Bearing Condition Prediction Using Enhanced Online Learning Fuzzy Neural Networks

Yongping Pan[1], Xiangyu Hu[1], Meng Joo Er[1], Xiang Li[2], and Rafael Gouriveau[3]

[1] School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798

[2] Singapore Institute of Manufacturing Technology, Singapore 638075

[3] FEMTO-ST Institute, 24 rue Alain Savary, Besancon 25000, France

Abstract

Machine health condition (MHC) prediction is useful for preventing unexpected failures and minimizing overall maintenance costs since it provides decision-making information for condition-based maintenance (CBM). This paper presents a novel bearing health condition prediction approach based on enhanced online sequential learning fuzzy neural networks (EOSL-FNNs). Based on extreme learning machine (ELM) theory, an online sequential learning strategy is developed to train the FNN. Taking advantage of the proposed learning strategy, a multi-step time-series direct prediction scheme is presented to forecast bearing health condition online. The proposed approach not only keeps all salient features of the ELM, including extremely fast learning speed, good generalization ability and elimination of tedious parameter design, but also solves the singular and ill-posed problems caused by the situation that the number of training data is smaller than the number of hidden nodes. Simulation studies using real-world data from the accelerated bearing life have demonstrated the effectiveness and superiority of the proposed approach.

Keywords:

Machine health condition (MHC); Fuzzy neural network (FNN); Time-series forecast; Prognosis; Online learning

1 INTRODUCTION

Machine health condition (MHC) prediction is useful for preventing unexpected failures and minimizing overall maintenance costs since it provides decision-making information for condition-based maintenance (CBM) [1]. Typically, MHC prediction methods can be divided into two categories, namely model-based data-driven methods [2]. Due to the difficulty of deriving an accurate fault propagation model [3], [4], researches have focused more on the data-driven method in recent years [5]. The neural network (NN)-based approach, which falls under the category of the data-driven method, have been considered to be very promising for MHC prediction due to the adaptability, nonlinearity, and universal function approximation capability of NNs [6]. Batch learning and sequential learning are two major training schemes of NNs. MHC prediction is essentially an online time-series forecasting problem which should perform real-time prediction while updating the NN. Thus, to save updating time and to maintain consistency of the NN, the sequential learning should be employed in such a problem.

The most popular NNs applied to MHC prediction are recurrent NNs (RNNs) and fuzzy NNs (FNNs). In [6], an extended RNN which contains both Elman and Jordan context layers was developed for gearbox health condition prediction. In [7], a FNN in [8] was applied to predict bearing health condition. In [9], an enhanced FNN was developed to forecast MHC. Next, in [10] and [11], a recurrent counterpart of the approach in [9] and a multi-step counterpart of the approach in [10] were presented to predict MHC, respectively. An interval type-2 FNN was also proposed to predict bearing health condition under noisy uncertainties in [12]. Note that the batch learning was employed in [6], [7], [12]. Common conclusions from [6], [7], [9]-[12] are that the RNN usually outperforms the feedforward NN, and the FNN usually outperforms the feedforward perceptron NN, feedforward radial-basis-function (RBF) NN, and RNN. Recently, to improve prediction performance under measurement noise, an integrated FNN and Bayesian estimation approach was proposed for predicting MHC in [13], where a FNN is employed to model fault propagation dynamics offline, and a first-order particle filter is utilized to update the confidence values of the MHC estimations online. In [14], a high-order particle filter was applied to the same framework of [13]. A question in the approaches of [13], [14] is that the FNNs should be trained by the system state data (rather than the output data) which are assumed to be immeasurable.

Extreme learning machine (ELM) is an emergent technique for training feedforward NNs with almost any type of nonlinear piecewise continuous hidden nodes [15]. The salient features of ELM are as follows [15]: *i)* All hidden node parameters of NNs are randomly generated without the knowledge of the training data; *ii)* it can be learned without iterative tuning, which implies that the hidden node parameters are fixed after generation and only output weight parameters need to be turned; *iii)* both training errors and weight parameters need to be minimized so that the generalization ability of NNs can be improved; *iv)* its learning speed is extremely fast for all types of learning schemes. ELM demonstrates great potential for MHC prediction due to these salient features. Nonetheless, the original ELM proposed in [15] is not appropriate for predicting MHC since it belongs to the batch learning scheme. To enhance the efficiency of ELM, online sequential ELM (OS-ELM) was developed in [16], and was further applied to train the FNN in [17]. Due to its extremely high learning speed, the OS-ELM-based FNN in [17] seems to be suitable for MHC prediction. Yet, there are two drawbacks in [17] as follows: *i)* It is not good to yield generalization models since only tracking errors are minimized; *ii)* it may encounter singular and ill-posed problems while the number of training data is smaller than the number of hidden notes.

To further improve the efficiency of MHC prediction, a novel FNN with an enhanced sequential learning strategy is proposed in this paper. The design procedure of the proposed approach is as follows: First, a ellipsoidal basic functions (EBFs) FNN is proposed; secondly, the FNN approximation problem is transformed into the bi-objective optimization problem; thirdly, an enhanced online sequential learning strategy based on the ELM is developed to train the FNN; finally, a multi-step direct prediction scheme based on the proposed learning strategy is presented for MHC prediction. The developed enhanced online sequential learning FNN (EOSL-FNN) is applied to predict bearing health condition by the use of real-world data from accelerated bearing life. Comparisons with other NN-based methods are carried out to show the effectiveness and superiority of the proposed approach.

The structures of the rest paper are as follows. The architecture of the FNN is described in Section II. The enhanced online sequential learning strategy based on the ELM is developed in Section III. The multi-step direct prediction scheme is given in Section IV. Simulation results based on real-world bearing data are provided in Section V. Conclusions are given in Section VI.

2 ARCHITECTURE OF FUZZY NEURAL NETWORK

For MHC prediction, we consider the *n*-input single-output system. Yet, the following results can be directly extended to the multi-input multi-output (MIMO) system. The FNN is built based on an EBF NN. It is functionally equivalent to a Takagi-Sugeno-Kang (TSK) fuzzy model that is described by the following fuzzy rules [18]:

$$\text{Rule } R^j : \text{IF } x_1 \textit{ is } A_{1j} \text{ and} \cdots \text{and } x_n \text{ is } A_{nj} \text{ THEN } \hat{y} \text{ is } w_j \tag{1}$$

where $x_i \in \mathbb{R}$ and $\hat{y} \in \mathbb{R}$ are the input variable and output variable, respectively, A_{ij} is the antecedent (linguistic variable) of the *i*th input variable in the *j*th fuzzy rule, w_j is the consequent (numerical variable) of the *j*th fuzzy rule, *i* = 1, 2, …, *n*, *j* = 1, 2, …, *L*, and *L* is the number of fuzzy rules.

As illustrated in Figure 1, there are in total four layers in the FNN. In Layer 1, each node is an input variable x_i and directly transmits its value to the next layer. In Layer 2, each node represents a Gaussian membership function (MF) of the corresponding A_{ij} as follows:

$$\mu_{A_{ij}}(x_i \mid c_{ij}, \sigma_{ij}) = \exp\left[-(x_i - c_{ij})^2 / 2\sigma_{ij}^2\right] \tag{2}$$

where $c_{ij} \in \mathbb{R}$ and $\sigma_{ij} \in \mathbb{R}^+$ are the center and width of the *i*th MF in the *j*th fuzzy rule, respectively. Note that the MF in (2) is an EBF since all its widths σ_{ij} are different [18]. In Layer 3, each node is an EBF unit that denotes a possible IF-part of the fuzzy rule. The output of the *j*th node is as follows:

$$\phi_j(\mathrm{x} \mid \mathrm{c}_j, \sigma_j) = \exp\left[-\sum\nolimits_{i=1}^{n}(x_i - c_{ij})^2 / \sigma_{ij}^2\right] \tag{3}$$

where $\mathrm{x} = [x_1, x_2, \cdots, x_n]^T \in \mathbb{R}^n$, $\mathrm{c}_j = [c_{1j}, c_{2j}, \cdots, c_{nj}] \in \mathbb{R}^n$, and $\sigma_j = [\sigma_{1j}, \sigma_{2j}, \cdots, \sigma_{nj}] \in \mathbb{R}^n$. In the last layer, the output $\hat{y}$ is obtained by the weighted summation of ϕ_j as follows:

$$\hat{y} = \hat{f}(\mathrm{x} \mid W, \mathrm{c}, \sigma) = \Phi(\mathrm{x} \mid \mathrm{c}, \sigma)W \tag{4}$$

where $\hat{f}(\cdot): \mathbb{R}^{n+L(1+2n)} \mapsto \mathbb{R}$, $\Phi = [\phi_1, \phi_2, \cdots, \phi_L] \in \mathbb{R}^L$, $\mathrm{c} = [\mathrm{c}_1, \mathrm{c}_2, \cdots, \mathrm{c}_L]^T \in \mathbb{R}^{L \cdot n}$, $\sigma = [\sigma_1, \sigma_2, \cdots, \sigma_L]^T \in \mathbb{R}^{L \cdot n}$, and $W = [w_1, w_2, \cdots, w_L]^T \in \mathbb{R}^L$.

For the TSK model, the THEN-part w_j is a polynomial of x_i which can be expressed as follows:

$$w_j = \alpha_{0j} + \alpha_{1j}x_1 + \cdots + \alpha_{nj}x_n \tag{5}$$

where $\alpha_{0j}, \alpha_{1j}, \cdots, \alpha_{nj} \in \mathbb{R}$ are weights of input variables in the *j*th fuzzy rule. The following lemma shows the universal function approximation property of the proposed FNN.

Lemma 1 [19]: For any given continuous function $f(\mathbf{x}) : \mathrm{D} \mapsto \mathbb{R}$ and arbitrary small constant $\varepsilon \in \mathbb{R}^+$, there exists a FNN in (4) with proper parameters W, c and σ such that

$$\sup\nolimits_{\mathbf{x} \in \mathrm{D}} \left| f(\mathbf{x}) - \hat{f}(\mathbf{x} \mid W, \mathrm{c}, \sigma) \right| < \varepsilon \tag{6}$$

where $\mathrm{D} \subset \mathbb{R}^n$ is an approximation region.

3 ONLINE SEQUENTIAL LEARNING STRATEGY

For training FNNs, consider a data set with *N* arbitrary distinct training samples: $\mathrm{N}_N = \{(\mathrm{x}_l, y_l)\}_{l=1}^N$, where $\mathbf{x}_l = [x_{l1}, x_{l2}, \cdots, x_{ln}]^T \in \mathbb{R}^n$, $y_l \in \mathbb{R}$, and *l* is the number of the sampling point. If a FNN with *L* hidden nodes can approximate these *N* samples with zero error, then there exist proper parameters W, c and σ such that

$$\Phi(\mathrm{x}_l \mid \mathrm{c}, \sigma)W = y_l \tag{7}$$

for all *l* = 1, 2, …, *N*. Since w_j in (5) can be rewritten into $w_j = \mathrm{x}_{le}^T\alpha_j$ with $\mathrm{x}_{le} = [1, \mathrm{x}_l^T]^T \in \mathbb{R}^{n+1}$ and $\alpha_j = [\alpha_{0j}, \alpha_{1j}, \cdots, \alpha_{nj}]^T \in \mathbb{R}^{n+1}$, one gets

$$W = [\mathrm{x}_{le}^T\alpha_1, \mathrm{x}_{le}^T\alpha_2, \cdots, \mathrm{x}_{le}^T\alpha_L]^T. \tag{8}$$

Substituting (8) into (7) for all *l* = 1, 2, …, *N*, applying the definition of Φ and making some manipulations, one gets

$$\begin{bmatrix} \mathrm{x}_{1e}^T(\phi_1\alpha_1 + \phi_2\alpha_2 + \cdots + \phi_L\alpha_L) \\ \mathrm{x}_{2e}^T(\phi_1\alpha_1 + \phi_2\alpha_2 + \cdots + \phi_L\alpha_L) \\ \vdots \\ \mathrm{x}_{Ne}^T(\phi_1\alpha_1 + \phi_2\alpha_2 + \cdots + \phi_L\alpha_L) \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}.$$

From the above expression, it is easy to show that

$$\begin{bmatrix} \mathrm{x}_{1e}^T\phi_1, \mathrm{x}_{1e}^T\phi_2, \cdots, \mathrm{x}_{1e}^T\phi_L \\ \mathrm{x}_{2e}^T\phi_1, \mathrm{x}_{2e}^T\phi_2, \cdots, \mathrm{x}_{2e}^T\phi_L \\ \vdots \\ \mathrm{x}_{Ne}^T\phi_1, \mathrm{x}_{Ne}^T\phi_2, \cdots, \mathrm{x}_{Ne}^T\phi_L \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_L \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

which can be written into the following compact form:

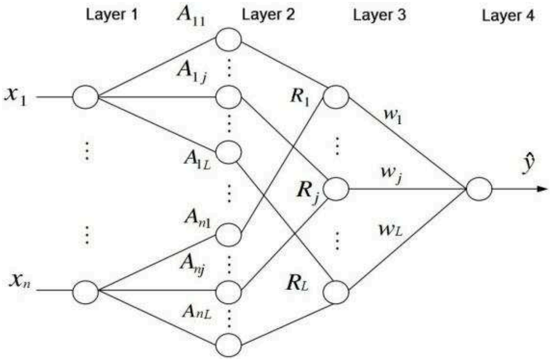


Figure 1: Architecture of fuzzy neural network.

$$H(\mathrm{X}, \mathrm{c}, \sigma)Q = Y \tag{9}$$

where $\mathrm{X} = [\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_N]^T \in \square^{N\times n}$, $Y = [y_1, y_2, \cdots, y_N]^T \in \square^{N\times 1}$, $Q = [\alpha_1^T, \alpha_2^T, \cdots, \alpha_L^T]^T \in \square^{(n+1)L\times 1}$ is the consequent parameter matrix, and $H \in \square^{N\times(n+1)L}$ is the hidden matrix weighted by the fired strength of fuzzy rules given by

$$H(\mathrm{X}, \mathrm{c}, \sigma) = \begin{bmatrix} \mathrm{x}_{1e}^T\phi_1(\mathrm{x}_1, \mathrm{c}_1, \sigma_1), \cdots, \mathrm{x}_{1e}^T\phi_L(\mathrm{x}_1, \mathrm{c}_L, \sigma_L) \\ \mathrm{x}_{2e}^T\phi_1(\mathrm{x}_2, \mathrm{c}_1, \sigma_1), \cdots, \mathrm{x}_{2e}^T\phi_L(\mathrm{x}_2, \mathrm{c}_L, \sigma_L) \\ \vdots \\ \mathrm{x}_{Ne}^T\phi_1(\mathrm{x}_N, \mathrm{c}_1, \sigma_1), \cdots, \mathrm{x}_{Ne}^T\phi_L(\mathrm{x}_N, \mathrm{c}_L, \sigma_L) \end{bmatrix}. \tag{10}$$

From ELM theory, the parameters c and σ in (10) can be randomly generated and fixed after generation, i.e. the updating of antecedent parameters is not necessary. Usually, the equality in (9) cannot be obtained due to the limitation of FNN scale. Consider the following minimizing problem:

$$\min_Q \left(\| HQ - Y \|^2 + \lambda \| Q \|^2 \right) \tag{11}$$

where $\| \cdot \|$ denotes the Euclidean norm, and λ is a real positive constant. The least-squares solution of *Q* in (11) is as follows:

$$\hat{Q} = (H^T H + \lambda I)^{-1} H^T Y. \tag{12}$$

Now, give an initial data set: $\mathrm{N}_0 = \{(\mathrm{x}_l, y_l)\}_{l=1}^{N_0}$. From (12), one immediately gets

$$\hat{Q}_0 = K_0^{-1} H_0^T Y_0 \tag{13}$$

$$K_0 = H_0^T H_0 + \lambda I \tag{14}$$

where $Y_0 = [y_1, y_2, \cdots, y_{N_0}]^T$, $H_0 = H(\mathrm{X}_0, \mathrm{c}, \sigma)$ and $\mathrm{X}_0 = [\mathrm{x}_1, \mathrm{x}_2, \cdots, \mathrm{x}_{N_0}]^T$. Let $\hat{y}_l$ be the estimation of y_l with $l = 1, 2, \cdots$. The FNN output at the initial phase is as follows:

$$\hat{Y}_0 = H_0 \hat{Q}_0 \tag{15}$$

where $\hat{Y}_0 = [\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_{N_0}]^T$.

Then, present the (*k*+1)th chuck of new observations: $\mathrm{N}_{k+1} = \{(\mathbf{x}_l, y_l)\}$ with $l = \sum_{j=0}^{k} N_j + 1, \sum_{j=0}^{k} N_j + 2, \cdots, \sum_{j=0}^{k+1} N_j$, where N_j denotes the number of observations in the (*k*+1)th chunk. From [16], one obtains the RLS solution for *Q* in (11) as follows:

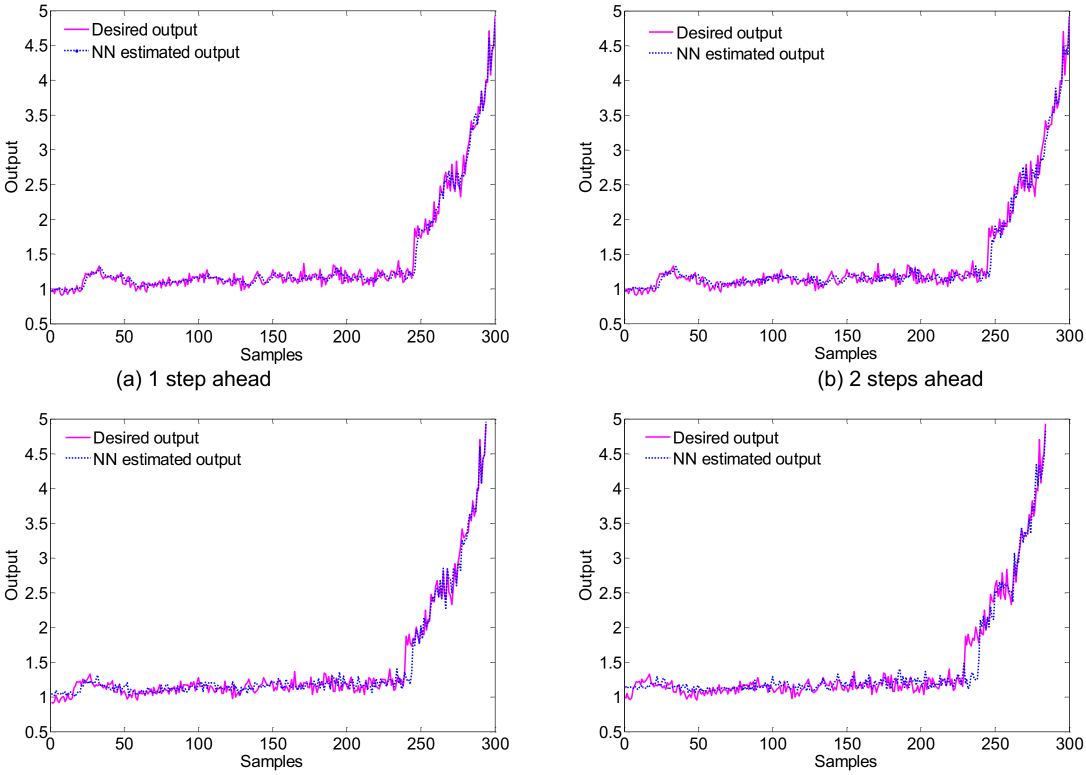
$$K_{k+1} = K_k + H_{k+1}^T H_{k+1} \tag{16}$$

$$\hat{Q}_{k+1} = \hat{Q}_k + K_{k+1}^{-1} H_{k+1}^T (Y_{k+1} - H_{k+1} \hat{Q}_k) \tag{17}$$

where $H_{k+1} = H(\mathrm{X}_{k+1}, \mathrm{c}, \sigma)$, $\mathrm{X}_{k+1} = [\mathrm{x}_{\sum_{j=0}^{k} N_j + 1}, \cdots, \mathrm{x}_{\sum_{j=0}^{k+1} N_j}]^T$ and $Y_{k+1} = [y_{\sum_{j=0}^{k} N_j + 1}, y_{\sum_{j=0}^{k} N_j + 2}, \cdots, y_{\sum_{j=0}^{k+1} N_j}]^T$. The FNN output at the learning phase is as follows:

$$\hat{Y}_{k+1} = H_{k+1} \hat{Q}_{k+1} \tag{18}$$

where $\hat{Y}_{k+1} = [\hat{y}_{\sum_{j=0}^{k} N_j + 1}, \hat{y}_{\sum_{j=0}^{k} N_j + 2}, \cdots, \hat{y}_{\sum_{j=0}^{k+1} N_j}]^T$.



(a) 1 step ahead

(b) 2 steps ahead

(c) 5 steps ahead

(d) 10 steps ahead

Figure 4: Initial training response of the proposed approach

To avoid the singular problem for the matrix inversion of K_{k+1} in (17) while $N_0 < L$, one makes $P_0 = K_0^{-1}$ and applies the Woodbury identity to calculate P_0 as follows [20]:

$$P_0 = I / \lambda - H_0^T (\lambda I + H_0 H_0^T)^{-1} H_0 / \lambda. \tag{19}$$

Similarly, to avoid the ill-posed problem so that the computational cost for the matrix inversion of K_{k+1} in (17) while N_i □ L can be reduced, one makes $P_k = K_k^{-1}$ and $P_{k+1} = K_{k+1}^{-1}$, and applies the updating law of $\hat{Q}_{k+1}$ as follows:

$$P_{k+1} = P_k - P_k H_{k+1}^T (I + H_{k+1} P_k H_{k+1}^T)^{-1} H_{k+1} P_k, \tag{20}$$

$$\hat{Q}_{k+1} = \hat{Q}_k + P_{k+1} H_{k+1}^T (Y_{k+1} - H_{k+1} \hat{Q}_k). \tag{21}$$

4 MULTI-STEP PREDICTION SCHEME

MHC prediction is essentially an online time-series prediction problem which should carry out updating and prediction concurrently. To carry out multi-step direct prediction, consider the nonlinear autoregressive with exogenous input (NARX) model as follows:

$$\begin{aligned} y_s(k+r) = f(&y_s(k), y_s(k-r), y_s(k-2r), \cdots, y_s(k-nr), \\ &x_s(k), x_t(k-r), y_s(k-2r), \cdots, x_s(k-nr)) \end{aligned} \tag{22}$$

where x_s and y_s are the input and target feature variables, respectively, r is the prediction step, n+1 is the maximum lag, i.e., the order of the system. Then, give a time-series data set: $\mathrm{T} = \{(x_s(i), y_s(i))\}_{i=1}^{\infty}$, its initial set: $\mathrm{T}_0 = \{(x_s(i), y_s(i))\}_{i=1}^{n_0}$ with $n_0 > (n+1)r$, and choose the root-mean-square error (RMSE) as the performance index. Based on the proposed learning strategy, the multi-step direct prediction scheme of time-series is presented as follows.

Step 1) **Offline Initialization**: Obtain the initial training data set: $\mathrm{N}_0 = \{(\mathrm{x}_l, y_l)\}_{l=1}^{N_0}$, where $N_0 = n_0 - (n+1)r$, and

$$\begin{aligned} \mathrm{x}_l = [&x_s(l), x_s(l+r), \cdots, x_s(l+nr), \\ &y_s(l), y_s(l+r), \cdots, y_s(l+nr)]^T, \end{aligned} \tag{23}$$

$$y_l = y_s(l + (1+n)r). \tag{24}$$

a) Randomly generate parameters: c and σ ;

b) Calculate $H_0 = H(\mathrm{X}_0, \mathrm{c}, \sigma)$ by (10), where $\mathrm{X}_0 = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{N_0}]^T$;

c) Calculate $\hat{Q}_0$ using (13) with (14) (if $N_0 \ge L$) or with (19) (if $N_0 < L$);

d) Calculate the initial training performance: RMSE $_{\mathrm{train}}(\hat{Y}_0, Y_0)$ with $\hat{Y}_0 = H_0 \hat{Q}_0$ and $Y_0 = [y_1, y_2, \cdots, y_{N_0}]^T$;

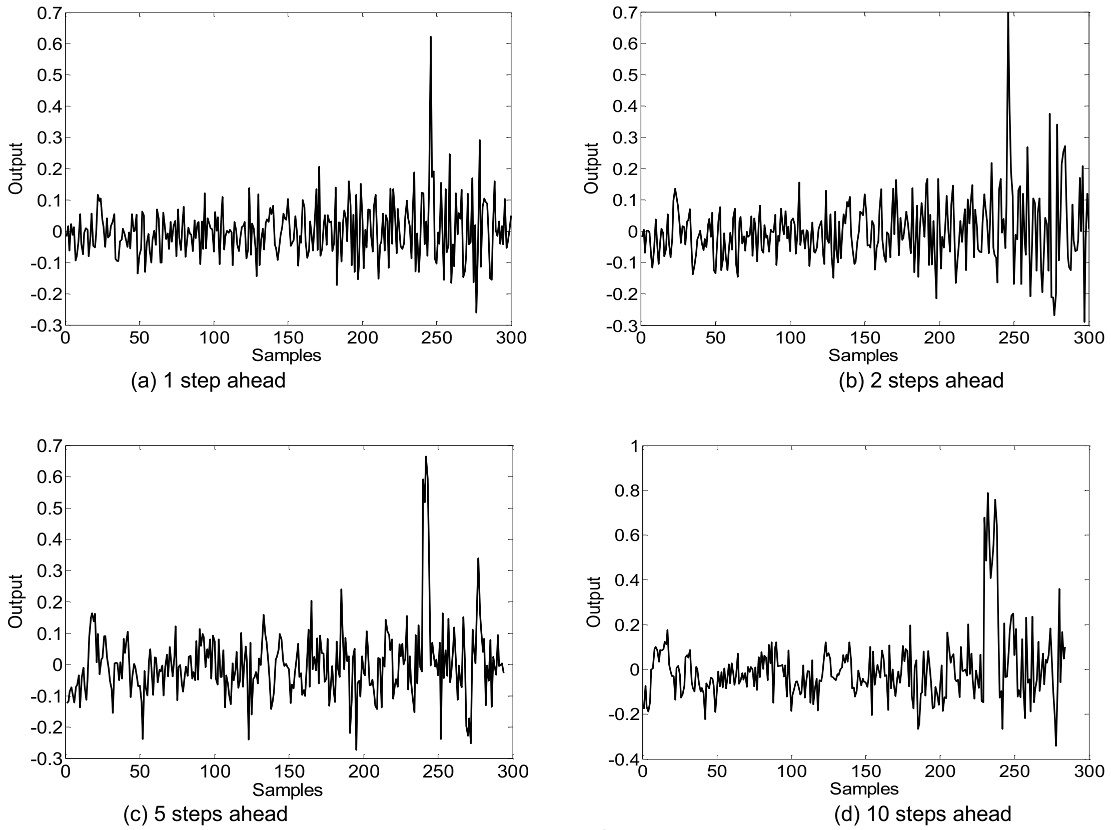
e) Predict the next *r* step's time-series:

$$\hat{y}_{N_0+r} = H(\mathbf{x}_{N_0+r}^T, \mathrm{c}, \sigma)\hat{Q}_0 \ ;$$

f) Let $Y_{10} = y_{N_0+1}$ and $\hat{Y}_{10} = \hat{y}_{N_0+1} = H(\mathbf{x}_{N_0+1}^T, \mathrm{c}, \sigma)\hat{Q}_0$;

g) Set the training step: *k* = 0.

Step 2) Online Sequential Prediction: Present the (*k*+1)th training data set: $\mathrm{N}_{k+1} = (\mathrm{x}_{N_0+k+1}, y_{N_0+k+1})$, where x_{N_0+k+1} and y_{N_0+k+1} are given by (23) and (24), respectively.



(a) 1 step ahead

(b) 2 steps ahead

(c) 5 steps ahead

(d) 10 steps ahead

Figure 5: Initial training errors of the proposed approach

a) Calculate $H_{k+1} = H(\mathbf{x}^T_{N_0+k+1}, \mathrm{c}, \sigma)$ by (10);

b) Update the prediction performance: $\mathrm{RMSE}_{\mathrm{Pred}}(\hat{Y}_{(k+1)k}, Y_{(k+1)k} = [Y^T_{k(k-1)}, y_{N_0+k+1}]^T$, $\hat{Y}_{(k+1)k} = [\hat{Y}^T_{k(k-1)}, \hat{y}_{N_0+k+1}]^T$ and $\hat{y}_{N_0+k+1} = H_{k+1}\hat{Q}_k$;

c) Update $\hat{Q}_{k+1}$ using (17) with (16) (if $N_{k+1} \geq L$) or by (21) with (20) (if $N_{k+1} < L$);

d) Update the training performance: $\mathrm{RMSE}_{\mathrm{train}}(\hat{Y}_{k+1}, Y_{k+1})$, $Y_{k+1} = [Y^T_k, y_{N_0+k+1}]^T$, $\hat{Y}_{k+1} = H(\mathrm{X}_{k+1}, \mathrm{c}, \sigma)\hat{Q}_{k+1}$, and $\mathrm{X}_{k+1} = [\mathrm{X}^T_k, \ \mathbf{x}_{N_0+k+1}]^T$;

e) Predict the next *r* step's time-series:

$$\hat{y}_{(N_0+k+1)+r} = H(\mathbf{x}^T_{(N_0+k+1)+r}, \mathrm{c}, \sigma)\hat{Q}_{k+1};$$

f) Set the training step: *k* = *k* + 1 and go to Step 2.

5 SIMULATION STUDIES

The applied MHC monitoring data were collected from PRONOSTIA, an experimental platform dedicated to test and validate bearings fault detection, diagnostic and prognostic approaches [21]. As shown in Figure 2, the PRONOSTIA is composed of three main parts: a rotating part, a degradation generation part and a measurement part. The main objective of PRONOSTIA is to provide real experimental data that characterize the degradation of ball bearings along their whole operational life. This platform allows accelerating bearing degradation in only few hours. An example of the vibration raw signal gathered during a whole experiment is shown in Figure 3. The non-trendable and non-periodical statistical properties of this type of signals increase the difficulty of MHC prediction [22].

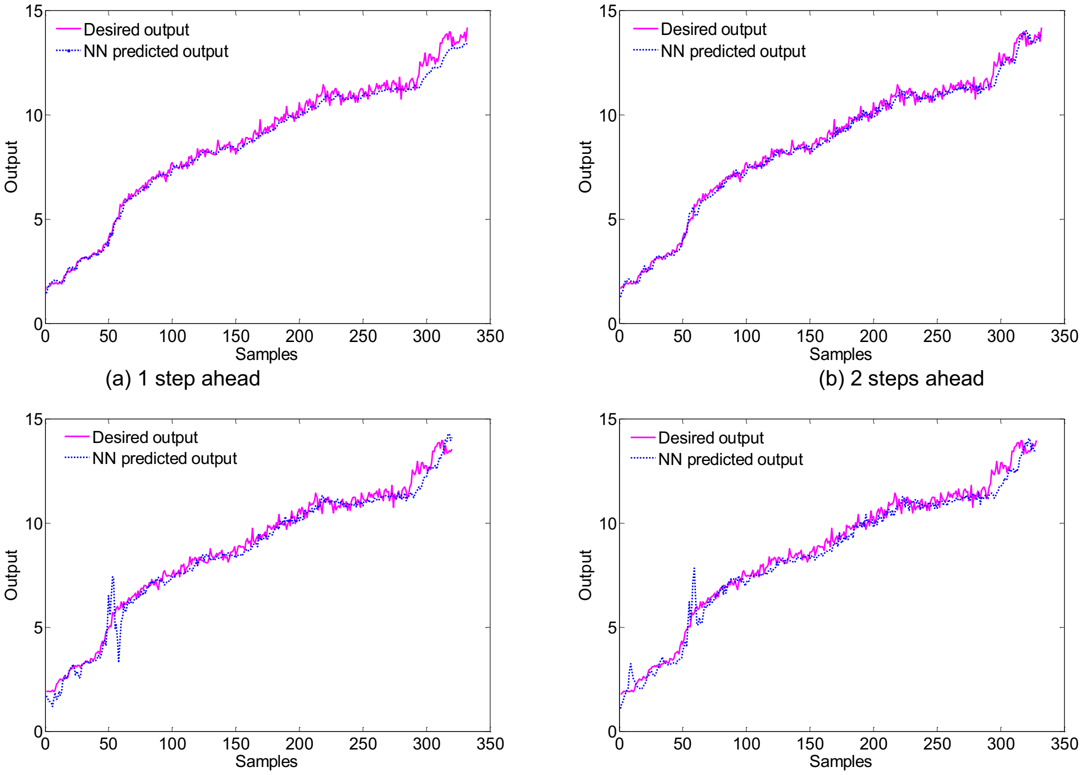
In this study, we choose two bearing data sets under the operating conditions: 1800 rpm speed and 4000 N load to carry out simulation. For the NARX model in (22), set *n* = 1, and *r* = 1, 2, 5 or 10, select x_s as the standard deviation (STD) of each vibration data set which consists of 2560 vibration signals, and y_s as the 5% trimmed mean of the vibration signal. The prediction procedure is as follows: First, the offline initialization is carried out based on one data set to obtain an intimal FNN model; second, the online prediction is carried out based on another data set to forecast time-series of *r* steps ahead. To demonstrate the superiority of the proposed EOSL-FNN, the OS-ELM in [16] and the NARX-NN are selected as the compared methods, where 10 notes is applied to the NARX-NN, and 100 notes with λ = 0.001 are applied to the EOSL-FNN and OS-ELM. Two performance indexes, namely the RMSE and the mean absolute percentage error (MAPE), are defined as follows:

$$\mathrm{RMSE}(\hat{Y}, Y) = [E((\hat{Y} - Y)^2)]^{1/2}, \tag{25}$$

$$\mathrm{MAPE} = \frac{1}{n}\left(\sum\nolimits_{t=1}^{n} |\,(y_t - \hat{y}_t)/y_t\,|\right) \times 100\%. \tag{26}$$

The Accuracy index is defined as $(100\% - \mathrm{MAPE})$.

The initial training and online prediction performance of the proposed EOSL-FNN are depicted in Figure 4 - 7. One observes that high training and predicting accuracy is obtained under small ahead step, and satisfied training and predicting accuracy is still obtained under large ahead step. The performance comparisons of all prediction methods in term of the time, RMSE, STD and accuracy are shown in Table I. Note that the results are obtained from averaging 10 times' simulation results. One observes that both the EOSL-FNN and the OS-ELM are extremely faster (with small training and predicting time) and more stable (with small STD) than the NARX-NN,



(a) 1 step ahead

(b) 2 steps ahead

(c) 5 steps ahead

(d) 10 steps ahead

Figure 6: Online prediction response of the proposed approach

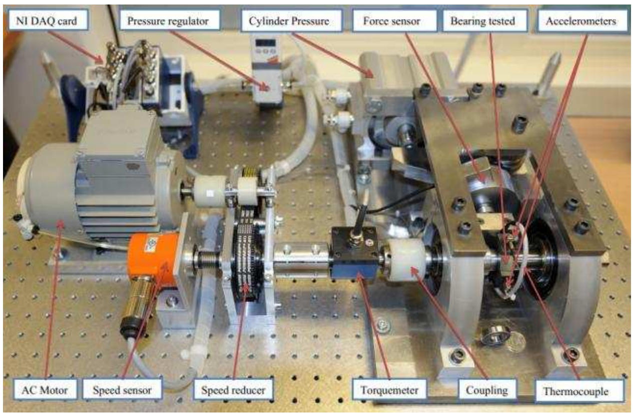


Figure 2: Experimental platform PRONOSTIA

the EOSL-FNN performs similar or better (with small RMSE and Accuracy) than the NARX-NN and OS-ELM, and the EOSL-FNN performs a little slower (with larger training and predicting time) than the OS-ELM since it contains more adjusting parameters.

6 CONCLUSIONS

In this paper, a novel EOSL-FNN has been developed and successfully applied to predict MHC. An online sequential learning strategy based on the ELM is developed to train the FNN. A multi-step time-series direct prediction scheme is presented to forecast bearing health condition online. The proposed approach not only keeps all salient features of the ELM, including extremely fast learning speed, good generalization ability and elimination of tedious parameter design, but also solves the singular and ill-posed problems caused by the situation that the number of training data is smaller than the number of hidden nodes. Simulation studies using real-world data from the accelerated bearing life have demonstrated the effectiveness and superiority of the proposed approach. Further work would focus on bearing long-term condition and remaining useful life prediction using online dynamic FNNs.

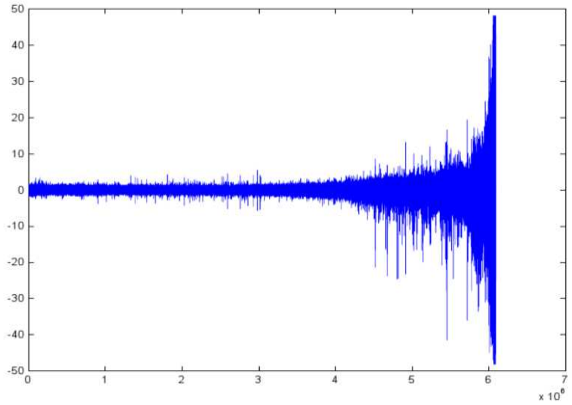
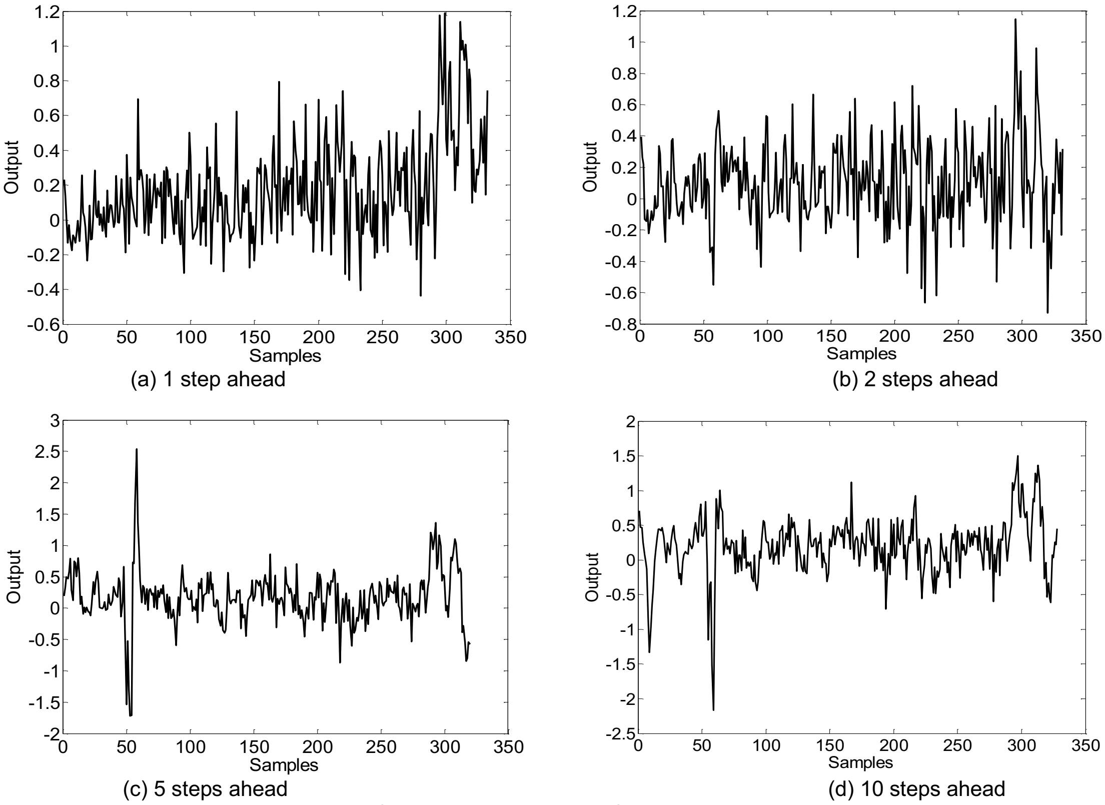


Figure 3: An example of the vibration raw signal

7 ACKNOWLEDGMENTS

This work is supported by the Science and Engineering Research Council, Agency for Science, Technology and Research (A*STAR), Singapore, under Grant no. 1122904016.



(a) 1 step ahead

(b) 2 steps ahead

(c) 5 steps ahead

(d) 10 steps ahead

Figure 7: Online prediction errors of the proposed approach

Step	NN Type	Training				Prediction			
		Time (s)	RMSE	STD	Accuracy (%)	Time (s)	RMSE	STD	Accuracy (%)
r = 1	ESL-FNN	0.0352	**0.0832**	54.010e-4	**97.197**	2.1145	**0.2343**	0.0354	**98.565**
	OS-ELM	**0.0312**	0.0865	34.100e-4	95.195	**2.0159**	0.2641	**0.0254**	97.548
	NARX-NN	1.5506	0.1153	**25.200e-4**	94.631	4.1824	0.3345	0.0191	96.744
r = 2	ESL-FNN	0.0334	**0.0987**	**5.6765e-4**	**97.120**	2.2387	**0.2645**	**0.0083**	**98.018**
	OS-ELM	**0.0250**	0.1056	6.9462e-4	94.585	**2.1141**	0.2837	0.0232	97.453
	NARX-NN	1.535	0.1220	197.00e-4	94.363	4.2151	0.4744	0.2707	95.970
r = 5	ESL-FNN	0.0388	**0.1054**	4.7654e-4	**95.078**	2.2416	**0.3879**	**0.0141**	**97.365**
	OS-ELM	**0.0324**	0.1181	**3.7799e-4**	94.044	**2.1541**	0.4562	0.0342	95.343
	NARX-NN	1.6427	0.1644	1474.0e-4	94.326	4.1434	0.4683	0.1815	95.832
r = 10	ESL-FNN	0.0295	**0.1250**	9.3490e-4	**94.418**	2.3015	**0.4561**	0.0355	**95.096**
	OS-ELM	**0.0264**	0.1441	**5.6543e-4**	93.317	**2.2784**	0.5441	**0.0341**	93.992
	NARX-NN	1.5085	0.1255	101.00e-4	94.285	4.0014	0.6344	0.1684	94.630

Table 1: Performance comparisons of all methods

8 REFERENCES

[1] G. J. Vachtsevanos, F. L. Lewis, M. Roemer, el al (2006) : Intelligent Fault Diagnosis and Prognosis for Engineering Systems. Hoboken, NJ: John Wiley & Sons.

[2] A. K. S. Jardine, D. M. Lin, and D. Banjevic (2006): A review on machinery diagnostics and prognostics implementing condition-based maintenance, Mech. Syst. Signal Proc., vol. 20, no. 7, pp. 1483-1510.

[3] M. Yu, D. Wang, M. Luo and Q. Chen (2012) : Fault detection, isolation and identification for hybrid systems with unknown mode changes and fault patterns, Expert Syst. Appl., vol.39, no.11, pp: 9955-9965.

[4] M. Yu, D. Wang, M. Luo and L. Huang (2011). Prognosis of hybrid systems with multiple incipient faults: augmented global analytical redundancy relations approach, IEEE Trans. Syst. Man Cybern. Part A, vol. 41, no. 3, pp. 540-551.

[5] X. S. Si, W. Wang, C. H. Hu, et al (2011): Remaining useful life estimation - A review on the statistical data driven approaches, Eur. J. Oper. Res., vol. 213, no. 1, pp. 1-14.

[6] Z. G. Tian and M. J. Zuo (2010) : Health condition prediction of gears using a recurrent neural network approach, IEEE Trans. Reliab., vol. 59, no. 4, pp. 700-705.

[7] F. G. Zhao, J. Chen, L. Guo, el al (2009): Neuro-fuzzy based condition prediction of bearing health, J. Vib. Control, vol. 15, no. 7, pp. 1079-1091.

[8] M. Brown and C. Harris (1994): Neuro-Fuzzy Adaptive Modeling and Control. Upper Saddle River, NJ: Prentice Hall.

[9] W. Wang, M. F. Golnaraghi, and F. Ismail (2004): Prognosis of machine health condition using neuro-fuzzy systems," Mech. Syst. Signal Proc., vol. 18, no. 4, pp. 813-831.

[10] W. Wang (2007): An adaptive predictor for dynamic system forecasting, Mech. Syst. Signal Proc., vol. 21, pp. 809-823.

[11] J. Liu, W. S. Wang, and F. Golnaraghi (2009): A multi-step predictor with a variable input pattern for system state forecasting, Mech. Syst. Signal Proc., vol. 23, no. 5, pp. 1586-1599.

[12] C. C. Chen and G. Vachtsevanos (2012) : Bearing condition prediction considering uncertainty: An interval type-2 fuzzy neural network approach, Robot. Comput.-Integr. Manuf., vol. 28, no. 4, pp. 509-516.

[13] C. C. Chen, B. Zhang, and G. Vachtsevanos (2012): Prediction of machine health condition using neuro-fuzzy and Bayesian algorithms, IEEE Trans. Instrum. Meas., vol. 61, no. 2, pp. 297-306.

[14] C. C. Chen, B. Zhang, G. Vachtsevanos, el al (2011) : Machine condition prediction based on adaptive neuro-fuzzy and high-order particle filtering, IEEE Trans. Ind. Electron., vol. 58, no. 9, pp. 4353-4364.

[15] G. B. Huang, Q. Y. Zhu, and C. K. Siew (2006) : Extreme learning machine: Theory and applications," Neurocomputing, vol. 70, no. 1-3, pp. 489-501.

[16] N. Y. Liang, G. B. Huang, P. Saratchandran, et al (2006) : A fast and accurate online sequential learning algorithm for feedforward networks," IEEE Trans Neural Netw, vol. 17, no. 6, pp. 1411-1423.

[17] H. J. Rong, G. B. Huang, N. Sundararajan, et al (2009) : Online sequential fuzzy extreme learning ma-chine for function approximation and classification problems," IEEE Trans Syst Man Cybern B Cybern, vol. 39, no. 4, pp. 1067-1072.

[18] S. Q. Wu, M. J. Er, and Y. Gao (2001) : A fast approach for automatic generation of fuzzy rules by generalized dynamic fuzzy neural networks, IEEE Trans. Fuzzy Syst., vol. 9, no. 4, pp. 578-594.

[19] Y. Lin and G. A. Cunningham III (1995) : A new approach to fuzzy-neural system modeling," IEEE Trans. Fuzzy Syst., vol. 3, no. 2, pp. 190-198.

[20] H. T. Huynh and Y. Won (2011) : Regularized online sequential learning algorithm for single-hidden layer feedforward neural networks," Pattern Recognit. Lett., vol. 32, no. 14, pp. 1930-1935.

[21] P. Nectoux, R. Gouriveau, K. Medjaher, et al (2012): PRONOSTIA: An experimental platform for bearings accelerated degradation tests, IEEE Int. Conf. Prognostics Health Management, Denver, Co, pp. 1-8.

[22] S. Porotsky (2012): Remaining useful life estimation for systems with non-trendability behaviour, IEEE Int. Conf. Prognostics Health Management , Denver, Co, pp. 1-6.