Computers and Mathematics with Applications 60 (2010) 1078–1087

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa

Predicting remaining useful life of rotating machinery based artificial neural network

Abd Kadir Mahamad [a,b,*], Sharifah Saon [b], Takashi Hiyama [a]

[a] Department of Computer Science and Electrical Engineering, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan
[b] Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Johor, Malaysia

ARTICLE INFO

Keywords:
RUL
ANN
Bearing
Prediction
FFNN

ABSTRACT

Accurate remaining useful life (RUL) prediction of machines is important for condition based maintenance (CBM) to improve the reliability and cost of maintenance. This paper proposes artificial neural network (ANN) as a method to improve accurate RUL prediction of bearing failure. For this purpose, ANN model uses time and fitted measurements Weibull hazard rates of root mean square (RMS) and kurtosis from its present and previous points as input. Meanwhile, the normalized life percentage is selected as output. By doing that, the noise of a degradation signal from a target bearing can be minimized and the accuracy of prognosis system can be improved. The ANN RUL prediction uses FeedForward Neural Network (FFNN) with Levenberg Marquardt of training algorithm. The results from the proposed method shows that better performance is achieved in order to predict bearing failure.

1. Introduction

The manufacturing and industrial sectors of the world are facing an exponentially increasing demand to produce goods at better quality while keeping their operating process at maximum yield. The manufacture of such typical products as textiles, aircraft, automobiles and appliances involve a large number of complex processes and nonlinear dynamic systems. Therefore, these processes are not well understood, and the operation is usually understood using experience rather than through the application of scientific principles. Failures occurring during production operation, results in several negative implications such as increase in downtime, low productivity and sometimes can even cause safety risks. Therefore, a method to detect machinery faults has evolved from preventive maintenance to condition based maintenance (CBM) in order to make sure that the production operation can reach maximum capacity.

CBM is growing in popularity in industries with significant increase in hardware and software. Nowadays, there is notable growth in the variety of forms of CBM techniques for electrical machine monitoring and fault prognosis. However, irrespective of the particular CBM technique used the principle of CBM is the same, condition data needs to be interpreted and appropriate actions should be taken accordingly. Therefore prognosis system is used to predict the RUL time of a machine failure.

The existing prognosis or RUL prediction methods can be classified into 3 categories, which are physics based prognosis models, data driven prognosis models and integration of reliability and prognosis system. The common physics based prognosis is a crack growth modeling, which combines mechanical knowledge, defect growth modeling and CBM data to provide sufficient knowledge of a prognosis output. Li et al. [1,2] proposed a method to estimate RUL of a bearing based on its defect growth. The fatigue crack propagation is then compared to the estimation from the diagnostic model. To validate the

* Corresponding author at: Department of Computer Science and Electrical Engineering, Kumamoto University, 2-39-1 Kurokami, Kumamoto 860-8555, Japan.
E-mail address: kadir@uthm.edu.my (A.K. Mahamad).

0898-1221/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2010.03.065

Nomenclature

C	Connection matrix from the hidden layer to the output layer.
B	Connection matrix from the input layer to the hidden layer.
K	Number of data points.
$F(t)$	Cumulative density function.
H	Hessian matrix.
J	Jacobian matrix.

proposed method, an experimental study with vibration measurement was performed. Finally, Li et al. concluded their work can performed effectively in order to predict the bearing defect process without requiring a prior knowledge of prognosis model parameters. Li and Lee [3] used Paris' law to model gear fatigue crack growth. This model was then tested with the data of other gear tests, where its performance error between actual and predicted RUL was smaller than 7%. The advantage of physics based prognosis is that it requires less data than that of data driven technique, but in real application, this technique is too stochastic and complex for modeling process.

On the other hand, data driven prognosis model does not require assumption of physics parameter, thus it is easy to apply. But this technique needs a large amount of data to make the system as close to real application as possible. ANN is most commonly found as a data driven technique in prognosis system. Normally, ANN consists of input layer, one or several hidden layers and output layer. Many researchers have proposed different types and structures of ANN to overcome their targeted problems. For instance Tian [4], developed ANN to achieve more accurate RUL of a pump bearing by selecting the age and multiple measurement values from condition monitoring as input for ANN. Meanwhile, the life percentage of a pump bearing is used as an output. Vachtsevanos and Wang [5] used dynamic wavelet neural network (DWNN) to predict rolling elements of a bearing failure which was compared to auto regression (AR) model. Satish and Sarma [6] demonstrated the methods which combines the ANN and fuzzy logic as a hybrid system to identify the condition of a bearing at present condition and its RUL. The merit of ANN is that it does not consider the analytical model of damage propagation. It only aims to model damage propagation based on the data collected during CBM.

The conventional data driven model can also be obtained from a simple model such as degradation features model of machine as it was done by Liao et al. [7]. They proposed the proportional hazard model and logistic regression model to predict the RUL of a machine. Meanwhile, Tran et al. [8] presented an approach to predict the condition of a machine which combines the classification and regression trees (CART) with adaptive neuro-fuzzy inference system (ANFIS). These combination methods are then associated with direct prediction technique to determine multi-step ahead prediction of a machine condition.

As a comparison, the integration of reliability and prognosis techniques, utilizes the available information more fully in increasing the accuracy of a prognosis system, which can be used in prognosis for the longer range. This technique requires both event and condition data for modeling process, thus the system becomes more complex. There are several papers which used these techniques as in [9–11]. Jozwiak [9] presented in his work that in order to solve the reliability of the system, the associated variables in the system must be considered. The Cox and Weibull models are studied, in which the method to estimate the parameters in these models is presented. Tian et al. [10] developed a method to find optimum maintenance scheme when the systems have multiple objective conditions. The decision maker from their system can gave good tradeoff between cost and reliability objective function. The reliability analysis of the condition monitoring takes into account the measurement information (vibration and temperature analysis, etc) in order to establish the optimal replacement scheme. This measurement information is considered as a covariate parameter [11], which is an important parameter in determining the RUL of a machine.

In this paper, we propose ANN to achieve accurate RUL of a predicted machine failure. To achieve this objective, the ANN model uses time and fitted measurements Weibull hazard rates of RMS and kurtosis from its present and previous points as input and normalized life percentage as output. By doing that, the noise due to degradation from a target bearing can be minimized and the accuracy of the prognosis can be improved. This research attempts to address the development of prognosis system which can predict the RUL of machine timely to avoid sudden bearing failure.

2. Experimental setup

In industries, bearings are important components and are regularly used. The root causes of bearing failures are normally attributed to improper installations, poor lubrication practices, excessive balance and alignment tolerances, poor storage and handling techniques. Monitoring the above failures is very important for early warning signs before the bearings approach failure stage. This will avoid serious damages which might lead to potentially hazardous situations.

Today, various methods are available to detect and monitor such failures. These include vibration and acoustic emission techniques. But in this research, the vibration will be utilized to acquire initial signal fault on a target bearing [4]. Features such as mean, kurtosis, skewness and RMS are utilized as important features representing fault signals of a bearing [12]. In literature, many methods such as ANN, fuzzy logic, evolutionary algorithm and several other methods have successfully been proven to diagnose bearing failures [13–15]. However, most of the papers discussed how to diagnose bearing failure only in

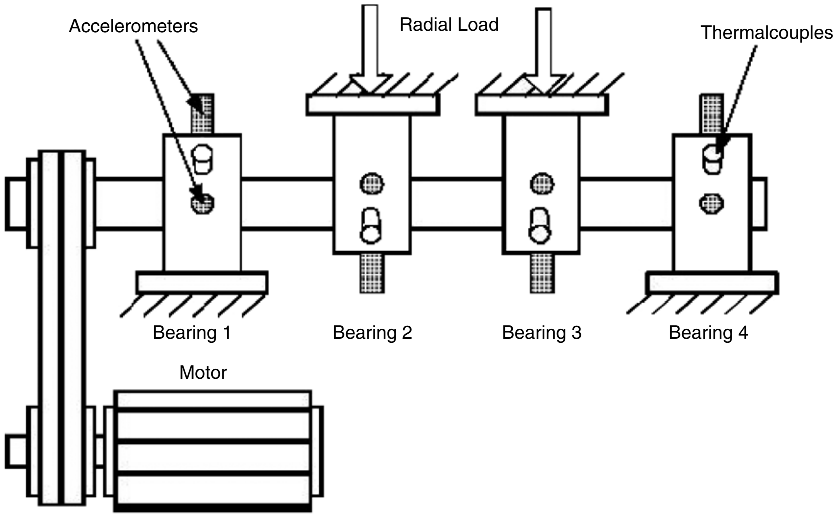

Fig. 1. Bearing test rig.

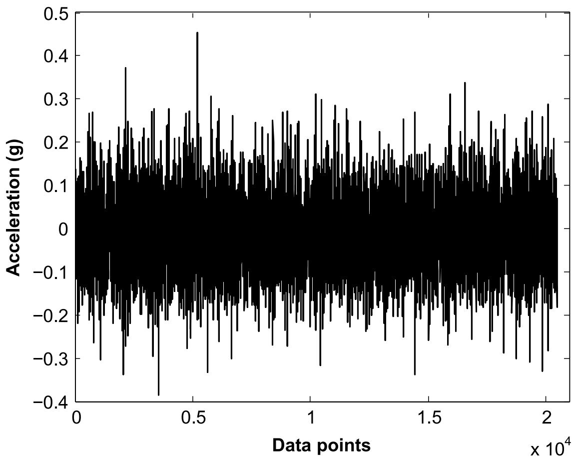

Fig. 2. Vibration signal.

specific failure boundary. In this work, fitted measurement Weibull hazard rate is used instead of the actual measurement to represent the fault signal from a target bearing. The hazard rate (instantaneous failure rate) is an appropriate analytical measure in assessing the reliability of a specific machine or component.

The vibration signals used in this paper were provided by the Center for Intelligent Maintenance Systems (IMS), University of Cincinnati. A schematic of experimental rig is shown as in Fig. 1 [16]. Four Rexnord ZA-2115 double row bearings are installed on one shaft. The shaft is driven by an AC motor and coupled by rub belts. The rotational speed is set at 2000 rpm, and its radial load is 6000 lbs which is applied to the shaft and bearing by a spring mechanism. The bearings have 16 rollers in each row, a pitch diameter of 2.815 inch, roller diameter of 0.331 inch and a tapered contact angle of 15.17°. All the bearings are force lubricated by using an oil circulation system. On each bearing, two accelerometers are installed in which one is at vertical Y and one at horizontal X. The vibration signal from each bearing is collected for one second every 10 min with the sampling rate of 20 kHz and the data length of 20,480 points as shown in Fig. 2. Data collection is done using a National Instruments LabVIEW program. It takes a total of 7 days until the bearing fail. At the end of the test, an outer race failure occurs on Bearing 1.

From the vibration signal, we can extract the measurement values or features which represent the degradation of the bearing. Two measurement values that commonly used in order to detect bearing failure [7], which are RMS and kurtosis can be defined as in (1) and (2),

$$RMS = \sqrt{\frac{\sum_{k=1}^{K}(x(k))^2}{K}} \tag{1}$$

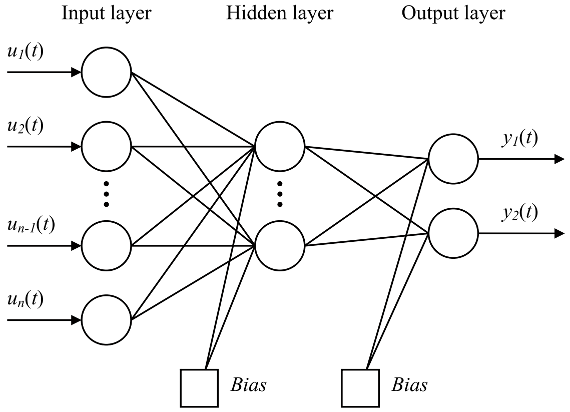

Fig. 3. Block diagram of FFNN.

$$kurtosis = \frac{\sum_{k=1}^{K}(x(k) - x_m)^4}{(K-1)x_{std}^4} \tag{2}$$

where $x(k)$ is a signal series for $k = 1, 2, \ldots, K$, and K is the number of data points. Standard deviation, x_{std} and mean value, x_m of the signal can be represented as in (3) and (4),

$$x_{std} = \sqrt{\frac{\sum_{k=1}^{K}(x(k) - x_m)^2}{K-1}} \tag{3}$$

$$x_m = \frac{1}{K}\sum_{k=1}^{K} x(k). \tag{4}$$

3. Review of FFNN

Three layer (sometimes called two layer) FFNN are commonly encountered models found in many scientific papers [17]. Fig. 3 shows configuration of FFNN, where the network is divided into 3 layers; input, hidden and output layers. Lines represents weighted connections and the bias thresholding nodes are represented by squares.

Mathematically, the typical FFNN can be expressed as

$$y_i = \varphi_o[C\varphi_h(Bu_i + b_h) + b_o] \tag{5}$$

where y_i is the output vector corresponding to input vector u_i, C is the connection matrix of weights between two nodes from the hidden layer to the output layer and B is the connection matrix from the input layer to the hidden layer. Meanwhile b_h and b_o are the bias vector for the hidden and output layer. $\varphi_h(\cdot)$ and $\varphi_o(\cdot)$ are the vector valued functions, which corresponds to the activation (transfer) functions of the nodes in the hidden and output layers, respectively. In Matlab, the transfer function can be logsig (log-sigmoid), tansig (tangent-sigmoid) and purelin (linear). FFNN models also have the general structure of

$$y_i = f(u) \tag{6}$$

where $f(\cdot)$ is a nonlinear mapping. FFNN is structurally similar to nonlinear regression models.

To introduce FFNN for identification of dynamic systems or prediction of time series, a vector comprising of a moving window of past input values (delayed coordinates) must be used as inputs to the network. This procedure yields a model analogous to a nonlinear finite impulse response model where

$$y_i = y_t \quad \textbf{and} \quad u_i = [u_t, u_{t-1}, \ldots, u_{t-m}] \quad \textbf{or} \quad y_t = f([u_t, u_{t-1}, \ldots, u_{t-m}]). \tag{7}$$

The lengths of the moving window must be long enough to support the system dynamics for each variable in practice. The duration of the data windows are determined by trial and error, and each individual input and output variable might have a separate data window for optimal performance.

Levenberg Marquardt (LM) learning algorithm is one of the earliest and the most common method used as a training algorithm for FFNN. LM algorithm is used to train nonlinear, multilayered networks (FFNN) to successfully solve many difficult and diverse problems.

Table 1
Results of parameter estimates.

Weibull hazard rate	Parameter estimates	
RMS	$\gamma_1 = 0.4077$	$\eta_1 = 1.2017$
Kurtosis	$\gamma_2 = 0.4360$	$\eta_2 = 1.2970$

4. The proposed ANN RUL prediction method

The RUL of a bearing is a nonlinear function. To predict it, we need the powerful tool which can determine the mapping relationship between the input data obtained from the bearing and the RUL of a bearing. To achieve this, the ANN is proposed; as it is a very powerful tool that can determine the nonlinear function of the system. The proposed ANN model uses the fitted measurement values as inputs instead of real measurement values. The reason is that, the measurement data taken from real application usually has external noise that can affect the measurement data [4]. Here, we propose the fitted measurements data as the input to represent the deterioration features of the bearing. The fitted measurement use the Weibull hazard rate function which is a very powerful function to represent the reliability of machine failure. The equation of Weibull hazard rate is as follows

$$h(t) = \frac{f(t)}{1 - F(t)} \tag{8}$$

where $f(t)$ is the probability density function and

$$F(t) = 1 - \exp\left[-\left(\frac{t}{\gamma}\right)^{\eta}\right] \quad t \geq 0 \tag{9}$$

is a cumulative density function, with γ and η being scale and shape parameters respectively. Table 1 shows the results of parameter estimates of Weibull hazard rate function of RMS and kurtosis values. Weibull hazard rate for RMS and kurtosis during current and previous inspection can be defined as follow;
Weibull hazard rate for RMS
Current inspection;

$$z_i^1 = \frac{f_i^1(t)}{1 - F_i^1(t)}. \tag{10}$$

Previous inspection;

$$z_{i-1}^1 = \frac{f_{i-1}^1(t)}{1 - F_{i-1}^1(t)}. \tag{11}$$

Weibull hazard rate for kurtosis
Current inspection;

$$z_i^2 = \frac{f_i^2(t)}{1 - F_i^2(t)}. \tag{12}$$

Previous inspection;

$$z_{i-1}^2 = \frac{f_{i-1}^2(t)}{1 - F_{i-1}^2(t)}. \tag{13}$$

For ANN training, there are 6 inputs fed into the network. Input, t_i and t_{i-1} are the time values at the current and previous inspection respectively. z_i^1 and z_{i-1}^1 are RMS values of fitted measurements at the current and previous inspection while z_i^2 and z_{i-1}^2 are kurtosis values of fitted measurements at the current and previous inspection respectively. In this work we take into account the time, RMS and kurtosis value. The time value at present and previous value is important for ANN in estimating the RUL of a bearing. Two other fitted measurements at present and previous are useful in representing the bearing's condition [7].

For output of ANN, the life percentage (normalized) is preferred and is denoted as T_i. The life percentage (normalized) is the best option in mapping the bearing's health condition, which is proportional to time. This means that the bearing is totally damaged when it reaches 100% of the life percentage. The structure of the proposed ANN model is shown in Fig. 4. While, the measurements value of RMS and kurtosis are depicted in Fig. 5(a) and (b).

Training neural networks is a data-analytic procedure. Under this condition, it is necessary to stop training once an overfit is indicated. The overfitting problem is defined as the model or system which gives good performance during training process, but when it is tested with unseen data, the model gives worse performance. One way to overcome this problem is by using cross-validation. Two different sets of data are required for training and validating the network. During overfit

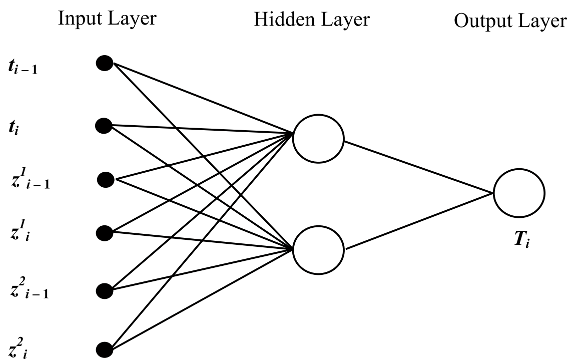

Fig. 4. Structure of the proposed ANN model.

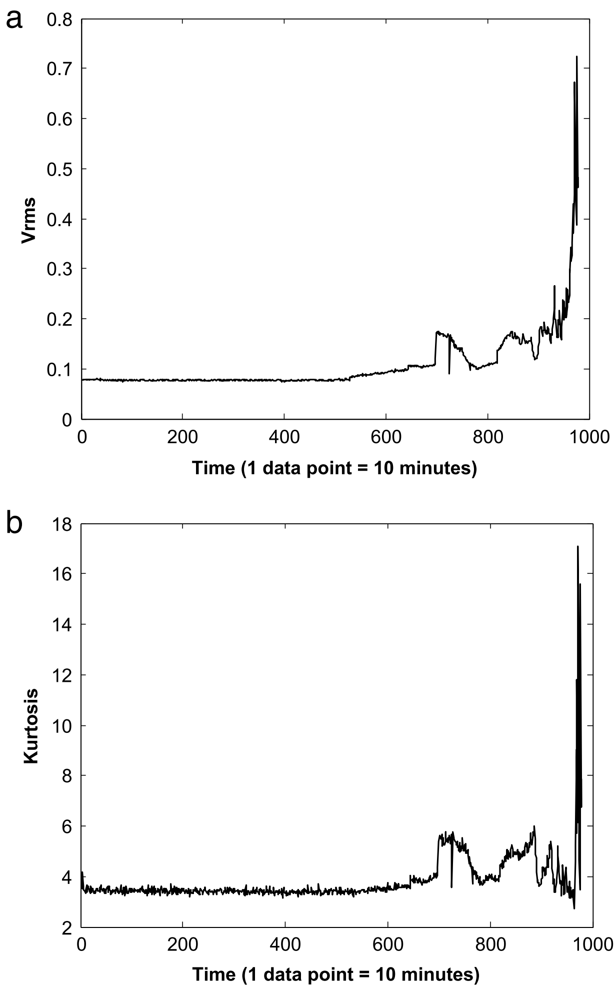

Fig. 5. (a) RMS and (b) kurtosis value.

situation, mean square error (MSE) for the validation set decreases first, but then comes to minimum and later increases though the MSE of the training set continues to decrease. When the MSE of the validation set increases, it is assumed that the regression algorithm is overfitting the training data. Thus, the training is stopped as soon as MSE over the validation set begins to increase. For the selection of FFNN topology, there is no specific method. Trial and error search method is the best option to select the optimum topology for the prediction.

In this work, we divide the data into two sets, the training and a validation sets in order to overcome the above problem. The training set uses the original data set from input but the validation set is perturbed with +5% of the feed. Furthermore, the actual FFNN output is normalized between 0 and 1 in order to get the same order of magnitude variables and to avoid numerical instability problems. The ANN is trained and validated in order to find the minimum validation error. The training and validation for ANN are setup from two to twenty hidden nodes. The network which produces a minimum validation error will be selected as the optimum one. The reason is that, the network can still give a good performance even when the input is perturbed within a certain feed.

The configuration of ANN model uses logsig (log-sigmoid) transfer function in its hidden layer and purelin (linear) transfer function in its output layer. With this combination, the network can approximate to any function. Meanwhile, the ANN network training technique for optimization, uses Levenberg Marquardt (trainlm) algorithm because it gives better performance among other network training algorithms.

This algorithm is designed for minimizing functions that are sums of squares of nonlinear functions and it has the ability to approach second-order convergence point without need to calculate the Hessian matrix. If the performance function reaches a sum of squares, then the approximation of Hessian matrix can be expressed as;

$$H = J^T J \tag{14}$$

and the gradient, $g = J^T e$ can be computed, where J is the Jacobian matrix, which contains first derivatives of network errors with respect to the weights and biases. This algorithm can be computed using a standard backpropagation technique that is easier than calculating using the Hessian matrix. While e is a vector of network errors.

The Levenberg–Marquardt algorithm uses the Hessian matrix approximation in the Newton-like update as follows;

$$x_{k+1} = x_k \left\{ - \left[J^T J + \mu I \right]^{-1} \right\} J^T e. \tag{15}$$

In case where scalar μ is zero, this update will resemble Newton's method using the approximate Hessian matrix. However, it becomes gradient descent method with a small step size, if μ is large. Newton's method is faster and more accurate near an error minimum; therefore, the main target is to shift toward Newton's method as quickly as possible. Therefore, during reduction of performance for every update step, μ is decreased and vice versa. As a conclusion, the performance function is always reduced at each iteration of the algorithm [18].

MSE is used to present the network performance in order to define the best network [17]. The equation is stated as;

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (e_i)^2 = \frac{1}{N} \sum_{i=1}^{N} (t_i - a_i)^2 \tag{16}$$

where

e_i = Error a_i = Actual value

t_i = Desired value N = Number of data.

The overall procedure of this proposed method can be illustrated in a flow chart as in Fig. 6 and explained as follows;

1. The input data used in this work contains time and measurement values (RMS and kurtosis).
2. Each measurement value is fitted with Weibull hazard rate function. The time and fitted measurements are used as training data set for ANN.
3. To validate the performance of ANN, we construct validation data set which is perturbed with +5% from the training data set. The validation data set is also used to avoid the overfitting problem during training process.
4. ANN is trained based on the training and validation data set, in which the Levenberg Marquardt algorithm is used as training algorithm.
5. After the ANN training is done, and when the minimum MSE error is met, the proposed network is used to predict the percentage of the bearing's life.
6. If the new measurement data (RMS and kurtosis) is available, it needs to be fitted with Weibull hazard rate function, before it is fed into a proposed ANN.

5. Result

The proposed networks are trained and validated to indicate their performance. The results for both training and validation are shown as in Table 2. The training error for this network is $9.99e^{-13}$ and the validation error is $3.1e^{-12}$. As it can be seen, the proposed network gave minimum error for both training and validation processes with 2 hidden nodes. The training and validation performance can be seen as in Fig. 7(a) and (b). The '+' sign indicates the actual output and the '•' sign indicates the predicted output. The x axis indicates the RUL of a bearing and y axis indicates the life percentage of a bearing in a normalized form. For instance, if we want to know the life percentage after the bearing is ran for 6000 min, the life percentage shows about 60%. The RUL of bearing would be around 40%. From Fig. 7, both (a) training and (b) validation performances show very good performance, in which the actual and predicted outputs are almost same.

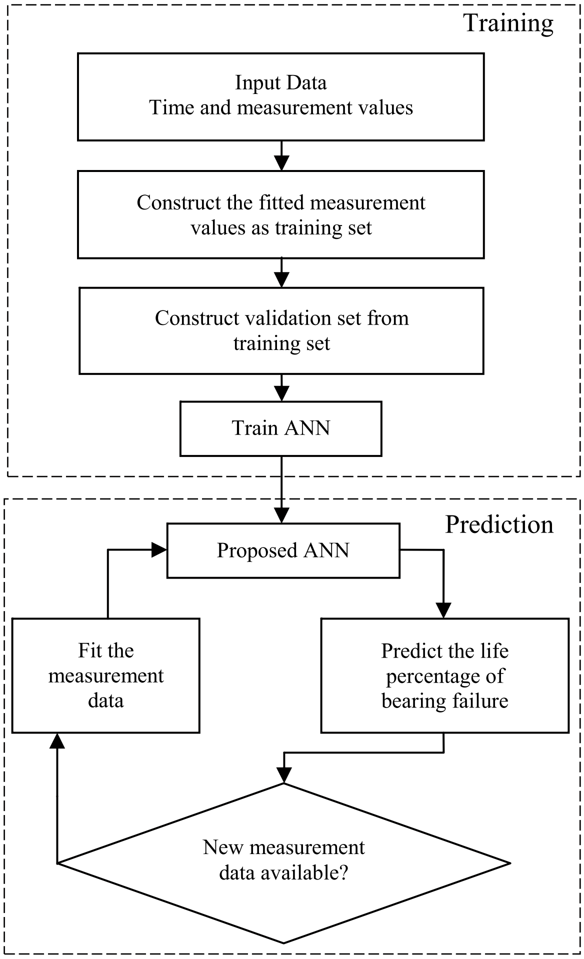

Fig. 6. Flow chart of the proposed method.

Table 2
Training and validation error of proposed model.

Network model	Hidden node	Training error	Validation error
FNNN	2	9.99E−13	3.10E−12

In order to confirm that our proposed method is the optimum network, we tested it with certain unseen data. In this process, we generated 100 data points randomly from the training set and perturbed it with +10% of feed. This data was then used as input for our FFNN model. The output performance of testing process can be seen as in Fig. 8. Observation from this figure shows that the actual and predicted outputs are same for every single output. This performance obviously shows that the proposed FFNN model, which uses time and hazard rates of RMS and kurtosis from present and previous points as input and normalized life percentage as output, is suitable to use as an optimum ANN model in predicting bearing failures.

As a discussion, this proposed ANN method predicts the life percentage of bearing failure, in which we do not consider the failure threshold as many researchers suggested in their papers. The bearing will totally fail after it reaches 100% of its life percentage.

6. Conclusions

The accurate RUL of a machine is important to CBM in order to improve the reliability and cost of maintenance. This paper proposes ANN in achieving more accurate estimate RUL of a bearing failure. In this case, the ANN model uses time and Weibull hazard rates of RMS and kurtosis from present and previous points as input for ANN. Furthermore, the normalized life percentage is selected as output. By doing this, the noise of degradation from a target bearing can be minimized and

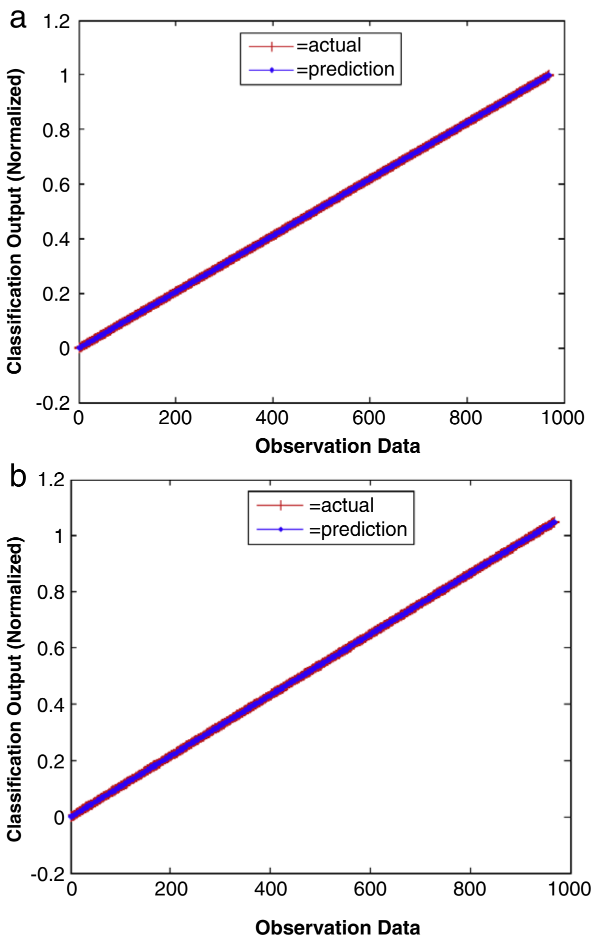

Fig. 7. Result of (a) training and (b) validation.

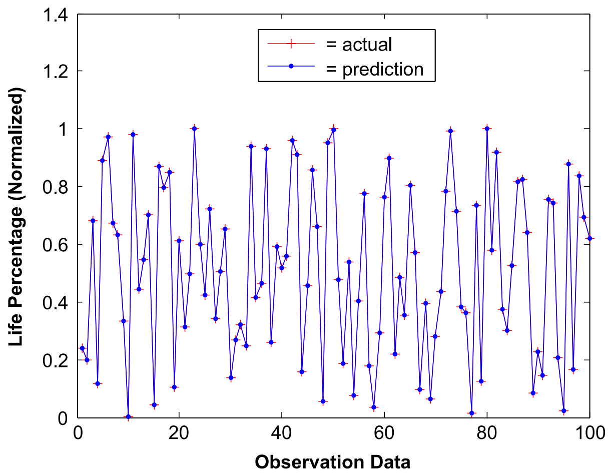

Fig. 8. The output performance of testing process.

the accuracy of prognosis can be improved. From the results, it shows that the proposed ANN gave good performance in predicting RUL of a bearing failure. In this paper we did not take into account the failure threshold, which many other researchers proposed. The bearing will fail after it reaches 100% of its life percentage.

References

[1] Y. Li, S. Billington, C. Zhang, T. Kurfess, S. Danyluk, S. Liang, Adaptive prognostics for rolling element bearing condition, Mechanical Systems and Signal Processing 13 (1999) 103–113.
[2] Y. Li, T.R. Kurfess, S.Y. Liang, Stochastic prognostics for rolling element bearings, Mechanical Systems and Signal Processing 14 (2000) 747–762.
[3] C.J. Li, H. Lee, Gear fatigue crack prognosis using embedded model, gear dynamic model and fracture mechanics, Mechanical Systems and Signal Processing (2005) 836–846.
[4] Z. Tian, An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring, Journal of Intelligent Manufacturing (2009).
[5] G. Vachtsevanos, P. Wang, Fault prognosis using dynamic wavelet neural networks, in: AUTOTESTCON Proceedings, IEEE Systems Readiness Technology Conference, 2001, pp. 857–870.
[6] B. Satish, N.D.R. Sarma, A fuzzy BP approach for diagnosis and prognosis of bearing faults in induction motors, in: IEEE Power Engineering Society General Meeting, vol. 3, 2005, pp. 2291–2294.
[7] H. Liao, W. Zhao, H. Guo, Predicting remaining useful life of an individual unit using proportional Hazards model and logistic regression model, in: Reliability and Maintainability Symposium, 2006, pp. 127–132.
[8] V.T. Tran, B.S. Yang, A.C.C. Tan, Multi-step ahead direct prediction for the machine condition prognosis using regression trees and neuro-fuzzy systems, Expert Systems with Applications 36 (2009) 9378–9387.
[9] I.J. Jóźwiak, An introduction to the studies of reliability of systems using the Weibull proportional hazards model, Microelectronics and Reliability 37 (1997) 915–918.
[10] Z. Tian, D. Lin, B. Wu, Condition based maintenance optimization considering multiple objectives, Journal of Intelligent Manufacturing (2009).
[11] D. Banjevic, A.K.S. Jardine, Calculation of reliability function and remaining useful life for a Markov failure time process, IMA Journal of Management Mathematics (2006) 115–130.
[12] R.B.W. Heng, M.J.M. Nor, Statistical analysis of sound and vibration signals for monitoring rolling element bearing condition, Applied Acoustics (1998) 211–226.
[13] Y. Lei, Z. He, Y. Zi, Q. Hu, Fault diagnosis of rotating machinery based on multiple ANFIS combination with GAs, Mechanical Systems and Signal Processing (2007) 2280–2294.
[14] J.M.F. Salido, S. Murakami, A comparison of two learning mechanisms for the automatic design of fuzzy diagnosis systems for rotating machinery, Applied Soft Computing (2004) 413–422.
[15] S. Rajakarunakaran, P. Venkumar, D. Devaraj, K.S.P. Rao, Artificial neural network approach for fault detection in rotary system, Applied Soft Computing (2008) 740–748.
[16] J. Lee, H. Qiu, G. Yu, J. Lin, Bearing data set, IMS, University of Cincinnati, NASA Ames Prognostics Data Repository, Rexnord Technical Services, 2007.
[17] A. Ahmad, M. Kamaruddin, Detection of multiple sensor faults in a palm oil fractionation plant using artificial neural network, in: Regional Symposium on Chemical Engineering, 2002.
[18] D. Howard, B. Mark, H. Martin, MATLAB Neural Network Toolbox 6, User's Guide.