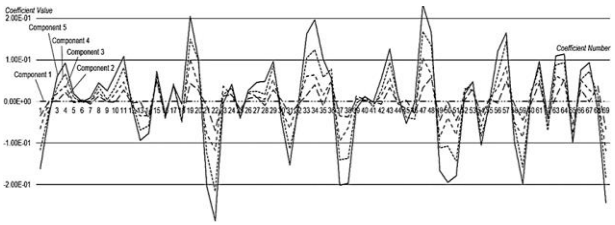

Figure 3. (a) Coefficient value – healthy bearing coefficient number (Class 1)

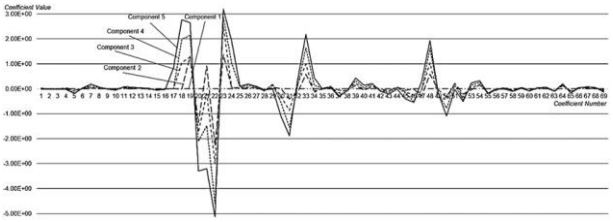

Figure 3 (b) Coefficient value – inner ring fault coefficient number (Class 2)

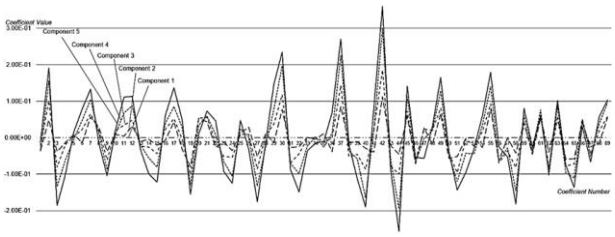

Figure 3. (c) Coefficient value – cage fault coefficient number (Class 3)

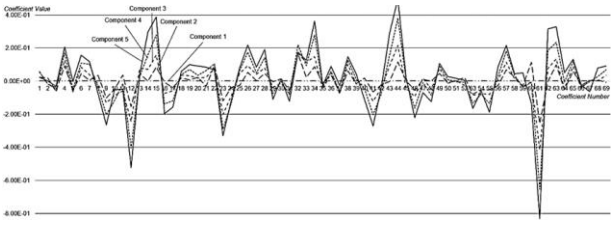

Figure 3. (d) Coefficient value – ball fault coefficient number (Class 4)

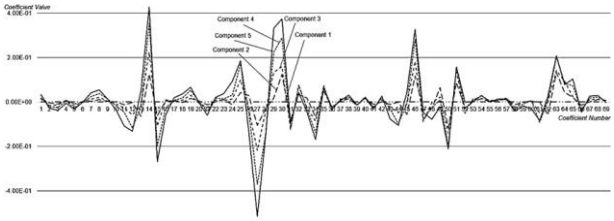

Figure 3. (e) Coefficient value – outer ring fault coefficient number (Class 5)

Table 2. Top 10 highest overall FFB detection rate

Sym	Archi-tecture	Function				DR	Sym	Archi-tecture	Function				DR
		TR	LR	PR	TF	(%)			TR	LR	PR	TF	(%)
A	5-7-1	LMF	GMB	MEP	HTS	83	F	5-9-1	RLF	GMB	MEP	LTR	87
B	5-7-1	RLF	GMB	MEP	HTS	83	G	5-9-1	RLF	GMB	MRI	HTS	87
C	5-8-1	LMF	GMB	MEP	HTS	83	H	5-9-1	RLF	GMB	MEP	HTS	90
D	5-8-1	RLF	GMB	MEP	HTS	87	I	5-10-1	RLF	GMB	MEP	HTS	87
E	5-9-1	LMF	GMB	MEP	HTS	87	J	5-10-1	LMF	GMB	MEP	HTS	83

Note: Sym – Symbol; TR – Training; LR – Learning; PR – Performance; TF – Transfer; DR – Detection Rate; RLF – Random order Incremental Training with Learning Functions; LMF – Levenberg-Marquardt Backpropagation; GMB – Gradient Descent with Momentum Weight and Bias Learning; MEP – Mean Squared Normalized Error Performance; MRI – Mean Squared Normalized Error Performance Function with Regularization Implemented; HTS – Hyperbolic Tangent Sigmoid; LTR – Linear Transfer

Table 3. Top 10 highest overall RNN detection rate

Sym	Archi-tecture	Function				DR	Sym	Archi-tecture	Function				DR
		TR	LR	PR	TF	(%)			TR	LR	PR	TF	(%)
A	5-10-1	LMF	GMB	MEP	HTS	60	F	5-7-1	RBP	GMB	MEP	HTS	40
B	5-9-1	LMF	GMB	MEP	HTS	60	G	5-10-1	RLF	GMB	MEP	HTS	40
C	5-8-1	LMF	GMB	MEP	HTS	60	H	5-7-1	SCB	GMB	MEP	HTS	40
D	5-7-1	LMF	GMB	MEP	HTS	70	I	5-10-1	LMF	GMB	MRI	HTS	40
E	5-7-1	RLF	GMB	MEP	HTS	60	J	5-10-1	LMF	GMB	MEP	HTS	40

Note: Sym – Symbol; TR – Training; LR – Learning; PR – Performance; TF – Transfer; DR – Detection Rate; RLF – Random order Incremental Training with Learning Functions; LMF – Levenberg-Marquardt Backpropagation; RBP – Resilient Backpropagation; SCB – Scaled Conjugate Gradient Backpropagation; GMB – Gradient Descent with Momentum Weight and Bias Learning; MEP – Mean Squared Normalized Error Performance; MRI – Mean Squared Normalized Error Performance Function with Regularization Implemented; HTS – Hyperbolic Tangent Sigmoid

list setup architectures, four-function combinations, and overall detection rates. Based on the tables, histograms are projected as in Figure 4(a) and Figure 5(a). These figures show the overall performance by detection rate value. The overall highest classifier score for FFB is 90% and 70% for RNN. These highest scores reflect the most effective architecture and function-combination used in this study. Hence, the most effective FFB architecture is 5-9-1 while for RNN is 5-7-1. Besides, the most effective FFB and RNN function-combinations are RLF-GMB-MEP-HTS and LMF-GMB-MEP-HTS.

In addition, based on these ANN classifier model simulations, each bearing fault detection rate is gathered. The performance of FFB classifier model in Figure 4(b) shows healthy bearing, cage fault, and inner ring fault are excellent for FFB. However, Figure 5(b) shows RNN classifier model is excellent for ball fault and outer ring fault. These figures also show the inefficient performance of faults detection rate for the classifier networks. Ball fault and outer ring fault are inefficient FFB classification performance. Healthy bearing, cage fault, and inner ring fault are inefficient for RNN classifier. The summary of detection rate by function-combination and bearing fault is presented in Table 4.

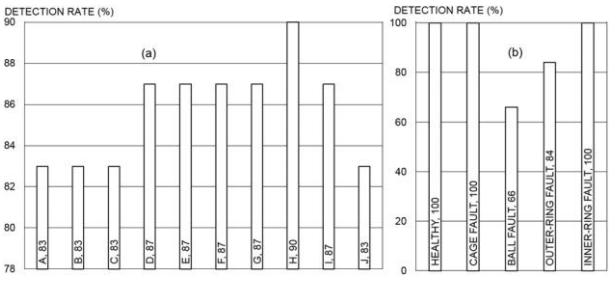

Figure 4. FFB Detection rate: (a) Overall; (b) Each Bearing Condition

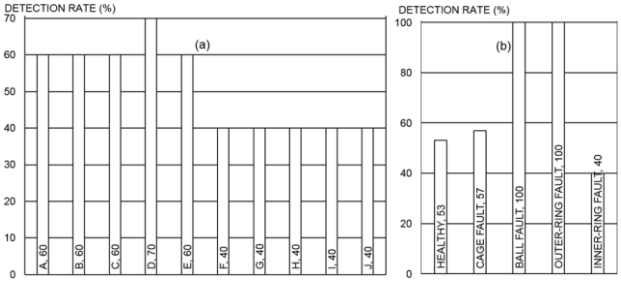

Figure 5. RNN Detection rate: (a) Overall; (b) Each Bearing Condition

7. Conclusions

The performance at 90% detection rate demonstrates that this study has successfully proposed an effective automated plastic bearing fault detection and classification system. CWT is used for signal transformation, CPA reduces the coefficient classes, EMA smoothens the coefficients, and these coefficients are classified by FFB and RNN. These set of sequence methods are well combined for an effective system of this study.

Table 4. Summary of FFB and RNN detection rate

	FFB		RNN	
	Symbol or Bearing Fault	DR (%)	Symbol or Bearing Fault	DR (%)
DR by function-combination	H	90	D	70
	D, E, F, G, & I	87	A, B, C, & E	60
	A, B, C, & J	83	F, G, H, I, & J	40
DR by bearing fault	Healthy	100	Healthy	53
	Cage Fault	100	Cage Fault	57
	Ball Fault	66	Ball Fault	100
	Outer-Ring Fault	84	Outer-Ring Fault	100
	Inner-Ring Fault	100	Inner-Ring Fault	40

Note: DR – Detection Rate; Please refer to table 2 and table 3 for the symbols usage

ANN classifier tests point a few number of thoughts. First, the two classifiers, i.e. FFB and RNN, show a classification of capable divergence. FFB classifies healthy bearing, cage fault, and inner ring fault effectively, while RNN classifies efficiently ball fault and outer ring fault. Second, architecture-wise, a more complex architecture does not necessarily reflect a better bearing faults classification. Moderate hidden-layer neurons are better classifier architecture. In this study, the best architectures are seven and nine neurons in hidden-layer. Third, ANN function-combination shows a significant point. The best function-combination for FFB is RLF-GMB-MEP-HTS and RNN is LMF-GMB-MEP-HTS. This also shows that the three functions, i.e. GMB-MEP-HTS are excellent for both classifiers. The functions represent class function of learning, performance, and transfer functions. For training function, certain function is effective for certain network. Industrial-wise, the findings suggest a new automated plastic bearing faults diagnosis system.

The used of the proposed system is significant for production line using machines installed with plastic bearing. The system with 90% detection rate is expected to detect and classify plastic bearing faults that occur in production line machines effectively. In line with this, early prevention action of machine losses can be taken by the faults detection. Corrective action is easier with the classified faults.

Based on the results, some recommendations can be made. For future study two stages of ANN classification are more effective. The successive stages can be a combination of FFB and RNN. This is because both networks are effective classifier on particular and contrast faults. Foregoing features are accountable for future model design such as the use of other functions and filters. Besides, deep learning classification approach is suggested because of the cascading data through multiple processing layer for classification. This feature is required for the proposed method for classification operation.

References

Boussaïd, I., Lepagnot, J. & Siarry, P. (2013). A survey on optimization metaheuristics. *Information Sciences*, *237*, 82-117.

Ali, J. B., Fnaiech, N., Saidi, L., Chebel-Morello, B., & Fnaiech, F. (2015). Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Applied Acoustics*, *89*, 16-27.

Anbazhagan, S., & Kumarappan, N. (2012). Day-ahead deregulated electricity market price forecasting using recurrent neural network. *IEEE Systems Journal*, *7*(4), 866-872.

Awheda, M. D., & Schwartz, H. M. (2013, April). Exponential moving average Q-learning algorithm. *Proceeding of the 2013 IEEE Symposium on Adaptive Dynamic Programming and Reinforcement Learning (ADPRL)* (pp. 31-38).

Belkacem, S., Bouafia, S., & Chabani, M. (2017). Study of oxytetracycline degradation by means of anodic oxidation process using platinized titanium (Ti/Pt) anode and modeling by artificial neural networks. *Process Safety and Environmental Protection*, *111*, 170-179.

Demuth, H., & Beale, M. (2009). Matlab neural network toolbox user's guide version 6. Portola Valley, CA: The MathWorks Inc.

Du, Y., Wang, W., & Wang, L. (2015). Hierarchical recurrent neural network for skeleton based action recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1110-1118).

Eren, L., & Devaney, M. J. (2004). Bearing damage detection via wavelet packet decomposition of the stator current. *Proceedings of the IEEE Transactions on Instrumentation and Measurement, 53(2)*, 431-436.

Graps, A. (1995). An introduction to wavelets. *Proceedings of the IEEE Computational Science and Engineering*, *2*(2), 50-61.

Grebenkov, D. S., & Serror, J. (2014). Following a trend with an exponential moving average: Analytical results for a Gaussian model. *Physica A: Statistical Mechanics and its Applications*, *394*, 288-303.

Guo, L., Li, N., Jia, F., Lei, Y., & Lin, J. (2017). A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing*, *240*, 98-109.

He, D., Li, R., & Zhu, J. (2012). Plastic bearing fault diagnosis based on a two-step data mining approach. *IEEE Transactions on Industrial Electronics*, *60*(8), 3429-3440.

Huang, J., Hu, X., & Yang, F. (2011). Support vector machine with genetic algorithm for machinery fault diagnosis of high voltage circuit breaker. *Measurement*, *44*(6), 1018-1027.

Immovilli, F., Cocconcelli, M., Bellini, A., & Rubini, R. (2009). Detection of generalized-roughness bearing fault by spectral-kurtosis energy of vibration or current signals. *IEEE Transactions on Industrial Electronics, 56(*11*)*, 4710-4717.

Immovilli, F., Bellini, A., Rubini, R., & Tassoni, C. (2010). Diagnosis bearing faults in induction machines by vibration or current signals: A critical comparison. *IEEE Transactions on Industry Applications*, *46*(4), 1350-1359.

Immovilli, F., Bianchini, C., Cocconcelli, M., Bellini, A., & Rubini, R. (2012). Bearing fault model induction motor with externally induced vibration. *IEEE Transactions on Industrial Electronics, 60(*8*)*, 3408-3418.

Jolliffe, I. (2011). *Principal component analysis* (pp. 1094-1096). Berlin, Germany: Springer.

Kant, G., & Sangwan, K. S. (2015). Predictive modelling and optimization of machining parameters to minimize surface roughness using artificial neural network coupled with genetic algorithm. *Procedia Cirp*, *31*, 453-458.

Klinker, F. (2011). Exponential moving average versus moving exponential average. *Mathematische Semesterberichte*, *58*(1), 97-107.

Koike, H., Honda, T., Kida, K., Santos, E. C., Kashima, Y., & Kanemasu, K. (2011). Influence of radial load on PEEK plastic bearings life cycle. *Advanced Materials Research*, *154*, 1288-1291

Konar, P., & Chattopadhyay, P. (2011). Bearing fault detection of induction motor using wavelet and Support Vector Machines (SVMs). *Applied Soft Computing*, *11*(6), 4203-4211.

Kumar, G. S., Natarajan, U., & Ananthan, S. S. (2012). Vision inspection system for the identification and classification of defects in MIG welding joints. *The International Journal of Advanced Manufacturing Technology*, *61*(9-12), 923-933.

Lau, E. C., & Ngan, H. W. (2010). Detection of motor bearing outer raceway defect by wavelet packet transformed motor current signature analysis. *IEEE Transactions on Instrumentation and measurement, 59(*10*)*, 2683-2690.

Li, Y. S. C. T. S., Billington, S., Zhang, C., Kurfess, T., Danyluk, S., & Liang, S. (1999). Adaptive prognostics for rolling element bearing condition. *Mechanical Systems and Signal Processing*, *13*(1), 103-113.

Liang, M., & Bozchalooi, I. S. (2010). An energy operator approach to joint application of amplitude and frequency-demodulations for bearing fault detection. *Mechanical Systems and Signal Processing*, *24(*5*)*, 1473-1494.

Lodewijks, G. (2004). Strategies for automated maintenance of belt conveyor systems. *Bulk Solids Handling*, *24*(1), 16-22.

Nahato, K. B., Harichandran, K. N., & Arputharaj, K. (2015). Knowledge mining from clinical datasets using rough sets and backpropagation neural network. *Computational and Mathematical Methods in Medicine, 2015.*

Nakano, M., Takahashi, A., & Takahashi, S. (2017). Generalized exponential moving average (EMA) model with particle filtering and anomaly detection. *Expert Systems with Applications*, *73*, 187-200.

Negnevitsky, M. (2011). *Artificial intelligence: A guide to intelligent systems* (3rd ed.). London, England: Pearson Education.

Nie, F., Huang, H., Ding, C., Luo, D., & Wang, H. (2011, June). Robust principal component analysis with non-greedy ℓ1-norm maximization. *Proceeding of Twenty-Second International Joint Conference on Artificial Intelligence*.

Ning, F., Cong, W., Qiu, J., Wei, J., & Wang, S. (2015). Additive manufacturing of carbon fiber reinforced thermoplastic composites using fused deposition modeling. *Composites Part B: Engineering*, *80*, 369-378.

Pan, Y., Chen, J., & Guo, L. (2009). Robust bearing performance degradation assessment method based on improved wavelet packet–support vector data description. *Mechanical Systems and Signal Processing, 23(*3*)*, 669-681.

Pan, Y., Chen, J., & Li, X. (2010). Bearing performance degradation assessment based on lifting wavelet packet decomposition & fuzzy c-means. *Mechanical Systems and Signal Processing, 24(2)*, 559-566.

Peter, W. T., & Wang, D. (2013). The design of a new sparsogram for fast bearing fault diagnosis: Part 1 of the two related manuscripts that have a joint title as "Two automatic vibration-based fault diagnostic methods using the novel sparsity measurement–Parts 1 and 2". *Mechanical Systems and Signal Processing*, *40(*2*)*, 499519.

Rafiee, J., Rafiee, M. A., & Tse, P. W. (2010). Application of mother wavelet functions for automatic gear and bearing fault diagnosis. *Expert Systems with Applications*, *37*(6), 4568-4579.

Riahi, K. L., & Stadlmayr, A. (2012). *U.S. Patent No. 8,226,297*. Washington, DC: U.S. Patent and Trademark Office.

Ringnér, M. (2008). What is principal component analysis?. *Nature Biotechnology*, *26*(3), 303.

Sadowsky, J. (1994). The continuous wavelet transform: a tool for signal investigation and understanding. *Johns Hopkins APL Technical Digest*, *15*, 306-306.

Shen, C., Wang, D., Kong, F., & Peter, W. T. (2013). Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier. *Measurement*, *46*(4), 1551-1564.

Shlens, J. (2014). A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*.

Smadi, H. J., & Kamrani, A. K. (2011). Product quality-based methodology for machine failure analysis and prediction. *International Journal of Industrial Engineering*, *18*(11).

Wu, J. D., & Liu, C. H. (2009). An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network. *Expert Systems with Applications*, *36*(3), 4278-4286.

Yang, W., & Court, R. (2013). Experimental study on the optimum time for conducting bearing maintenance. *Measurement*, *46*(8), 2781-2791.

Zahran, O., Kasban, H., El-Kordy, M., & El-Samie, F. A. (2013). Automatic weld defect identification from radiographic images. *Ndt & E International*, *57*, 26-35.

Zain, A. M., Haron, H., & Sharif, S. (2010). Prediction of surface roughness in the end milling machining using Artificial Neural Network. *Expert Systems with Applications*, *37*(2), 1755-1768.

Zarei, J., Tajeddini, M. A., & Karimi, H. R. (2014). Vibration analysis for bearing fault detection and classification using an intelligent filter. *Mechatronics*, *24*(2), 151-157.

Zhang, L., Xiong, G., Liu, H., Zou, H., & Guo, W. (2010). Bearing fault diagnosis using multi-scale entropy and adaptive neuro-fuzzy inference. *Expert Systems with Applications*, *37(*8*)*, 6077-6085.