A Deep Adversarial Learning Prognostics Model for Remaining Useful Life Prediction of Rolling Bearing

Bi-Liang Lu, Zhao-Hua Liu, *Member, IEEE*, Hua-Liang Wei, Lei Chen, Hongqiang Zhang, and Xiao-Hua Li

***Abstract*—Remaining useful life (RUL) prediction for condition-based maintenance decision making plays a key role in prognostics and health management (PHM). Accurately predicting RUL of the rotating components of complex machines becomes a challenging task for PHM. For many existing methods the current prediction error of RUL prediction may be accumulated into the future predictions, and thus can lead to a prediction error superposition problem. In this paper, the formation mechanism of prediction error superposition is analyzed, and for the first time a deep adversarial long short-term memory (LSTM) prognostic framework is proposed to overcome the major issue related to prediction error superposition. In the proposed framework, a generative adversarial network (GAN) architecture combining the LSTM network and auto-encoder (AE) is investigated for bearing RUL monitoring. In the proposed deep adversarial learning prediction framework, due to the potential involvement of long-term and complex tasks, the LSTM network (generator) is used to predict the degradation process of rolling bearings based on available historical data, and a simple but useful AE (discriminator) is used to determine and refine the accuracy of the prediction. Therefore, the AE plays the adversarial role of the LSTM network, and the prediction accuracy of the LSTM network can be significantly improved. For illustration purpose, two practical case studies, which use a series of bearing degradation data and the IEEE PHM 2012 PRONOSTIA datasets, respectively, are presented to show the prediction performance of the proposed method. Experimental results show that the proposed method works very well for vibration monitoring and performs better in comparison with the reference machine learning and deep learning approaches.**

***Impact Statement* — The damage of rolling bearing usually leads to a significant consequence to industrial production process. However, the existing remaining useful life prediction methods for rolling bearing have a prediction error superposition problem that can affect the multi-step prediction performance. The new adversarial learning prognostics model proposed in this paper can overcome the problem. The proposed method uses long short-term memory network as a generator to predict remaining useful life for rolling bearing, and uses auto-encoder as a discriminator to estimate the prediction accuracy. The method can significantly improve the multi-step prediction accuracy of remaining useful life for rolling bearing, and provides reliable and scientific strategy in prognostics and health management of mechatronics equipment.**

***Index Terms*—Auto-encoder (AE), condition monitoring, deep adversarial learning, deep learning, generative adversarial network (GAN), long short-term memory (LSTM), prediction error superposition, prognostics and health management, remaining useful life (RUL) prediction, rolling bearings.**

I. Introduction

ROLLING bearings are important and expensive rotating components in many industrial machinery systems, such as high-speed railway, induction motors, and wind turbine drivetrain systems. The harsh working environment of rotating machinery makes rolling bearings expose to high-low temperature, high pressure, and humid working conditions, and all this can quickly cause damage to the rolling bearings. Any failure of the rolling bearings can cause the entire machine to malfunction and result in high-priced maintenance costs. If rolling bearing faults can be predicted in advance, the entire rotating machinery shutdown caused by bearing faults can be avoided. With the development of industrial Internet, large amounts of data have been collected from the prognostics and health management (PHM) systems in rotating machines. Remaining useful life (RUL) prediction for degrading systems plays a key role in any PHM. Accurately predicting the RUL of the rotating components from monitoring data with new advanced approaches has attracted increasingly more attention in recent years [1]-[4].

The existing RUL prediction methods proposed for PHM can be categorized into two types: model-based methods and data-driven methods [5], [6]. Model-based methods require to establish a mathematical or physical model of the machine to describe the degradation process, and joint measured data to determine model parameters. For example, Li *et al.* [7] proposed an improved exponential model based on adaptive first predicting time (FPT) selection approach for predicting

Manuscript received December 29, 2020; revised April 3, 2021 and May 29, 2021; accepted July 10, 2021. This work was supported in part by the National Key Research and Development Project of China under Grant 2019YFE0105300, in part by the National Natural Science Foundation of China under Grant 61972443 and Grant 61503134, in part by the Hunan Provincial Hu-Xiang Young Talents Project of China under Grant 2018RS3095, in part by the Hunan Provincial Natural Science Foundation of China under Grant 2018JJ2134 and Grant 2020JJ5199, and in part by the Outstanding Youth Project of Education Department of Hunan Province of China under Grant 19B200. *(Corresponding author: Zhao-Hua Liu.)*

B.-L. Lu, Z.-H. Liu, L. Chen, H. Zhang, and X.-H. Li are with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, 411201, China (e-mail: 1197393632@qq.com; zhaohualiu2009@hotmail.com; chenlei@hnust.edu.cn; hongqiangzhang@hnust.edu.cn; lixiaohua_0227@163.com).

H.-L. Wei is with the Department of Automatic Control and Systems Engineering, the University of Sheffield, Sheffield S1 3JD, U.K (e-mail: w.hualiang@sheffield.ac.uk).

RUL of rolling bearings. Lei *et al.* [8] proposed a model-based method for RUL prediction of machinery. In addition, model-based RUL prediction methods have also been employed in many other applications [9]-[11]. One advantage of model-based RUL prediction methods is that they can combine the expert knowledge with actual operational data. As a result, such a method has favorable performance when dealing with RUL prediction for rolling bearings. However, there are two main defects for rolling bearings using existing model-based RUL prediction methods: 1) The physical or mathematical model of rolling bearing drive systems could be tremendously complicated, even the mechanical principle of rotating machinery itself is a highly complex system whose model can be very complex. A reduced mathematical model used for RUL prediction may not be able to sufficiently represent the real system of interest, as a consequence, a model-based RUL prediction method may fail. 2) Different rolling bearings need different models. For those rolling bearings whose mathematical models have already been available, the cost of undertaking model-based RUL prediction may be small and acceptable, but for rolling bearings whose models are not available, the cost of building a mathematical model could be extremely expensive.

Data-driven approach tends to infer the rolling bearing degradation process (RBDP) from historical actual data through machine learning techniques [12]-[15]. Therefore, data-driven RUL prediction methods are mainly affected by two factors: 1) the quantity and quality of the historical actual data, and 2) the prediction performance of the machine learning models. Recently, with the rapid development of sensor technology and especially deep learning in Industry 4.0, highly accurate and robust sensors can capture high-quality and sufficient quantity data from complex real-world applications, and deep learning models such as deep belief network (DBN) and long-short term memory (LSTM) network can use these captured data to characterize the RBDP. Specially, Zhang *et al.* [16] proposed a multi-objective DBN in which the multi-objective evolutionary algorithm embedded with the traditional DBN learning technique to achieve RUL estimation. Guo *et al.* [17] proposed a recurrent neural network based health indicator for RUL prediction of bearings, where data-driven method can incorporate both measured information and intelligent deep learning model. Consequently, data-driven approach becomes an indispensable alternative for RUL prediction of bearings.

Data-driven RUL prediction methods are usually implemented through one-step time-series prediction or iterative multi-step ahead RUL prediction. The former needs to know the entire RBDP, and uses the measured data to make prediction. The latter only needs to know the first half of the RBDP, and the latter half is inferred by the prediction model constructed using deep learning methods. Obviously, a data-driven RUL prediction method based on one-step time-series prediction is simpler and more accurate, and detailed results on this may be found in [18] and [19]. In [18], a one-step time-series RUL estimation was proposed based on support vector regression (SVR). In [19], a SVR method was proposed for the monitoring of rolling bearings, where the estimation of the RUL is achieved by a one-step time-series predictor based on SVR. More work on SVR-based prediction methods can be found in [20]-[23]. One-step time-series prediction method can be beneficial if the entire RBDP is available. Nevertheless, the dependence on whether the entire RBDP is available is challenging for practical applications. In contrast, iterative multi-step ahead RUL prediction is more suitable for actual working conditions.

Iterative multi-step ahead RUL prediction forecasts one or several steps from previously observed data, then adds the predicted points to the historical data to predict the next part, and so as to iteratively get the entire RBDP. Because of this characteristic, such an approach is more challenging compared with a one-step time-series prediction based method, but it has attracted more and more attention in recent years, due to its effectiveness. For example, Xia *et al.* [24] applied deep neural network architecture to develop a two-stage prognostics method for the RUL prediction of bearings. Sun *et al.* [25] proposed a sparse auto-encoder (AE) based on deep transfer learning method for RUL prediction. It should be noted that although these methods have good performance, there is still a nonnegligible issue, that is, the prediction error superposition problem, which has not been resolved. This is a case where the previous prediction error is accumulated into the next prediction. This is related to the iterative multi-step ahead RUL prediction method mentioned above.

Recently, deep learning techniques [26], especially generative adversarial networks (GAN), have been widely used in natural language processing (NLP), computer vision (CV), etc. GANs normally consist of the generative model *G* and discriminative model *D*, leading them to Nash equilibrium [27]-[30]. The *G* aims to estimate the data distribution, whilst the *D* is designed to evaluate the probability; the two against each other, and this mechanism can be used to improve the performance of the GAN. The core idea of the 'generator' and 'discriminator' proposed in GANs can be used to solve the prediction error superposition problem. In this paper, a deep adversarial LSTM framework is proposed for rolling bearing RUL prediction using a two-stage scheme. In the first stage, the generator is used to predict the RBDP when historical data are given. In the second stage, the discriminator decides whether the RBDP is derived from real historical data or predicted RBDP. In this way, the prediction error superposition problem can be well solved.

The main contributions of this study are as follows:

1) Traditional RUL prediction method uses one-step prediction. In this paper, a iterative multi-step ahead RUL prediction method based on multi-step time series prediction is proposed, which is more suitable for real application of RUL prediction. Based on the observation and analysis of the prediction error superposition problem, a novel deep adversarial LSTM-based prognostics framework for bearing RUL prediction is developed to reduce the effect of error propagation in iterative multi-step ahead RUL prediction.
2) In order to validate the efficacy of the proposed

framework, a deep adversarial LSTM prognostics model is proposed in the experiment stage. Firstly, a LSTM network is used to generate RBDP based on the available measured data. The AE then discriminates whether the inputted RBDP belongs to the measured data or the predicted RBDP, and the two against each other. The experimental results show that the proposed model works well for RUL prediction.

3) Two practical case studies are carried out based on two real datasets: the first dataset for accelerated degradation tests of rolling element bearings and the second one for the public PRONOSTIA bearing. Detailed analysis results are reported.

The remaining parts of this paper are summarized as follows. The related work is presented in Section II. In Section III, a novel deep adversarial LSTM approach is provided for RUL prediction of rolling bearings. Detailed analysis and comparisons of experimental results are shown in Section IV. At last, the main conclusions and future work are given in Section V.

II. The Proposed Adversarial Learning Architecture for Prediction Error Superposition Problem

A. Prediction Error Superposition Problem

As mentioned earlier, many data-driven RUL prediction methods have been reported in [12]-[15]. This work considers data-driven RUL prediction method and takes into account the superposition problem of prediction errors in this stage.

1) Problem Analysis of Time Series Prediction

For rolling bearing, the degradation process (i.e. normal status becomes malfunctioning) is essentially a time series. Therefore, one-step time-series prediction can be interpreted as: given a time-series sample set *X*, where $X=\{x_1,x_2,x_3,......,x_n\}$, using $\hat{x}_n$ to represent a predicted value, the process of making one-step time-series prediction is shown as follows:

$$\begin{cases}\{x_1,x_2,x_3,......,x_{n-1}\}\rightarrow\hat{x}_n\\ \{x_2,x_3,x_4,......,x_n\}\rightarrow\hat{x}_{n+1}\\ \{x_3,x_4,x_5,......,x_{n+1}\}\rightarrow\hat{x}_{n+2}\\ \{x_4,x_5,x_6,......,x_{n+2}\}\rightarrow\hat{x}_{n+3}\end{cases}\qquad(1)$$

where the symbol ‘$\mathrm{A}\rightarrow\mathrm{B}$’ means value B is predicted from value A. Specifically, A denotes the measured data and B denotes the predicted value from the measured RBDP. Unlike one-step time-series prediction, iterative multi-step ahead RUL prediction can be expressed as:

$$\begin{cases}\{x_1,x_2,x_3,......,x_{n-1}\}\rightarrow\hat{x}_n\\ \{x_2,x_3,x_4,......,\hat{x}_n\}\rightarrow\hat{x}_{n+1}\\ \{x_3,x_4,......,\hat{x}_n,\hat{x}_{n+1}\}\rightarrow\hat{x}_{n+2}\\ \{x_4,......,\hat{x}_n,\hat{x}_{n+1},\hat{x}_{n+2}\}\rightarrow\hat{x}_{n+3}\end{cases}\qquad(2)$$

The prediction error superposition problem is clearly reflected in (2). For graphical illustration, it is also depicted in Fig. 1. Obviously, there is a prediction error in $\{x_1,x_2,x_3,......,x_{n-1}\}\rightarrow\hat{x}_n$, because usually $x_n\neq\hat{x}_n$. With $\hat{x}_n$ being used to predict the next point, the result would be that the prediction error would become larger and larger after many steps. Such an error propagation phenomenon ubiquitously exists in time series prediction.

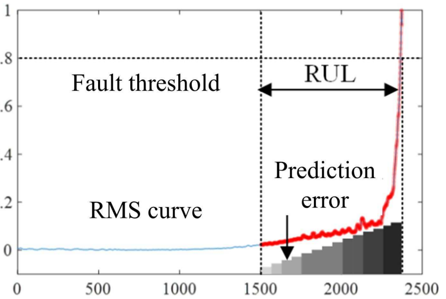


Fig. 1. An illustration of prediction error superposition.

2) Problem Definition

Let $X=\{x_1,x_2,...,x_m,...,x_n\}$ be a time series consisting of *n* samples. Suppose that the first half, $\mathcal{T}=\{x_1,x_2,...,x_m\}$, is used as the training set to train the model, and the second half, $\mathcal{P}=\{x_{m+1},x_{m+2},...,x_n\}$, is used as the prediction set to assess the performance. Let $\mathcal{T}_i\in\mathbb{R}^m$ and $\mathcal{P}_j$ denote the *i*-th *m*-dimensional training sample and the *j*-th prediction sample, respectively. This study aims at minimizing the prediction error superposition problem. This is equivalent to maximizing the conditional probability of $\mathcal{P}_j$ when the training sample $\mathcal{T}_i$ is given. The above conditional probability can be defined as follows:

$$\begin{cases}p(\mathcal{P}_1|\mathcal{T}_1)=p(x_{m+1}|x_1,...,x_m)\\ p(\mathcal{P}_2|\mathcal{T}_2)=p(x_{m+2}|x_2,...,x_{m+1})\\ \qquad\cdots\\ p(\mathcal{P}_n|\mathcal{T}_n)=p(x_n|x_{n-m},...,x_{n-1})\end{cases}\qquad(3)$$

This problem is similar to natural language processing (NLP) and video captioning (VC). A major difference is that NLP and VC use $\mathcal{T}$ to predict the whole $\mathcal{P}$, but in RUL prediction, only one point is predicted each time and finally superimposed to form the entire RBDP.

B. A Solution to the Prediction Error Superposition Problem

In order to overcome the error propagation issue, in this section, the adversarial learning architecture-based LSTM is presented to generate RUL prediction under the influence of the GAN. The structure of the adversarial learning for predicting RUL of rolling bearings is shown in Fig. 2. Specifically, the adversarial learning architecture-based LSTM (named LSTM-GAN) is made up of two parts: the generator (denoted by *G*) and the discriminator (denoted by *D*). The *G* gives the guidelines that generate a series of the prediction values given a short training set $\mathcal{T}$. The *D* compresses the input sequence and outputs the corresponding label, indicating whether the generated prediction values are correct and reasonable and meet the forecast requirements.

The objective function of the designed architecture is

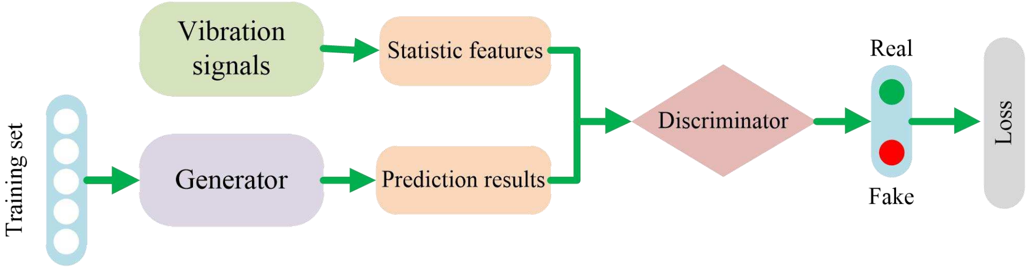


Fig. 2. An adversarial learning architecture.

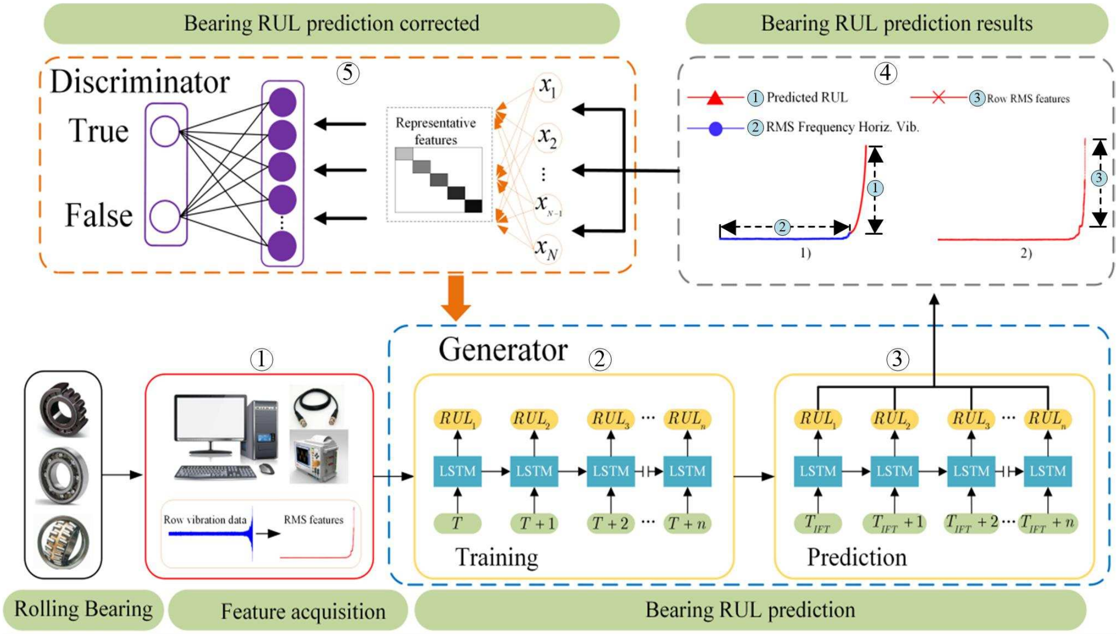


Fig. 3. The proposed adversarial learning architecture-based LSTM for bearing RUL prediction. $\{T, T+1, \ldots\ldots, T+n\}$ represents the training set, $\{T_{IFT}, T_{IFT}+1, \ldots\ldots, T_{IFT}+n\}$ means the prediction part, and IFT is the incipient failure threshold. Meanwhile, the notation 'Horiz. Vib.' means the horizon vibration, and the steps of the training process given in Fig. 3, that is, ①→②→③→④→⑤→②...→⑤ ..., give the entire iteration and adversarial training process.

introduced as follows. For a given training set $\mathcal{T}$, the conditional probability of $\mathcal{P}$ can be expressed as:

$$\begin{aligned} p(\mathcal{P}|\mathcal{T}) &= p(x_{m+1},\ldots,x_n|x_1,\ldots,x_m) \\ &= \prod_{i=1}^{n-m} p(x_{m+i}|x_i, x_{i+1},\ldots,x_{m+i-1}) \end{aligned} \tag{4}$$

In this study, a number of internal parameters are considered to minimize the prediction error superposition. The conditional probability can be used to obtain optional parameters as follows:

$$\theta^* = \arg\max_{\theta} \sum_{i=1}^{n-m} \log p(x_{m+i}|x_i, x_{i+1},\ldots,x_{m+i-1};\theta) \tag{5}$$

where θ is a vector of the parameters (weight, bias, etc.) of the designed model. For D, the main goal is to find the proper discriminator that can be used to process the input sequence and map input features to the output domain $D(\mathcal{P}) \in [0,1]$, where $D(\bullet)$ is a function indicating that the probability that $\mathcal{P}$ comes from real RBDP, rather than from the generative model G. Then, the objective function of D can be integrated into a loss using the following equation:

$$\begin{aligned} \mathcal{L}_D(y, D(\mathcal{P})) = -\frac{1}{m}\sum_{i=1}^{m}[(y^{(i)}\log(D(\mathcal{P}_i)) \\ +(1-(y^{(i)})*\log(1\text{-}D(\mathcal{P}_i)))] \end{aligned} \tag{6}$$

where m represents the quantity of input samples. $y^{(i)}$ and $D(\mathcal{P}_i)$ represent the true and discriminative labels, respectively. The main goal of training process is to minimize the log likelihood, which can be represented as follows:

$$\begin{aligned} \min_G \max_D = \mathbb{E}_{x\sim P_{(\mathcal{P})}}[\log p(\mathcal{P}|\mathcal{T})] \\ + \mathbb{E}_{x\sim P_{(\mathcal{T})}}[\log(1-D(G(\mathcal{P})))] \end{aligned} \tag{7}$$

III. The Proposed Adversarial Learning Architecture-Based LSTM Prognostics Framework for RUL Prediction

A solution to the prediction error propagation problem is proposed in this section. A graphical illustration of the proposed solution is shown in Fig. 3, where the generator G (LSTM) and discriminator D (AE) form the main framework of the adversarial learning architecture. The LSTM network uses a prediction-overlay-prediction method to generate the relevant RBDP. Meanwhile, the AE network, as the discriminator, is used for assessing whether the generated time series in sequence is correct. More detailed explanations for Fig. 3 are given below.

A. Generative Model Based on LSTM

Traditional recurrent neural network (RNN) models [31], [32] aim to capture complex temporal dynamical behavior by mapping the samples to the nodes directionally connected into a ring and exhibiting dynamic time behavior by its internal state:

$$\begin{aligned} h_t &= f(W_h x_t + \psi_h h_{t-1} + b_h) \\ o_t &= f(\psi_h h_t + b_0) \end{aligned} \tag{8}$$

where W_h and ψ_h are the weight parameters, b represents the bias parameters, and f is the activation function such as tanh; x_t is the input sequence, $h_t \in \mathbb{R}^M$ is the hidden layer state which is obtained by weighting the input sequence through the hidden unit, and o_t means the output sequence at the t moment.

The RNN has achieved great success in text generation and speech recognition. However, due to the exploding and vanishing gradient problem, it is difficult to deal with long distance dependence problem. The LSTM network [33] has been proven to avoid gradient disappearance and exploding problem. This is because that it combines a memory cell and a forget gate to help the network learn long-range temporal dependencies, while the forget gate can remove invalid hidden states and update them with new information in a timely manner. The basic structure of the LSTM network, consisting of a memory cell and several control gates, is illustrated in Fig. 4. Let x_t , c_t , and h_t represent the time-series input, cell state, and the hidden layer state at the t moment, respectively. Given a time series $X = \{x_1, x_2, x_3, ..., x_T\}$, two sequences will be calculated by the LSTM network: the hidden state sequence $h_t = \{h_1, h_2, h_3, ..., h_T\}$, and the memory cell sequence $c_t = \{c_1, c_2, c_3, ..., c_T\}$. Formally, the forward propagation algorithm of the LSTM network is expressed as:

$$\begin{aligned} i_t &= \sigma(W_i[h_{t-1}, x_t] + b_i) \\ f_t &= \sigma(W_f[h_{t-1}, x_t] + b_f) \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\ g_t &= \sigma(W_g[h_{t-1}, x_t] + b_g) \\ c_t &= f_t \otimes c_{t-1} + i_t \otimes g_t \\ h_t &= o_t \otimes \tanh(c_t) \end{aligned} \tag{9}$$

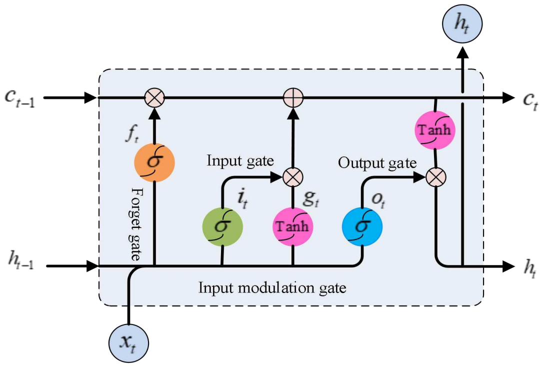


Fig. 4. The basic structure of LSTM.

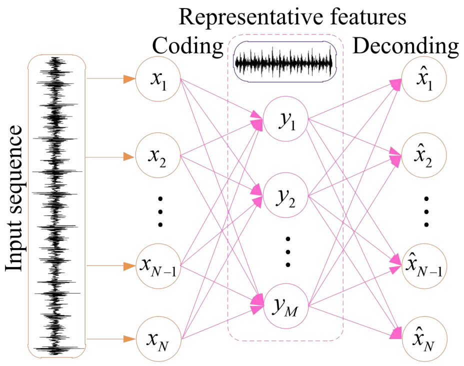


Fig. 5. The auto-encoder processor of input sequence in discriminative model.

where the symbol $\otimes$ represents point-wise product and the W is matrix used in the LSTM weight parameters. Finally, σ and tanh are sigmoid and hyperbolic tangent activation functions, respectively.

B. Discriminative Model Based on Auto-Encoder

Typically, the AE consists of a three-layer neural network, which together forms the encoding layer and decoding layer [34]-[36]. The encoding layer compresses the original input sequence into the representative features by reducing the number of neurons. Afterwards, the representative features are converted by the decoder into a reconstructed sequence with the same dimension as the input sequence. The main purpose of AE is to rebuild the input sequence at the decoding layer; it is mainly achieved by optimizing the loss value between the reconstruction sequence and the input sequence. The general architecture of the AE is illustrated in Fig. 5.

The AE maps the input sequence $x \in \mathbb{R}^n$ to the representative feature space $y \in \mathbb{R}^r (r < n)$ by the activation function $f(x)$ and AE network parameters, which can be defined as

$$y = f(Wx + b) \tag{10}$$

where W is the weight parameters with the $r \times n$ dimensions, and b means the bias parameters. Correspondingly, the decoding part is to restore the representative features y to the same dimension as the input sequence by using:

$$\hat{x} = f(\hat{W}y + \hat{b}) \tag{11}$$

where $\hat{W}$ represents the weight parameters of decoding part with $n \times r$ dimensions, and $\hat{b}$ means the bias parameters of decoding part. Then, the training of the entire AE network updates W , $\hat{W}$, b , and $\hat{b}$ through the following loss function:

$$\begin{aligned} \theta^*, \hat{\theta}^* &= \arg\min_{\theta, \hat{\theta}} \frac{1}{2m} \sum_{i=1}^{m} \mathcal{L}(x^{(i)}, \hat{x}^{(i)}) \\ &= \arg\min_{\theta, \hat{\theta}} \frac{1}{2m} \sum_{i=1}^{m} \mathcal{L}(x^{(i)}, f(\hat{W}(f(Wx + b)) + \hat{b})) \end{aligned} \tag{12}$$

where $\theta \triangleq \{W, b\}$, $\hat{\theta} \triangleq \{\hat{W}, \hat{b}\}$, the symbol $\mathcal{L}$ represents the formula such as $\mathcal{L}(x, x') = \| x - x' \|^2$, and m is the number of samples involved in the input sequence.

C. Representative Features Classifier

The structure of the multiple hidden layers enables AE to extract good representative features. These features can be used to implement a multi-classification task using multi-classifier such as softmax [37], [38]. Given an input sequence with m samples, $\left\{x^{(i)}\right\}_{i=1}^{m}$, where $x^{(i)} \in \mathbb{R}^n$, involving k types of labels, that is $y^{(i)} \in \left\{1,2,......,k\right\}$, , and $i = 1,2,\cdots,m$, the main function of softmax is to estimate the probability that each sample belongs to each category, and takes the category with the highest probability as the class of the sample. This probability is given by:

$$p(y_{i=j}^{(i)} \left| x^{(i)};\theta) = \left[p(y^{(i)} = 1 \left| x^{(i)};\theta),..., p(y^{(i)} = k \left| x^{(i)};\theta) \right]^T \right.\right.\right.$$

$$= \frac{1}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \cdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \tag{13}$$

where $y_{i=j}^{(i)} \in \left\{1,2,......,k\right\}$. $\theta_1,\theta_2,......,\theta_k$ are the model parameters with k types of labels, and $1 / \sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}$ is a hypothesis function to regularize the probability distribution.

The loss function of softmax model can be expressed as:

$$\mathcal{L}_{J_\theta} = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_j^T x^{(i)}}}\right] \tag{14}$$

where the symbol '1{}' represents the indicator function in which 1{a true statement}=1 and 1{a false statement}=0. In AE network, the softmax regression is placed at the end of the structure to classify the representative features of the AE extracted by minimizing $\mathcal{L}_{J_\theta}$.

D. Optimization Solution

In this study, a novel adversarial learning architecture-based LSTM is investigated for rolling bearing RUL prediction, and the stochastic gradient descent (SGD) algorithm is selected to search the optimal parameters. The main process is briefly listed in Algorithm 1 below.

IV. Experiment Tests

In this section, the proposed LSTM-GAN method is applied to bearing RUL prediction. The results of one-step time-series prediction and iterative multi-step ahead RUL prediction are calculated based on two real datasets, that is, a series of bearing degradation data [8], and the IEEE PHM 2012 PRONOSTIA dataset [39]. In this paper, two forecasting methods, namely, one-step time-series prediction method and iterative multi-step ahead RUL prediction, are considered. For the former, a sliding window scheme is adopted to train the predictive model that is later used. This means that the model will use historical data to predict the next observation. For the latter, the main idea is to directly predict the remaining useful life based on the training data, which is more challenging than the one-step prediction method.

Algorithm 1: The adversarial learning algorithm-based LSTM prognostics model for bearing RUL prediction.

```
Input: The training set $\mathcal{T} = \{x_1,x_2,...,x_m\}$ and the RBDP $X = \{x_1,x_2,...,x_m,...,x_n\}$ for discriminative model $D$ to output $D(\mathcal{P}) \in [0,1]$.
Output: The prediction set $\mathcal{P} = \{x_{m+1},x_{m+2},...,x_n\}$ .
1.1: Set the LSTM parameters:
●    activation ="relu', loss='mean squared error', optimizer='sgd', epochs=10000, batchsize=64,
●    #hidden layer =1,
●    #nodes=20 per hidden layer ;
1.2: Set the AE discriminator:
●    Trade-off parameter for the weight decay term $\lambda = 0.001$ ;
2: Obtain the predicted vectors $\left\{\mathcal{P}\right\}_{i=m+1}^{n} = trained\_LSTM\left(\mathcal{T}\right)$;
3: Optimize the generative model $G$ using SGD on the predicted
    vector $\left\{\mathcal{P}\right\}_{i=m+1}^{n}$ only; optimize the discriminative model $D$ using
    SGD on the predicted vector $\left\{\mathcal{P}\right\}_{i=m+1}^{n}$ and the
    RBDP $X = \{x_1,x_2,...,x_m,...,x_n\}$ ; the stochastic gradient of two
    optimization processes can be written as follow:
        $\nabla_{\theta_d} \frac{1}{m}\sum_{i=1}^{m}[\log D(\mathcal{T}) + \log(1 - D(G(\mathcal{P})))]$
        $\nabla_{\theta_g} \frac{1}{m}\sum_{i=1}^{m}[\log(1 - D(G(\mathcal{P})))]$        (15)
4: This process is repeated until the AE input features
    vector $\left\{\mathcal{P}\right\}_{i=m+1}^{n}$ meets prediction performance requirement.
    Gradient-based updates can be implemented using any standard based gradient learning rules.
```

A. Dataset Description

In this section, the IEEE PHM 2012 dataset is used to validate the proposed deep adversarial learning prognostics framework with two prediction methods. The dataset was collected using a laboratory experimental platform (PRONOSTIA) that can accelerate the degradation of bearings under constant and/or variable operating conditions, and collect online health monitoring data (including vibration, temperature, rotating speed, and load force). This experimental platform can complete bearing degradation within a few hours to provide real experimental data and characterize the degradation of the ball bearing throughout its useful life. A view of the equipment is shown in Fig. 6. For vibration signals (horizontal and

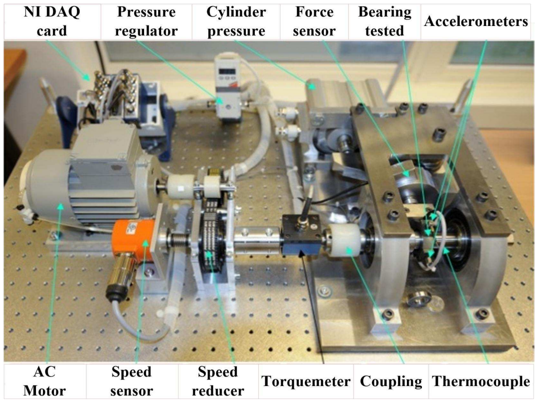


Fig. 6. The PRONOSTIA experimental platform [39].

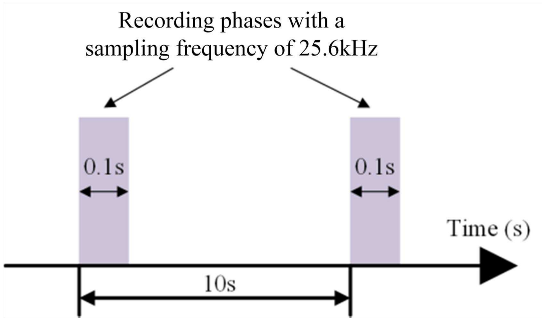


Fig. 7. An illustration of acquisition parameters for vibration signals.

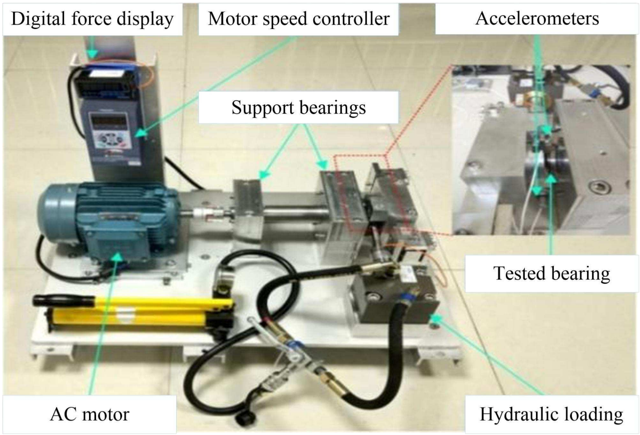


Fig. 8. Test platform of the accelerated degradation of bearing [8].

vertical), the sampling frequency is 25.6 kHz, so 25,600 samples are recorded per second. In this study, a total of 1560 samples are considered (see Fig. 7). These samples contain information on three working conditions: 1800 rpm and 4000 N, 1650 rpm and 4200 N, and 1500 rpm and 5000 N. More details are shown in Table I. Another dataset was collected from the measured equipment as shown in Fig. 8. The sampling frequency is the same as that of the IEEE PHM 2012 dataset. In this study, a total of 32,768 samples (i.e. 1.28 s) are used to verify the performance of the iterative multi-step ahead RUL prediction model. Three working conditions are shown in Table II.

TABLE I
DATASETS OF IEEE 2012 PHM PROGNOSTIC CHALLENGE

Datasets	Operating Conditions		
	Condition 1	Condition 2	Condition 3
Learning set	Bearing 1-1 Bearing 1-2	Bearing 2-1 Bearing 2-2	Bearing 3-1 Bearing 3-2
Test set	Bearing 1-3 Bearing 1-4 Bearing 1-5 Bearing 1-6 Bearing 1-7	Bearing 2-3 Bearing 2-4 Bearing 2-5 Bearing 2-6 Bearing 2-7	Bearing 3-3

TABLE II
BEARING TESTS UNDER DIFFERENT WORKING CONDITIONS

Operating Conditions	Radial force (N)	Rotating speed (rpm)	Bearing dataset	
Condition 1	1200	2100	Bearing 1-1 Bearing 1-3 Bearing 1-5	Bearing 1-2 Bearing 1-4
Condition 2	1100	2250	Bearing 2-1 Bearing 2-3 Bearing 2-5	Bearing 2-2 Bearing 2-4
Condition 3	1000	2400	Bearing 3-1 Bearing 3-3 Bearing 3-5	Bearing 3-2 Bearing 3-4

B. Implementation Details

In this paper, the bearing degradation data and the IEEE PHM 2012 PRONOSTIA dataset are considered to test the performance of the proposed adversarial learning scheme (i.e. LSTM-GAN), where multiple RUL prediction tasks are performed. These tasks include both one-step prediction and iterative multi-step ahead RUL prediction. The structure parameters of LSTM-GAN under two situations are given in Table III. For comparison purposes, six state-of-the-art RUL prediction methods are used in the experiments. For one-step time-series prediction, both conventional machine learning and deep learning methods, including SVR [18], radical basis function neural network (RBF) [40], and LSTM [33] are employed. For iterative multi-step ahead RUL prediction, the back-propagation neural network (BP) is used as a reference method. Meanwhile, two DBN-based methods, namely, DBN+SVR and DBN+BP are also used for verification. More specifically, the experimental settings and details of these methods are summarized as follows.

TABLE III
STRUCTURE PARAMETERS OF LSTM-GAN

Bearing degradation data		IEEE PHM 2012 PRONOSTIA dataset	
Parameter	Value	Parameter	Value
LSTM structure parameters			
Epochs	10000	*Epochs*	20000
Dropout μ	0.1	*Dropout* μ	0.2
Inputs size	20	*Inputs size*	20
Hidden layer	2	*Hidden layer*	2
AE structure parameters			
Epochs	10000	*Epochs*	22000
Learning rate ε	1×10^{-3}	*Learning rate* ε	5×10^{-4}
Hidden layer	1	*Hidden layer*	1

1) For SVR, radical basis function was used and the experiments were carried out using LIBSVM. For SVM, the running mode is set to e-SVR, meaning that the model optimization goal is to achieve a minimum value of root mean square error (RMSE).
2) For RBF, the activation function of the hidden layer is Gaussian function, where the number of hidden layers is 1, the numbers of input and output are 20 and 1, respectively.
3) For BP, the number of hidden layers is 1, the number of hidden nodes is 140, and the learning rate is 0.01.
4) For LSTM, the network has a double-layer, and each hidden layer has 50 nodes. The activation function is ReLU, the loss function is defined as the mean squared error (MSE), and the training algorithm is SGD.
5) For DBN+SVR and DBN+BP, the DBN has a four-layer structure, where the two hyper-parameters are learning rate and weight decay, which are set to 0.01 and 0.0008, respectively.
6) For LSTM-GAN, the details about the parameters are shown in Table III. LSTM network is used as a generator, and AE+softmax is used as a discriminator for adversarial training. The LSTM network has a double-layer structure and it is consistent with the above method 4); the input size of AE is the predicted RUL points of the generator.

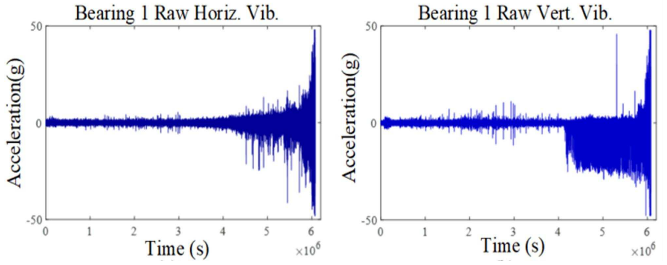
The experiments for the above four methods are implemented in MATLAB 2016, while the last two methods are implemented with PyCharm and Keras. The main purpose of parameter optimization is to select the best model parameters based on the analysis of regression prediction results. For all the test methods, the model parameters are optimized as follows. Initially, the model parameters are roughly determined. The roughly determined parameters are then refined in the reduced area. Finally, the regression prediction is used to analyze the best parameters for network training.

C. Case 1: One-Step Time-Series Prediction

In this stage, the IEEE PHM 2012 dataset is used to test the proposed method (named LSTM-GAN) with one-step time-series prediction method. Five rolling bearing degradation process indicated as bearing 1, bearing 2, bearing 3, bearing 4, and bearing 5 under working condition 1 as test targets, and the row horizontal and vertical vibration of bearing is shown in Fig. 9. The row horizontal vibration, as the prediction target, is preprocessed and converted to root mean square curve (RMS), and the RMS curves from bearings 1-5 are shown in Fig. 10. An LSTM-GAN model, consisting of the LSTM and AE networks, is proposed for the five bearing samples. The number of input layer neurons is 50, and the number of neurons in the output layer of the LSTM network is 1, which is the predicted operating health state. Therefore, the AE network is only used to distinguish the source of the features, which is simply set to one layer. In the study, the first half of the RMS curves is used for training and the second half is used to predict the degradation process. The set points with the first point to be predicted are 1800, 1000, 2250, 1800, and 1600 from bearing 1 to bearing 5 respectively. Finally, for comparison purpose, three existing methods, namely, SVR, RBF, and the proposed method without using the adversarial part (i.e. LSTM) are also applied to the same training and test data. The experimental results are depicted in Fig. 11, where it can be seen that the LSTM-GAN performs better than SVR and RBF in the average RMSE value, but its superiority over LSTM is not obvious. This is because LSTM has an outstanding ability to learn long-distance dependence, and the prediction error does not accumulate in the one-step time-series prediction mode. In addition, to make it easier to compare the predictive performance of each method on different bearings, we calculate the value of root mean square error (RMSE) based on the difference between the predicted and actual values. In particular, RMSE is computed as follows:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \tag{16}$$

where m is the number of samples included in the test dataset, y_i and $\hat{y}_i$ are the predicted value and actual value, respectively. The RMSE values of different methods are shown in the Table IV, from which it can be seen that the proposed method has the smallest average RMSE for the five bearings. However, in terms of training time, due to the deep structure of LSTM-GAN and the large number of iterations, the time complexity of the proposed method is far greater than other the machine learning methods. Therefore, this is where we need to improve in future work. For iterative multi-step ahead RUL prediction, however, LSTM-GAN shows a better performance than LSTM.



(a) (b)

Fig. 9. Whole lifetime vibration signals of bearing 1 from (a) horizontal direction and (b) vertical direction.

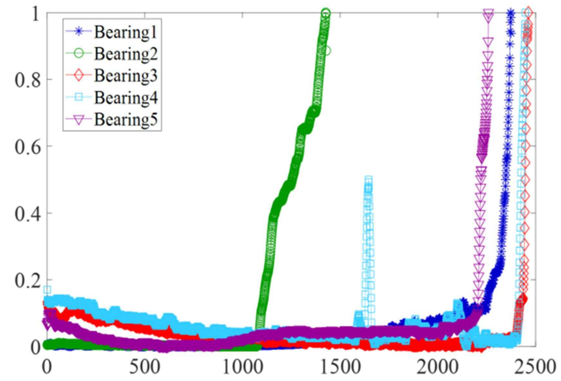


Fig. 10. The root mean square curve (RMS) of the frequency features for bearings 1-5.

TABLE IV
THE RMSE VALUES OF ONE-STEP TIME SERIES PREDICTION WITH DIFFERENT METHODS

Method Bearing	SVR	RBF	LSTM	LSTM-GAN
1	0.9606	0.0042	0.0111	**0.0037**
2	0.5406	0.7036	0.4861	**0.1427**
3	0.9895	**0.0130**	0.0170	0.0137
4	**0.0094**	0.0314	0.0383	0.0159
5	0.0278	0.0695	0.0340	**0.0104**
Average	0.5056	0.1651	0.1165	**0.0373**
Time (s)	3.4718	**2.8573**	185.9691	1765.5118

D. Case 2: Iterative multi-step ahead RUL Prediction

In this section, the bearing degradation data are used to test the LSTM-GAN with iterative multi-step ahead RUL prediction method. Normally, several statistical features need to be obtained from the experimental data to observe the overall trend of the data, such as RMS value, Kurtosis, and mean value [41]. Following [24], these statistical features are used to make RUL prediction for bearings; a graphical illustration of these statistical features are shown in Fig. 12. Note that the inclusion of these statistical features in the input vector to the network may bring more complex tasks to the LSTM-GAN model,

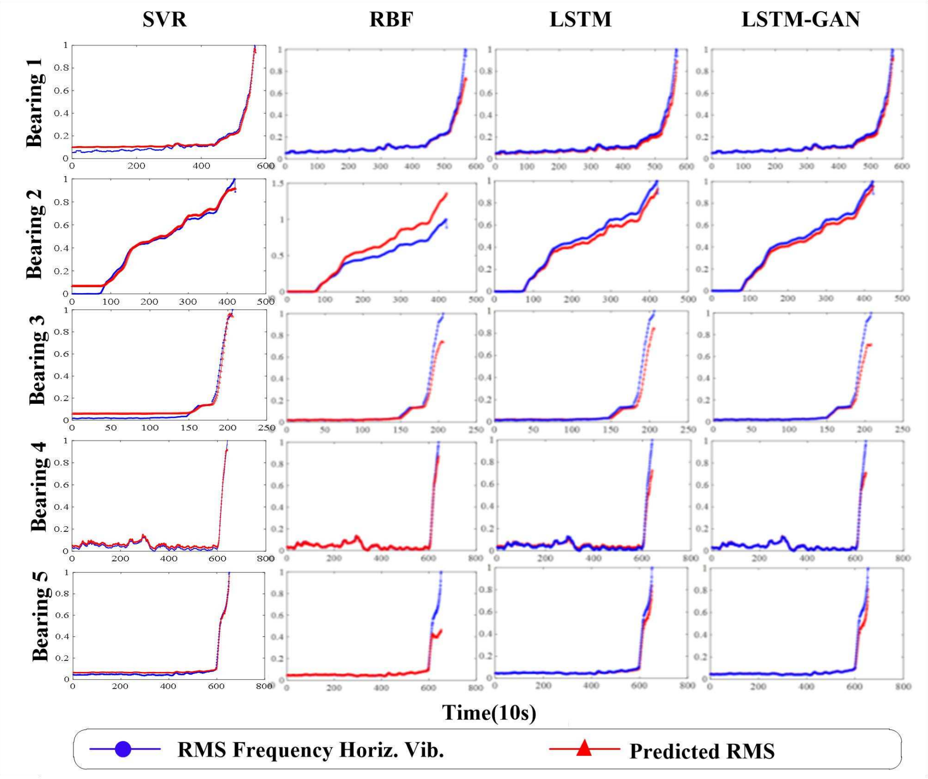


Fig. 11. One-step time-series prediction results of four the methods (SVR, SVM, LSTM, and LSTM-GAN) for bearings 1-5.

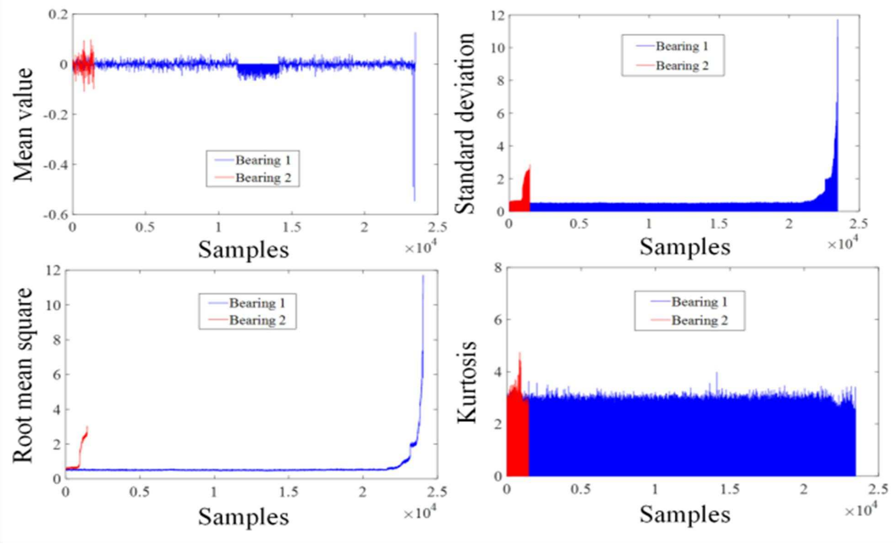


Fig. 12. An illustration of the basic statistics for Bearing 1 and Bearing 2.

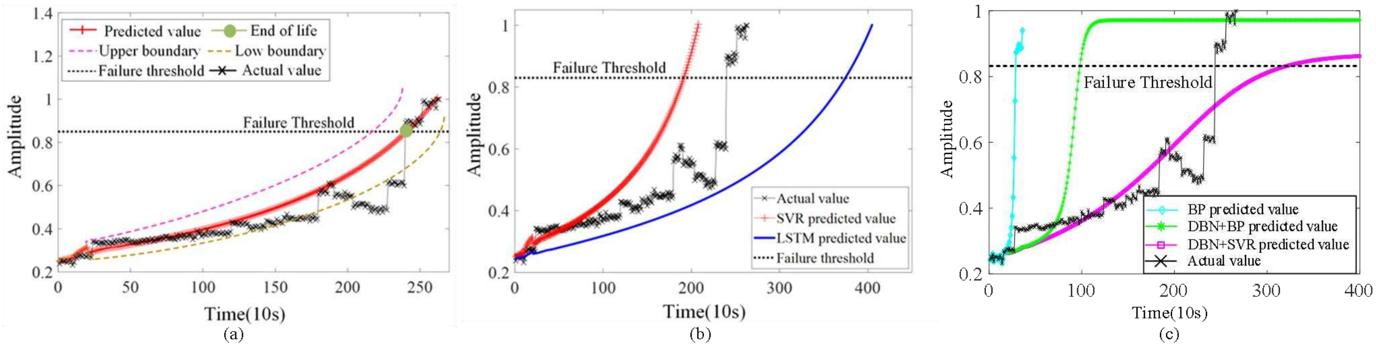
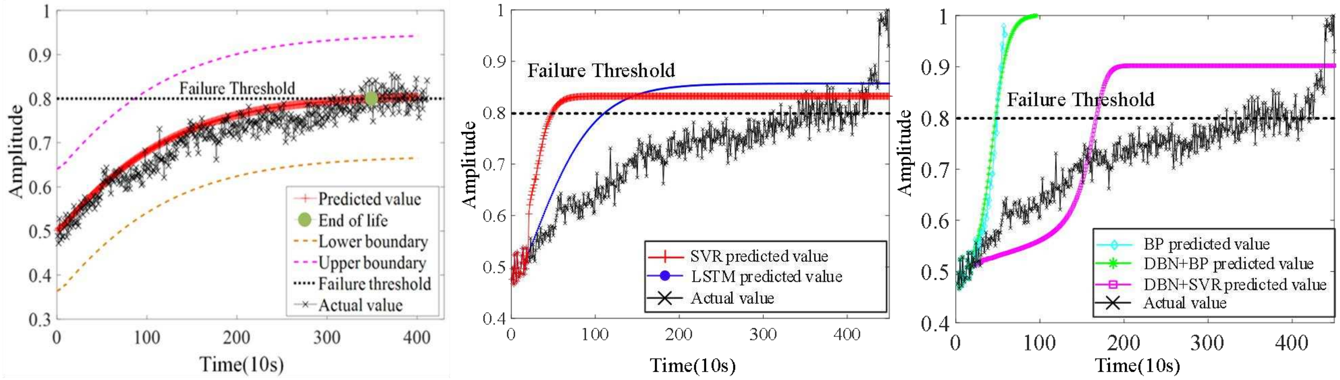


Fig. 13. Iterative multi-step ahead RUL estimation for Bearing 1 using six methods (starting at 0.25 incipient failure threshold). (a) The proposed method (LSTM-GAN). (b) SVR and LSTM. (c) BP, DBN+SVR, and DBN+BP.



(a) (b) (c)

Fig. 14. Iterative multi-step ahead RUL estimation for Bearing 2 using six methods (starting at 0.5 incipient failure threshold). (a) The proposed method (LSTM-GAN). (b) SVR and LSTM. (c) BP, DBN+SVR, and DBN+BP.

TABLE V
A Comparison of Different Methods for Iterative Multi-Step Ahead RUL Prediction

Testing dataset	Current time (s)	Actual RUL (s)	SVR RUL (s)	LSTM RUL (s)	BP RUL (s)	DBN+SVR RUL (s)	DBN+BP RUL (s)	LSTM-GAN RUL (s)
Bearing 1	237950	2390	620	1920	290	3080	980	2400
Bearing 2	10200	3310	3750	1130	420	1680	480	3500
	Score		0.0338	0.2983	0.0498	0.1163	0.0926	0.5971
	MAE		1012.5	1325	2715	1160	2120	50
	NRMSE		0.9755	0.5533	0.8863	0.4393	0.7845	0.0456
	Time (s)		1.4583	12.6650	20.4371	22.1538	278.7295	1818.7975

which is not desirable for the LSTM-GAN network. As a trade-off, it is better to infer the RUL by using a single statistical feature. Therefore, the RMS values of Bearing 3-1 (named Bearing 1) and Bearing 1-4 (named Bearing 2) are used as the statistical features. This is more difficult for the method using multi-statistical features. Two shallow LSTM-GAN networks are constructed and used to predict the RUL of the two bearings, separately. The parameters of the LSTM-GAN are updated using the training samples. Two sets of testing samples, relating to each of the operating conditions, have different incipient failure threshold (i.e. 0.5 and 0.25). Finally, for comparison purpose, five existing methods, namely, SVR, BP, DBN+SVR, DBN+BP, and LSTM, are applied to the same training and test data. Fig. 13 and Fig. 14 show a comparison of the RUL estimation results and the upper and lower bounds of the 95% confidence interval. The 95% confidence interval is calculated by:

$$x_{ci} = x \pm z\frac{s}{\sqrt{n}} \tag{17}$$

where x_{ci} is the value of the upper and lower bounds of the confidence interval, z is the standard deviation, s is the deviation of the sample, and n is the sample size.

Comparing Fig. 13(a), (b), and (c), the LSTM-GAN is significantly better than the five compared methods. Particularly, it is interesting to notice the obvious difference between LSTM-GAN and LSTM, which shows that the proposed method can effectively correct the prediction curve and reduce the propagation of prediction errors. The same observation can be noticed in Fig. 14. The LSTM-GAN approach provides much better results than the five compared methods. Note that initially the compared methods produce good prediction results, but with the continuous degradation process, these compared methods gradually deviate from the expected trajectory. Finally, all the compared methods have poor performance in the iterative multi-step ahead RUL prediction.

In order to comprehensively assess the performance of the proposed approach, the following score function is employed to calculate the accuracy of RUL prediction [39]:

$$Score = \frac{1}{2}\sum_{i=1}^{2} A_i \tag{18}$$

where

$$A_i = \begin{cases} \exp(-\ln(0.5).(Er_i/5)), Er_i \le 0 \\ \exp(+\ln(0.5).(Er_i/20)), Er_i > 0 \end{cases} \tag{19}$$

with Er_i being the percent error on experiment i, which is defined by:

$$Er_i(\%) = \frac{ActRUL_i - \widehat{RUL_i}}{ActRUL_i} \times 100 \tag{20}$$

where $ActRUL_i$ and $\widehat{RUL_i}$ represent the actual RUL and the estimated RUL for the *i*-th test bearing, respectively. In addition to the score function, the common performance metrics such as mean absolute error (MAE) and normalized root mean square error (NRMSE) are also used for further comparison. MAE and NRMSE are calculated as:

$$MAE = \frac{1}{2}\sum_{i=1}^{2}\left|ActRUL_i - \widehat{RUL}_i\right| \tag{21}$$

$$NRMSE = \frac{\sqrt{\frac{1}{2}\sum_{i=1}^{2}\left(ActRUL_{i} - \widehat{RUL}_{i}\right)^{2}}}{\left(\frac{1}{2}\sum_{i=1}^{2}\widehat{RUL}_{i}\right)} \tag{22}$$

The RUL estimation results for the two bearings are depicted in Table V. Compared with traditional SVR, LSTM, BP, DBN+SVR, and DBN+BP, the enhanced LSTM-GAN RUL prediction method generates small fluctuation in RUL prediction and small prediction error, and the LSTM-GAN gets the highest score among all the methods. Especially, for iterative multi-step ahead RUL prediction, the traditional RUL prediction methods fail since they do not compensate for the prediction error superposition problem. In contrast, the LSTM-GAN shows good performance in iterative multi-step ahead RUL prediction. The comparison results further verify the effectiveness of the LSTM-GAN for improved iterative multi-step ahead RUL prediction of bearings. Finally, the time complexity of all methods is listed in Table V. It can be observed from Table V that the time complexity of the iterative multi-step ahead prediction approach is similar to the case of one-step time series prediction (shown in Table IV). The training time used by LSTM-GAN is by far longer than the other methods. Therefore, the use of multi-GPU parallel computing to accelerate model training and improve the real-time performance of the method is key point to be investigated in our future work.

V. Conclusion

In this study, a deep adversarial learning prognostics approach based on the LSTM network and GAN was proposed for solving the prediction error superposition problem. Specifically, a LSTM-GAN structure was designed, taking advantages of both the LSTM network and GAN. One of the advantages of the proposed LSTM-GAN is that it is able to learn more representative features including the common statistic features, and therefore improves the RUL prediction accuracy. The proposed deep adversarial LSTM prognostics framework provides a reliable RUL prediction strategy in prognostics and health management of mechatronics equipment. However, the improved accuracy is achieved at the expense of taking more time on training the LSTM-GAN network. Therefore, part of the future work will be to investigate and design new implementation algorithms to accelerate the training speed of the proposed network and meanwhile to further improve its overall performance for rolling bearing RUL prediction.

References

[1] S. Nandi, H. A. Toliyat, and X. Li, “Condition monitoring and fault diagnosis of electrical motors—A review,” *IEEE Trans. Energy Convers.*,vol. 20, no. 4, pp. 719-729, Dec. 2005.

[2] P. Zhang, Y. Du, T. G. Habetler, and B. Lu, “A survey of condition monitoring and protection methods for medium-voltage induction motors,” *IEEE Trans. Ind. Electron.*, vol. 47, no. 1, pp. 34-46, Jan./Feb. 2011.

[3] Y. N. Qian, R. Q. Yan, and S. J. Hu, “Bearing degradation evaluation using recurrence quantification analysis and Kalman filter,” *IEEE Trans. Instrum. Meas.*, vol. 63, no. 11, pp. 2599-2610, Nov. 2014.

[4] F. P. G. Márquez, A. M. Tobias, J. M. P. Pérez, and M. Papaelias, “Condition monitoring of wind turbines: Techniques and methods,” *Renewable Energy*, vol. 46, pp. 169-178, 2012.

[5] A. Malhi, R. Q. Yan, and R. X. Gao, “Prognosis of defect propagation based on recurrent neural networks,” *IEEE Trans. Instrum. Meas.*, vol. 60, no. 3, pp. 703-711, Mar. 2011.

[6] R. K. Singleton, E. G. Strangas, and S. Aviyente, “The use of bearing currents and vibrations in lifetime estimation of bearings,” *IEEE Trans. Ind. Inform.*, vol. 13, no. 3, pp. 1301-1309, Jun. 2017.

[7] N. P. Li, Y. G. Lei, J. Lin, and S. X. Ding, “An improved exponential model for predicting remaining useful life of rolling element bearings,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7762-7773, Dec. 2015.

[8] Y. G. Le. (2019). XJTU-SY Bearing Datasets. [Online]. Available: https://www.mediafire.com/folder/m3sij67rizpb4/XJTU#0m0g59yk1pmh5.

[9] D. T. Liu, J. B. Zhou, H. T. Liao, Y. Peng, and X. Y. Peng, “A health indicator extraction and optimization framework for lithium-ion battery degradation modeling and prognostics,” *IEEE Trans. Syst., Man, Cybern.: Syst.*, vo. 45, no. 6, pp. 915-928, Jun. 2015.

[10] Z. G. Tian and H. T. Liao, “Condition based maintenance optimization for multi-component systems using proportional hazards model,” *Rel. Eng. Syst. Safety*, vol. 96, no. 5, pp. 581-589, May 2011.

[11] H. Tischmacher, I. Tsoumas, and S. Gattermann, “Probability model for discharge activities in bearings of converter-fed electric motors,” in *Proc. Int. Conf. Elect. Mach.*, 2014, pp. 1818-1824.

[12] J. Deutsch and D. He, “Using deep learning-based approach to predict remaining useful life of rotating components,” *IEEE Trans. Syst., Man, Cybern.: Syst.*, vo. 48 , no. 1, pp. 11-20, Jan. 2018.

[13] K. Javed, R. Gouriveau, and N. Zerhouni, “A new multivariate approach for prognostics based on extreme learning machine and fuzzy clustering,” *IEEE Trans. Cybern.*, vol. 45, no. 12, pp. 2626-2639, Dec. 2015.

[14] F. D. Maio, K. L. Tsui, and E. Zio, “Combining relevance vector machines and exponential regression for bearing residual life estimation,” *Mech. Syst. Signal Process.*, vol. 31, pp. 405-427, Aug. 2012.

[15] J. Zhu, N. Chen, and W. W. Peng, “Estimation of bearing remaining useful life based on multiscale convolutional neural network,” *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3208-3216, Apr. 2019.

[16] C. Zhang, P. Lim, A. K. Qin, and K. C. Tan, “Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 10, pp. 2306-2318, Oct. 2017.

[17] L. Guo, N. P. Li, F. Jia, Y. G. Lei, and J. Lin, “A recurrent neural network based health indicator for remaining useful life prediction of bearings,” *Neurocomputing*, vol. 240, pp. 98-109, May 2017.

[18] R. Khelif, B. Chebel-Morello, S. Malinowski, E. Laajili, F. Fnaiechand, and N. Zerhouni, “Direct remaining useful life estimation based on support vector regression,” *IEEE Trans. Ind. Electron.*, vol. 64, no. 3, pp. 2276-2285, Mar. 2017.

[19] A. Soualhi, K. Medjaher, and N. Zerhouni, “Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression,” *IEEE Trans. Instrum. Meas.*,vol. 64, no. 1, pp. 52-62, Jan. 2015.

[20] T. H. Loutas, D. Roulias, and G. Georgoulas, “Remaining useful life estimation in rolling bearings utilizing data-driven probabilistic e-support vectors regression,” *IEEE Trans. Rel.*, vol. 62, no. 4, pp. 821-832, Dec. 2013.

[21] C. Q. Shen, F. Hu, F. Liu, A. Zhang, and F. R. Kong, “Quantitative recognition of rolling element bearing fault through an intelligent model based on support vector regression,” in *Proc. 2013 Fourth International Conference on Intelligent Control and Information Processing*, Beijing, China, Jun. 2013, pp. 842-847.

[22] T. Benkedjouh, K. Medjaher, N. Zerhouni, and S. Rechakc, “Remaining useful life estimation based on nonlinear feature reduction and support vector regression,” *Engineering Applications of Artificial Intelligence*, vol. 26, no. 7, pp. 1751-1760, Aug. 2013.

[23] J. Gokulachandran and K. Mohandas, “Comparative study of two soft computing techniques for the prediction of remaining useful life of cutting tools,” *Journal of Intelligent Manufacturing*, vol. 26, no. 2, pp. 255-268, Apr. 2015.

[24] M. Xia, T. Li, T. X. Shu, J. F. Wan, C. W. de Silva, and Z. R. Wang, “A two-stage approach for the remaining useful life prediction of bearings

using deep neural networks," *IEEE Trans. Ind. Inform.*, vol. 15, no. 6, pp. 3703-3711, Jun. 2019.

[25] C. Sun, M. Ma, Z. B. Zhao , S. H. Tian, R. Q Yan, and X. F. Chen, "Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing," *IEEE Trans. Ind. Inform.*, vol. 15, no. 4, pp. 2416-2425, Apr. 2019.

[26] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436-444, May 2015.

[27] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Proc. NIPS*, pp. 2672-2680, Jun. 2014.

[28] E. Tzeng, J. Hoffman, K. Saenko, T. Darrell, "Adversarial discriminative domain adaptation," in *Proc. IEEE Conference CVPR*, pp. 2962-2971, July. 2017.

[29] K. Seeliger, U. Güçlü, L. Ambrogioni, Y. Güçlütürk, and M. A. J. van Gerven, "Generative adversarial networks for reconstructing natural images from brain activity," *Neuro. Image.*, vol. 181, pp. 775-785, Nov. 2018.

[30] B. Dai, D. Lin, R. Urtasun, and S. Fidler. (2017). "Towards diverse and natural image descriptions via a conditional GAN." [Online]. Available:https://arxiv.org/abs/1703.06029.

[31] Y. Zhang and S. S. Ge, "Design and analysis of a general recurrent neural network model for time-varying matrix inversion," *IEEE Trans. Neural Netw.*,vol. 16, no. 6, pp. 1477-1490, Nov. 2005.

[32] Y. N. Zhang, D. C Jiang, J. Wang, "A recurrent neural network for solving Sylvester equation with time-varying coefficients," *IEEE Trans. Neural Netw.*, vol. 13, no. 5, pp. 1053-1063, Sep. 2002.

[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, 1997.

[34] R. Thirukovalluru, S. Dixit, R. K. Sevakula, N. K. Verma, and A. Salour, "Generating feature sets for fault diagnosis using denoising stacked auto-encoder," in *Proc. IEEE Conference ICPHM*, pp. 1-7, Aug. 2016.

[35] H. D. Shao, H. K. Jiang, H. W. Zhao, and F. Wang, "A novel deep autoencoder feature learning method for rotating machineryfault diagnosis," *Mechanical Systems and Signal Processing.*, vol. 95, pp. 187-204, Oct. 2017.

[36] Z. H. Liu, B. L. Lu, H. L. Wei, X. H. Liu, and L. Chen, "Fault diagnosis for electromechanical drivetrains using a joint distribution optimal deep domain adaptation approach," *IEEE sensor J.*, vol. 19, no. 24, pp. 12261-12270, Dec. 2019.

[37] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Trans. Syst., Man, Cybern.: Syst.*, vol. 49, no. 1, pp. 136-144, Jan. 2019.

[38] C. Kandaswamy, et al., "Improving transfer learning accuracy by reusing stacked denoising autoencoders," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, San Diego, USA, Oct. 2014, pp. 1380-1387.

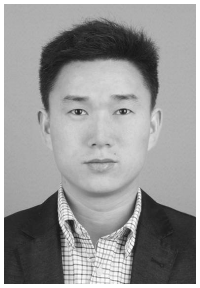
[39] P. Nectoux, et al., "Pronostia: An experimental platform for bearings accelerated degradation tests," in *Proc. IEEE Int. Conf. Prognostics Health Manage.*, 2012, pp. 1-8.

[40] X. K. Zhang, F. Q. Sun, and X. Y. Li, "A degradation interval prediction method based on RBF neural network," in *Proc. 2014 10th International Conference on Reliability, Maintainability and Safety*, Guangzhou, China, Aug. 2014, pp. 310-315.

[41] R. Zhao, D. Z Wang, R. Q Yan, K. Z. Mao, F. Shen, and J. J Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539-1548, Feb. 2018.

Bi-Liang Lu received the B.Eng. degree in electrical engineering and automation, the M.Sc. degree in automatic control and electrical engineering from the Hunan university of science and technology, Xiangtan, China, in 2017 and 2020, respectively.

His current research interests include deep learning algorithm design, and condition monitoring and fault diagnosis for electric power equipment.



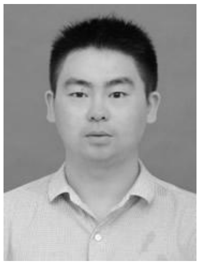
Zhao-Hua Liu (M'16) received the M.Sc. degree in computer science and engineering, and the Ph.D. degree in automatic control and electrical engineering from the Hunan University, China, in 2010 and 2012, respectively. He worked as a visiting researcher in the Department of Automatic Control and Systems Engineering at the University of Sheffield, United Kingdom, from 2015 to 2016.

He is currently an Associate Professor with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. His current research interests include artificial intelligence and machine learning algorithm design, parameter estimation and control of permanent-magnet synchronous machine drives, and condition monitoring and fault diagnosis for electric power equipment.

Dr. Liu has published a monograph in the field of Biological immune system inspired hybrid intelligent algorithm and its applications, and published more than 30 research papers in refereed journals and conferences, including IEEE TRANSACTIONS/JOURNAL/MAGAZINE. He is a regular reviewer for several international journals and conferences.

Hua-Liang Wei received the Ph.D. degree in automatic control from the University of Sheffield, Sheffield, U.K., in 2004.

He is currently a senior lecturer with the Department of Automatic Control and Systems Engineering, the University of Sheffield, Sheffield, UK. His research focuses on evolutionary algorithms, identification and modelling for complex nonlinear systems, applications and developments of signal processing, system identification and data modelling to control engineering.



Lei Chen received the M.S. degree in computer science and engineering, and the Ph.D. degree in automatic control and electrical engineering from the Hunan University, China, in 2012 and 2017, respectively.

He is currently a Lecturer with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. His current research interests include deep learning, network representation learning, information security of industrial control system and big data analysis.

Hongqiang Zhang received the B.S., M.S. and Ph.D. degrees in control science from Hunan University of Science and Technology in 2001, Hunan University (HNU) in 2004 and HNU in 2016, respectively.

He is currently a Lecturer with the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. His research interests are swarm robotics system, swarm intelligence, optimization, and intelligent control.

Xiao-Hua Li received the B.Eng. degree in computer science and technology from Hunan University of Science and Engineering, Yong zhou, China, in 2007 and the M.Sc. degree in computer science from Hunan University, Changsha, China, in 2010. Currently, She is currently a lecturer in the School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, China. Her interests are in evolutionary computation.