

Received March 2, 2020, accepted March 9, 2020, date of publication March 12, 2020, date of current version March 23, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2980244

Bayesian Long Short-Term Memory Model for Fault Early Warning of Nuclear Power Turbine

GAOJUN LIU¹, HAIXIA GU¹, XIAOCHENG SHEN², AND DONGDONG YOU¹

¹State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment, China Nuclear Power Engineering Company Ltd., Shenzhen 518172, China

²School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640, China

Corresponding author: Dongdong You (youdd@scut.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 51875209, in part by the Open Funds of State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment, and in part by the Natural Science Foundation of Guangdong Province under Grant 2017A030313320.


ABSTRACT Fault early warning of equipment in nuclear power plant can effectively reduce unplanned forced shutdown and avoid significant safety accidents. This paper presents a Bayesian Long Short-Term Memory (LSTM) neural network method for fault early warning method of nuclear power turbine. The Long Short-Term Memory neural network prediction model is developed to address data uncertainty while taking into account complicated situation of the equipment operation. Quantitative reliability validation method is established based on Bayesian inference. A wavelet packet multi-scale time-frequency analysis is employed for data denoising. A Probabilistic Principal Component Analysis (PPCA) method combined with key factor analysis is proposed for dimension reduction and dealing with the data uncertainty. The principal component inverse search method is developed to identify the critical factors mainly contributing to the turbine fault. Numerical results indicate that the proposed novel model is validated with Bayesian confidence of 92% by using the real-world steam turbine data and the model can provide accurate warning in the early creep stage of the fault.

INDEX TERMS Bayesian inference, long short-term memory, discrete wavelet packet transform, nuclear power turbine, probabilistic principal component analysis.

I. INTRODUCTION

Early warning and Remaining Useful Life (RUL) prediction of large turbo machines such as pumps and steam turbines in a nuclear plant is often carried out at the early stage of equipment failure through real-time condition monitoring, thus reducing unplanned forced shutdown and accidents [1], [2]. Ma and Jiang [1] introduced the Fault Detection and Diagnosis (FDD) method to enhance the safety, reliability and availability of a nuclear power plant. The method is applied to monitor six areas of the plant, i.e. instrument calibration, instrument channel dynamic performance, equipment, reactor core, loose part and transient identification.

In recent years, big data technology and artificial intelligence algorithms have been widely used to improve the reliability and safety of a nuclear power through real-time monitoring, damage diagnosis and prognosis, automatic

The associate editor coordinating the review of this manuscript and approving it for publication was Baoping Cai .

condition assessment and RUL prediction. These techniques include advanced signal processing and multivariate data analysis. This subject has become one of the cutting-edge hot spots in the energy field which attract numerous researchers. Condition monitoring and fault prediction of nuclear power equipment require addressing the data uncertainty from multiple sources and the complexity of modeling, which makes it very challenging. Gong *et al.* [2] presented a fault diagnosis method for the main coolant system of a nuclear power plant based on Dempster/Shافر (D-S) evidence theory. Peng *et al.* [3] presented a fault diagnosis method for nuclear power equipment by combining correlation analysis and deep belief network and compared the fault diagnostic results of the proposed method with the back-propagation neural network and support vector machine. Dong [4] proposed a Boolean network with its linear representation for describing the fault propagation among sensors. Theoretic analysis is conducted to select the diagnosis-oriented sensors for a nuclear steam supply system. Wu *et al.* [5] introduced a FDD

technology framework based on a Bayesian network (BN) for large-scale complex systems in nuclear power plants. Numerical results have demonstrated its advantages in terms of easy visualization, parameter uncertainty representation and ability of incomplete data handling.

The key to the success of fault early warning includes two factors: accurate prediction model and effective warning logics. The prediction model needs to be established and trained from a set of historical data without any known anomaly to reflect the behavior of the system under normal operation conditions. The model is validated and then used to predict the system response under unknown conditions. A decision logic is developed to judge the system operation status based on the model prediction and actual measurement of the system response. An alarm is triggered when the indicator representing the system state exceeds a predefined threshold. In the field of nuclear power, there are a few methods for fault early warning. Welz *et al.* [6] developed the early warning function in online monitoring system for nuclear power plants by using the Auto-Association Kernel Regression (AAKR) technique and present three cases that represent the functions and roles of the system. Min *et al.* [7] investigates the influence of integrating maintenance information of nuclear plant equipment on prognostic model prediction accuracy, and the developed maintenance-dependent models can greatly improve the accuracy.

Recently, various data-driven non-parametric system identification methods have been widely used in the fault prediction of rotating machinery such as bearings and gears. These methods include time series analysis, regression analysis, grey system theory or artificial neural network. For example, Li *et al.* [8] used an improved multi-scale symbolic dynamic entropy method (MMSDE) to quantify the regularity of time series, which improves the prediction accuracy and effectively diagnoses the fault of planetary gearbox. Zhen *et al.* [9] proposed a multi-objective sparse implicit regression (MSIR) method, which can simultaneously establish the internal correlation and complex nonlinear input-output relationship between targets in one framework, and can flexibly handle this highly complex relationship. Alexander *et al.* [10] developed a method to judge the accuracy of a regression model and by analyzing the shortcomings of the standard method for conventional test fitting, and applied R-2 in model fitting statistics. Ying *et al.* [11] employed a gray relation algorithm for fault pattern recognition of a rolling bearing using the extracted feature vectors. Jiang and his coauthors [12]–[14] developed a variety of non-parametric methods for system identification, damage detection and condition monitoring.

In recent years, the Artificial Neural Network (ANN) method has been developed rapidly, which is able to solve practical problems with complex uncertainty and time-varying. The ANN prediction method can capture the non-linearity in the time series. For example, Jiang *et al.* [15] proposed a damage prediction model of turbomachinery based on dynamic fuzzy Stochastic Neural Network (SNN) for non-parametric system identification, which can capture

the nonlinear and random characteristics of the system and carry out fault early warning for the collected information. The reliability of the model is verified by Bayesian hypothesis test. Martínez-Martínez *et al.* [16] established an expert system based on artificial neural network, which uses single point vibration signals of the machine to predict the state of multiple rotating parts of agricultural machinery. Ayodeji *et al.* [17] proposed a testing scheme to establish a nuclear power plant operator support system by using Principal Component Analysis (PCA) and ANN, and evaluated the prediction ability of Elman neural network and radial basis function neural network. On the bench pressurized water reactor simulator, the process is validated by using the data collected from different failure scenarios. Cabrera *et al.* [18] proposed a method of feature extraction and fault severity estimation. The deep convolution neural network is trained by the stack convolution self-encoder. Under the conditions of constant or varying speed of rotating machinery, the data sets with different fault severity are used to demonstrate the robustness and accuracy of this method. Dou and Zhou [19] compared and analyzed four direct classification methods like probabilistic neural network for intelligent fault diagnosis of rotating machine. Amare *et al.* [20] combined Auto Associative Neural Network (AANN), nested Machine Learning Classifier (MLC) and Multi-Layer Perceptron (MLP), and proposed a hybrid intelligent technology to diagnose three simultaneous failures in a double axis of the industrial gas turbine engine.

In past decade deep learning methods have been broadly applied for fault diagnosis and health management of mechanical equipment, such as Recurrent Neural Network (RNN) [21], [22] and Long-Short Term Memory (LSTM) neural network methods. LSTM approach is an evolutionary version of RNN, which effectively addresses the issue of long-term dependence of effective information in time series. In comparison to other models, the LSTM method is more sensitive to the trivial features in the historical data, easier to capture the details, suitable for big data processing, and more accurate in time series prediction. Lee *et al.* [23] proposed a Function based Hierarchical Framework (FHF) based LSTM method for autonomous operation of a nuclear power plant safety system. A small nuclear simulator was used to train and validate the LSTM network. According to the characteristics of reciprocating vibration signals, Tian *et al.* [24] proposed a hybrid prediction modeling strategy by combining the autocorrelation local characteristic-scale decomposition and the improved LSTM neural network. Zhang *et al.* [25] proposed a LSTM approach for bearing performance degradation evaluation. Numerical results show that the proposed LSTM method can effectively predict the remaining service life of the bearing. Yang *et al.* [26] presented a new method of fault detection and isolation of electromagnetic bearings. A LSTM neural network was used to build a sensor data model, which can effectively process time series data. A sliding window is employed to improve the performance of the LSTM model applied to fault isolation. Li *et al.* [27] proposed

a LSTM method for fault diagnosis and isolation of wind turbine, where stochastic forest algorithm is applied to make decision, and validated by numerical comparison with a wind turbine benchmark Simulink model.

In order to improve the data quality and facilitate the subsequent data analysis and mining as well as modeling, data preprocessing is conducted by identifying and imputing the bad data. Recently, Lee *et al.* [28] proposed a comprehensive state-of-the-art review on diagnostics and health management design for rotary machinery systems, emphasizing the key role of signal processing technology in the damage diagnosis and prediction of rotary machinery. Montalvo and García-berrocal [29] applied the Discrete Wavelet Transform (DWT) method to monitor and diagnose resistance temperature detector related to the safety of a nuclear power plant. The method was demonstrated to effectively reduce the dispersion of estimated response time, and eliminate the non-stationary features such as trend and peak in the signal. Upadhyaya *et al.* [30] enhanced the sensor measurement by using the wavelet transform to filter the low and high frequency components of nuclear power plant sensor data. Wavelet transform is used for signal conditioning to minimize signal bandwidth distortion, which provides an effective method for data preprocessing. Baraldi *et al.* [31] proposed a hybrid method to diagnose faults from similar transients by combining Haar wavelet transform, fuzzy similarity, spectral clustering and the Fuzzy C-Means algorithm, and illustrated its effectiveness by applying it to the transient process resulted from different faults of nuclear power reactor regulators.

Recently principal component analysis-based dimension reduction technique has been widely applied to speed up model convergence and improve prediction accuracy in fault detection and isolation of turbomachinery. Yong-Kuo *et al.* [32] combined PCA, directed graph G (SDG) and Elman Neural Network (ENN) methods for fault detection, fault isolation and severity estimation of nuclear power plants. Park *et al.* [33] developed a PCA-based transient monitoring system for the secondary system of nuclear power plant, where the operation data was reduced to smaller dimension. Li *et al.* [34] proposed two PCA models based on statistics and iteration for sensor fault detection of nuclear power plant. The combined model made the fault monitoring more effective and efficient. Wu *et al.* [5] proposed a PCA-based FDD framework for nuclear power plant pressurized water reactor. The fuzzy theory along with PCA was used as data fusion approach to integrate the data of multiple sensors into one node.

When the abovementioned methods are applied for fault early warning and diagnosis of large-scale nuclear power machinery, they need to be improved to address the following issues: 1) incomplete data, 2) impact of multivariate correlation, 3) nonlinear and stochastic behavior of nuclear power rotating machinery, and 4) quantitative evaluation of prediction model. This study attempts to propose an intelligent probability method to establish an accurate predictive model

for fault identification of nuclear power machinery to deal with the above four issues.

This paper ingeniously combined advanced signal processing, pattern recognition, intelligent algorithm and probabilistic decision-making methods, and proposed a Bayesian LSTM algorithm for intelligent fault early-warning of nuclear power machinery. The wavelet packet threshold denoising and PPCA are used to preprocess the raw data collected from the nuclear plant. The principal component inverse search method is developed to quickly find out the abnormal signal. The LSTM model is proposed to predict the system response of the nuclear plant. A Bayesian framework is developed to quantitatively validate the predictive model. The early warning logic is developed to identify fault creep stage and monitoring system fault by making full use of the prior information of the historical data.

II. DATA PROCESSING AND DATA INTEGRATION TRANSFORMATION

A. WAVELET PACKET DENOISING

Wavelet packet analysis is employed to remove the noise from the raw data collected from the turbomachine system. Compared with discrete wavelet transform, which only transforms the low-pass filtering results, wavelet packet transform is more precise, which decomposes both low-pass and high-frequency parts, as shown in Fig. 1. After a noisy signal is decomposed into multiple layers, its energy is mainly concentrated in the partial wavelet packet decomposition coefficients. Whereas the noise energy is distributed in the coefficients of the whole wavelet domain, and the amplitude of the wavelet packet transformation coefficients of the signal itself is larger than that of the noise [35]. Thus, the wavelet packet transform coefficients can be split out by setting an appropriate threshold to filter out the noise part in the decomposed coefficients. Accordingly, wavelet reconstruction is carried out to obtain the cleaned signal.

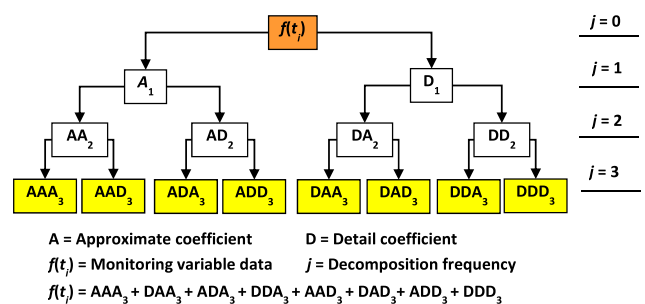


FIGURE 1. Schematic diagram of three-level wavelet decomposition.

The threshold plays a key role on the denoising effect of wavelet packet transform method. The traditional wavelet denoise method utilizes general, sure or GCV threshold. None takes into consideration the prior information of the original signal wavelet coefficient. This study employs Bayesian inference to determine the threshold, which takes the advantage of prior information.

In Discrete Wavelet Packet Transform (DWPT), a signal $f(t)$ with n discrete data points can be represented as $f(t) = \sum_{j=-\infty}^{+\infty} f_j(t)$ by wavelet decomposition, where $f_j(t)$ is the coefficient series at j level, and j is the position index representing the scale and also the level to be further decomposed. Therefore, the wavelet coefficients $f_j(t)$ can be further decomposed into multiresolution wavelet components:

$$f_j(t) = \sum_{m=0}^{2^k-1} f_{j-k}^m(t) = \sum_{m=0}^{2^k-1} \sum_{l \in Z} d_l^{j-k,n} w_n(2^{j-k}t - l), \quad n=2^k+m \quad (1)$$

where k represents the number of successive decomposition levels, $0 \leq k \leq j$. Generally, the number of levels is 3-5. The coefficient $\{d_l^{j-k,n}\}$ is called the orthogonal wavelet packet decomposition coefficient with the resolution ratio of $j-k$, l is the corresponding coefficient length, and $w_n(t)$ is the wavelet function [36]. The wavelet functions commonly used for denoising are dBn (Daubechies) wavelet, symN (Symlet) wavelet and coifN (Coiflet) wavelet. The next level decomposition coefficient $\{d_l^{j,2n}\}$ and $\{d_l^{j,2n+1}\}$ of $\{d_l^{j+1,n}\}$ can be obtained by the following iterative formula:

$$\begin{cases} d_l^{j,2n} = \sum_{k \in Z} h_{k-2l} d_k^{j+1,n} \\ d_l^{j,2n+1} = \sum_{k \in Z} g_{k-2l} d_k^{j+1,n} \end{cases} \quad (2)$$

The initial value of Eq. (2) is the discrete data itself. After the wavelet coefficients are cleaned by thresholding at each decomposition level, a cleansed signal can be reconstructed by:

$$d_l^{j+1,n} = \sum_{k \in Z} [h_{l-2k} d_k^{j,2n} + g_{l-2k} d_k^{j,2n+1}] \quad (3)$$

The Bayesian shrinking threshold estimation method [37] is utilized to determine the threshold value. The calculation formula is $T(j, n) = \sigma^2/\sigma_x$, where σ^2 is the variance of contaminated signals, and σ_x is the variance of the original signal. By using the soft threshold function, the wavelet coefficients can be filtered:

$$d_l^{j,n} = \begin{cases} \text{sgn}(d_l^{j,n})(|d_l^{j,n}| - T(j, n)), & |d_l^{j,n}| \geq T(j, n) \\ 0, & |d_l^{j,n}| \leq T(j, n) \end{cases} \quad (4)$$

B. BAYES PROBABILISTIC PRINCIPAL COMPONENT ANALYSIS

For a pump or steam turbine in a nuclear plant under the condition monitoring, there are hundreds of parameters to be monitored such as rotating speed, vibration, temperature, eccentricity, and axis displacement. In this study Bayesian PPCA [38] is employed for dimension reduction and uncertainty processing of the data prior to model prediction.

Different from traditional PCA the PPCA keeps part of the discarded information through Gaussian noise thus retaining more useful information from the original signal in the principal components.

In PPCA the data sample $\mathbf{X}_{d \times n}$ with d variables and n points is expressed as $\mathbf{X} = \mathbf{W}\mathbf{z} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}$, where \mathbf{W} is weight vector with the dimension of $d \times q$, ($q \leq d$); \mathbf{z} is the $q \times n$ reduced principal components of \mathbf{X} in which $\mathbf{z} \sim N(0, \mathbf{I}_q)$ is the random Gaussian vector; $\boldsymbol{\mu}$ is the sample mean, and $\boldsymbol{\varepsilon}$ is the noise, which is assumed to obey the Gaussian distribution with variance σ^2 . By using Bayes formula, the posterior probability $P(\mathbf{z}|\mathbf{X}) \sim N(\mathbf{M}^{-1}\mathbf{W}^T(\mathbf{X} - \boldsymbol{\mu}), \sigma^2\mathbf{M}^{-1})$ of \mathbf{z} can be obtained, where $\mathbf{M} = (\sigma^2\mathbf{I}_q + \mathbf{W}^T\mathbf{W})^{-1}$ [38]. Thus, the expected $\mathbf{M}^{-1}\mathbf{W}^T(\mathbf{X} - \boldsymbol{\mu})$ of the posterior probability of \mathbf{z} is treated as of the reduced principals for \mathbf{X} . The unknown parameters \mathbf{W} and σ^2 are estimated by the maximum likelihood function as follows:

$$\widehat{\mathbf{W}} = \mathbf{U}_q (\boldsymbol{\Lambda}_q - \sigma^2\mathbf{I}_q)^{1/2} \mathbf{R} \quad (5)$$

$$\widehat{\sigma}^2 = \frac{1}{d-q} \sum_{j=q+1}^d \lambda_j \quad (6)$$

where λ_j is obtained by decomposing the covariance matrix of sample \mathbf{X} according to the eigenvalue, that is, $\mathbf{C}v_j = \lambda_j v_j$, where v_j is the eigenvector with $\mathbf{U}_q = (v_1, v_2, \dots, v_q)$, $\boldsymbol{\Lambda}_q = \text{diag}(\lambda_1, \dots, \lambda_q)$. Then the dimension reduced variable \mathbf{z} can be expressed as:

$$\mathbf{z} = \mathbf{M}^{-1}\widehat{\mathbf{W}}(\mathbf{X} - \boldsymbol{\mu}) \quad (7)$$

Notes that when $q = d$, $\mathbf{W}^{-1}\mathbf{X} = \mathbf{z} + \mathbf{W}^{-1}\boldsymbol{\mu} + \mathbf{W}^{-1}\boldsymbol{\varepsilon}$. The principal component inverse search is carried out for preliminary fault diagnosis by using \mathbf{W}^{-1} . For example, \mathbf{W}^{-1} can be expressed as:

$$\mathbf{W}^{-1} = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1d} \\ w_{21} & w_{22} & \cdots & w_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ w_{d1} & w_{d2} & \cdots & w_{dd} \end{bmatrix} \quad (8)$$

where w_{12} is the contribution rate of the second dimension signal in the data sample $\mathbf{X}_{d \times n}$ to the first principal component. Thus, the contribution of all signals to the first principal component can be obtained. When the principal component is abnormal, the signals with high contribution rate are identified to fail. Therefore, the PPCA is used in this study to identify the anomaly in the signals effectively.

III. LSTM PREDICTION MODEL

A. INPUT AND OUTPUT LAYER DETERMINATION

The principal component obtained from the Bayesian PPCA method is likely to be chaotic time series. The prediction of chaotic time series has become an important research topic in the field of signal processing. Takens [39] proposed a delay phase space reconstruction and embedding theorem for nonlinear time series. In the reconstructed phase space, ANN which has strong nonlinear mapping ability is usually

applied to approximate the mapping relationship between the present and the future states to predict the chaotic time series. In this study, the input and output of the neural network are determined by the reconstructed embedding dimension m and the delay time τ .

Kim *et al.* [40] proposed the correlation integral Cross-Correlation (C-C) method to estimate the delay time τ and embedding dimension m simultaneously, and considered that they are related to the delay window τ_w , expressed as $\tau_w = (m - 1)\tau$. The original signal $\{x_1, x_2, x_3, \dots\}$ can be constructed as a matrix of disjoint time series:

$$\mathbf{X} = \begin{bmatrix} x(1) & x(1 + \tau) & x(1 + 2\tau) & \cdots & x(1 + (m - 1)\tau) \\ x(2) & x(2 + \tau) & x(2 + 2\tau) & \cdots & x(2 + (m - 1)\tau) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x(\tau) & x(\tau + \tau) & x(\tau + 2\tau) & \cdots & x(\tau + (m - 1)\tau) \end{bmatrix} \quad (9)$$

Each row in the time series matrix is a time series vector. The element of each row in the matrix is used as the input node of the neural network, and the next τ time of the last element in the row is used as the output node, thus yielding the input and output layers. Furthermore, with the embedding dimension and delay time, Lyapunov exponent may be used to judge whether the signal is a chaotic time series. If Lyapunov exponent is regular, the phase space reconstruction is needed, otherwise $\tau = 1$, and m can be determined by enumeration method during the neural network training.

B. ESTABLISHING LSTM MODEL

Different from the traditional RNN, LSTM is able to solve the issues of gradient disappearance and gradient explosion in the training process, yielding more accurate prediction of a long-term time series. A LSTM model is built to predict the response value of the system under investigation. A single LSTM unit is shown in Fig. 2, where x_t is the response value at time t , \mathbf{h}_{t-1} is the transferred output of the last state, and \mathbf{c}_{t-1} is the historical information output of the last state. The LSTM method splits x_t and \mathbf{h}_{t-1} into a vector composition

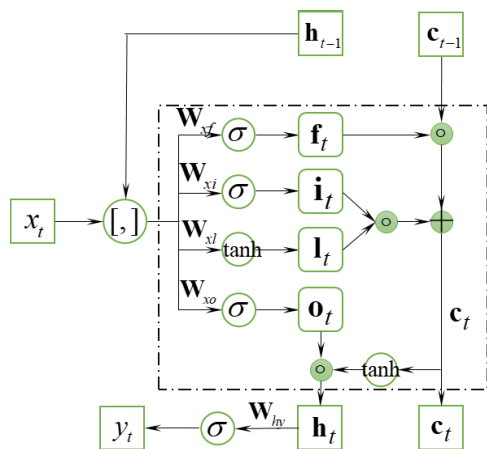


FIGURE 2. LSTM unit structure.

output, then multiplies four different weight matrices and adds bias. Three of them are converted into values from 0 to 1 by *sigmoid* function to obtain \mathbf{f}_t , \mathbf{i}_t and \mathbf{o}_t which corresponds to forgetting gate, input gate and output gate, respectively. The other is converted into values from -1 to 1 by *tanh* function to get \mathbf{l}_t . The output \mathbf{h}_{t-1} , \mathbf{c}_{t-1} , y_t of the current state are obtained by the Eq. (10) [41]:

$$\begin{cases} \mathbf{c}_t = \mathbf{i}_t \odot \mathbf{l}_t \oplus \mathbf{c}_{t-1} \odot \mathbf{f}_t \\ \mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t) \\ y_t = \sigma(\mathbf{w}_{hy}\mathbf{h}_t) \end{cases} \quad (10)$$

where, \odot represents the multiplication of corresponding elements of matrix and \oplus represents the addition of corresponding elements, so two matrices are required to be of the same type.

To predict the time series, multiple units are connected and the output is given in the last unit. The number of units is determined by the embedding dimension m . The LSTM prediction model corresponding to Eq. (9) is shown in Figure 3. The initial input \mathbf{h}^0 and \mathbf{c}^0 of the first unit of each group are random values, and the last unit of each group outputs the prediction value \hat{x} , which corresponds to the next τ time point of the input time point of the last unit in the group.

The back-propagation algorithm is employed to update the weights. For the LSTM prediction model, four groups of weight matrices need to learn, which are the weight matrix \mathbf{W}_f of forgetting gate, the weight matrix \mathbf{W}_i of input gate, the weight matrix \mathbf{W}_o of output gate, and the weight matrix \mathbf{W}_l of calculating unit state and their respective offsets term. At time t , the output value of LSTM is \mathbf{h}_t . The error term $\delta_{t-\tau}$ of time $t - \tau$ can be expressed as Eq. (11) by the error of time t , and E is the cross-entropy loss function expressed as $E = -[x \ln \hat{x} + (1 - x) \ln(1 - \hat{x})]$. For the derivation of $\partial \mathbf{h}_t / \partial \mathbf{h}_{t-\tau}$, i.e., the parameter training, please refer to [42].

$$\delta_{t-\tau} = \frac{\partial E}{\partial \mathbf{h}_{t-\tau}} = \delta_t \frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_{t-\tau}} \quad (11)$$

C. MODEL VALIDATION

The data set is divided into three parts: training, validation and testing set. The training data set is used to build the model. The validation set is used to evaluate the accuracy of the model prediction. Most commonly used model validation metrics include Mean Square Error (MSE) and decision coefficient R^2 . MSE explains the degree of deviation between the model prediction and the actual data. The decision coefficient R^2 indicates the superiority of the model compared with the direct mean value. The R^2 value falls in between 0-1, with one indicating an ideal model. However, these two metrics couldn't take into account the uncertainties in the actual measurement and model prediction. A Bayesian hypothesis test method [43]–[45] is developed to address the issues of the abovementioned two metrics. The Bayesian approach accounts for the data uncertainties, quantifies confidence on the model reliability as well as considers the prior information of the training set.

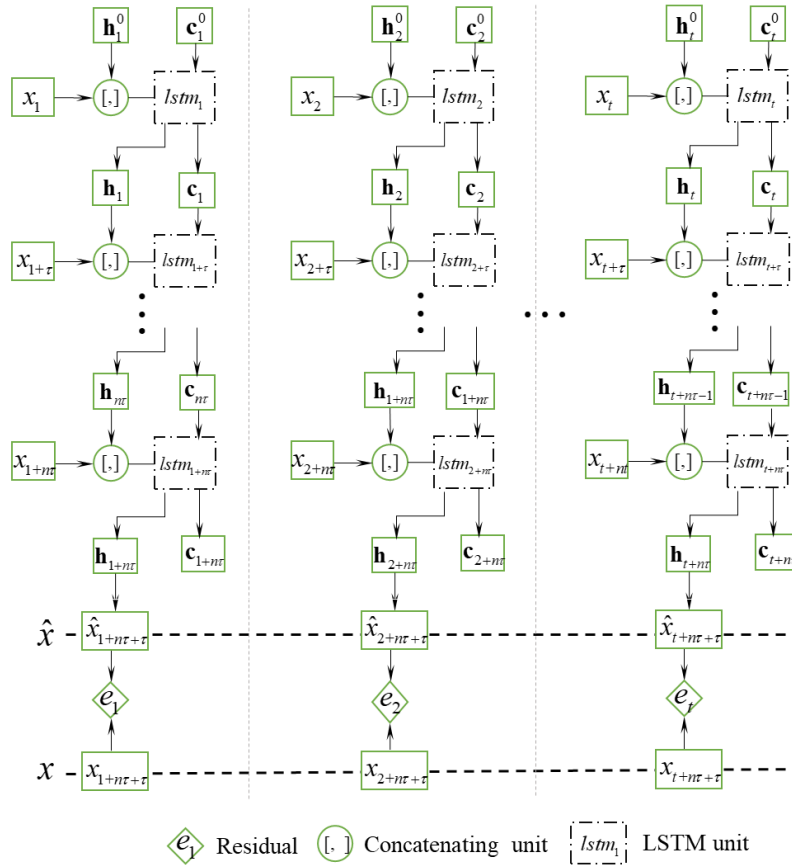


FIGURE 3. LSTM prediction model after phase space reconstruction of time series.

It is assumed that the model prediction is represented by $y_{pred} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$, and the corresponding actual data is $y_{exp} = \{y_1, y_2, \dots, y_n\}$, where n is the number of data. Let $e_i = y_i - \hat{y}_i$ represent the residual of the i -th actual signal and the i -th output signal, we have the n residuals $\varepsilon = \{e_1, e_2, \dots, e_n\}$. Generally, the residual variable e_i is assumed to follow the normal distribution $N(\mu, \sigma_1^2)$, and thus the mean value $\bar{\varepsilon}$ of residuals would follow the normal distribution $N(\mu, \sigma_1^2/n)$. The original hypothesis H_0 and alternative hypothesis H_1 is established:

$$H_0 : \mu = 0, H_1 : \mu \neq 0$$

Assuming that the prior probability density of μ is $N(0, \sigma_0^2)$, the training set is usually used as the prior information for the selection of σ_0^2 . In other words, the mean value of the prediction error of the training set is calculated by sections with a certain size of window, and the variance of the mean value is taken as σ_0^2 . By combining with the regularity of normal distribution, the edge density function of \bar{e}_i to $g(\mu)$ is [14]:

$$\begin{aligned}
 m(\bar{x}) &= \int_{-\infty}^{\infty} p(\bar{\varepsilon}|\mu)g(\mu)d\mu \\
 &= \frac{1}{\sqrt{2\pi(\sigma_0^2 + \sigma_1^2/n)}} \exp\left\{-\frac{\bar{\varepsilon}^2}{2(\sigma_0^2 + \sigma_1^2/n)}\right\} \quad (12)
 \end{aligned}$$

The Bayesian factor can be expressed as:

$$\begin{aligned}
 B^\pi(\bar{\varepsilon}) &= \frac{p(\bar{\varepsilon}|\mu)}{m(\bar{\varepsilon})} \\
 &= \sqrt{1 + \frac{n\sigma_0^2}{\sigma_1^2}} \exp\left\{\frac{n\bar{\varepsilon}^2}{2}\left(\frac{1}{\sigma_1^2} + \frac{1}{n\sigma_0^2 + \sigma_1^2}\right)\right\} \quad (13)
 \end{aligned}$$

By using Eq. (13), a posterior probability can be obtained:

$$\lambda = \pi(\mu|\bar{\varepsilon}) = \left[1 + \frac{1 - \pi_0}{\pi_0} \frac{1}{B^\pi(\bar{\varepsilon})}\right]^{-1} \quad (14)$$

where λ is the confidence on the reliability of the prediction model. When $B^\pi(\bar{\varepsilon}) \rightarrow 0$, $\lambda \rightarrow 0$, which indicates the zero confidence of supporting the model, that is, the model is invalid. Conversely, when $B^\pi(\bar{\varepsilon}) \rightarrow \infty$, $\lambda \rightarrow \infty$, which indicates the full confidence of supporting the model, that is, the model is reliable.

D. FAULT EARLY WARNING

After the model is validated to provide accurate prediction, the real-time prediction results are next used to evaluate the health status of the system under monitoring. As shown in Fig. 4, when the component in a machinery or monitoring system has fault at a certain time, the collected signals will demonstrate anomaly trend or feature, which can be detected by the monitoring system. The anomaly is usually represented

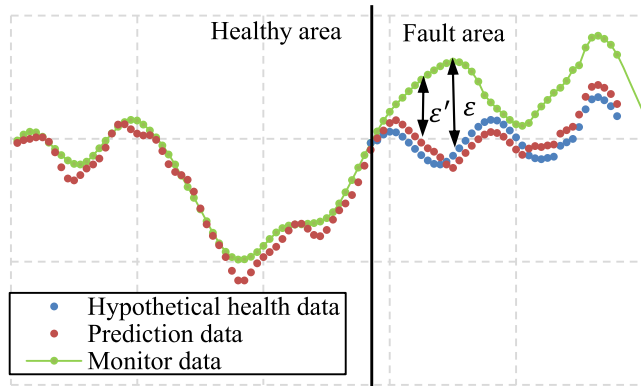


FIGURE 4. Error between fault signal, Hypothetical health signal and prediction signal.

by a large error ε in the signals under the hypothetical healthy state. Before the failure time, the predictive model established from the historical health data is used to produce the healthy response of the system. There will be a larger gap ε' between the model prediction and actual data during the monitoring to indicate the anomaly in the system, as shown in the figure 4. An early warning alarm can be triggered by identifying this kind of large deviation in the monitoring and predefining a threshold.

In this study a threshold value ζ is predefined as the maximum absolute value e_{max} of the error between the model prediction and the actual data of the validation set. The validation data set is the historical time series collected under the healthy condition of the turbomachine unit. If the model prediction deviation exceeds the maximum error of the validation for a certain period during real-time condition monitoring, the unit can be judged to have potential fault in some components which need to be inspected.

IV. ILLUSTRATIVE EXAMPLE

A. DESCRIPTION OF DATA SET

In this study, one-month operation data collected in the hour interval from a high and intermediate pressure (HIP) cylinder of a steam turbine in a real-world nuclear power plant is employed to illustrate the effectiveness and feasibility of the proposed methodology. Steam turbine is a large and high-fidelity high-speed rotating machine. In this example, the proposed methodology is demonstrated to provide early warning for the turbine during the creep period of steam turbine failure.

There are 29 critical factors of a steam turbine investigated in this example, including vibration, differential expansion (difference of expansion between cylinder and rotor), rotor shaft displacement, temperature and rotation speed, as defined in Table 1, and 720 data points for each factor. These factors are used to represent the overall health status of the steam turbine. The monitoring positions are shown in Fig. 5. The nominal maximum power of the steam turbine is 1118 MW, the rated rotation speed is 1500 rpm, the steam pressure prior to the high-pressure stop valve is 6.4 MPa,

TABLE 1. Description of sampling signal.

Single type	Quantity	Monitoring object
Vibration	8	absolute and relative vibration in horizontal and vertical directions of HP and IP cylinder
Differential expansion	1	difference of expansion between cylinder and rotor
Temperature	15	8 positions of HP cylinder and 7 positions of IP cylinder
Rotation speed	4	1 rotor rotation speed, 3 rotation speeds of bearing groups
Displacement	1	rotor shaft
Total	29	

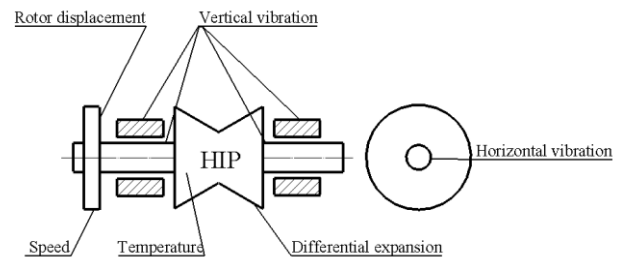


FIGURE 5. Schematic diagram of monitoring points.

TABLE 2. Sample data of HIP cylinder.

Cylinder		Pressure (MPa)	Temperature (°C)
High pressure	Air intake	6.11	277
	exhaust	0.97	179
Intermediate pressure	Air intake	0.94	269
	exhaust	0.3	153

and the steam temperature is 280 °C. Table 2 shows the sampling data and the factor description of the HIP cylinder. The sample period is from 0 a.m. on April 1, 2019 to 23 p.m. on April 30, 2019. In this example, the data before April 18 is used for model training, the data from April 19 to April 24 is used for model validation, and the data after April 24 is used for model testing.

B. DATA ANALYSIS

Data imputation is first conducted to improve the data quality. Fig. 6 shows the comparison of the raw data and processed data which demonstrates the defect of the speed signal for the rotor shaft. The training and validation data sets are imputed by the regression interpolation method, with the purpose of building an accurate predictive model. The two data points before and after the missing value are employed to impute the missing one via the interpolation and the bad data in the testing set will be used as the diagnostic object.

The wavelet threshold denoising method is applied for vibration signals with different noise degrees. The three-level discrete wavelet packet analysis with db8 function is

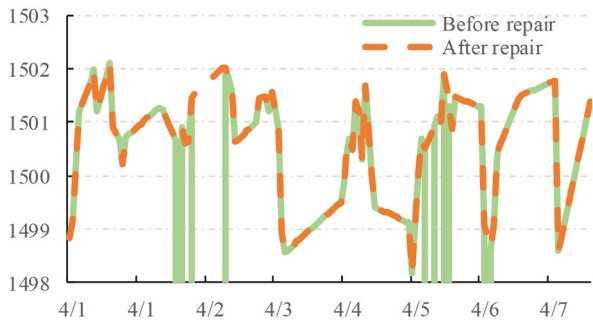


FIGURE 6. Comparison of raw and processed data for rotor shaft speed signal.

conducted for the time series signal to obtain 8 wavelet coefficients. The Bayesian threshold method is then applied to each wavelet coefficient for denoising. The cleansed vibration signal is reconstructed from the denoised wavelet coefficients. Fig. 7 compares the raw and denoised data in both the time (Fig. 7a) and frequency (Fig. 7b) domains. It can be observed that the Bayesian wavelet denoising algorithm provides an effective approach to filter the noise while retaining useful information in the signal.

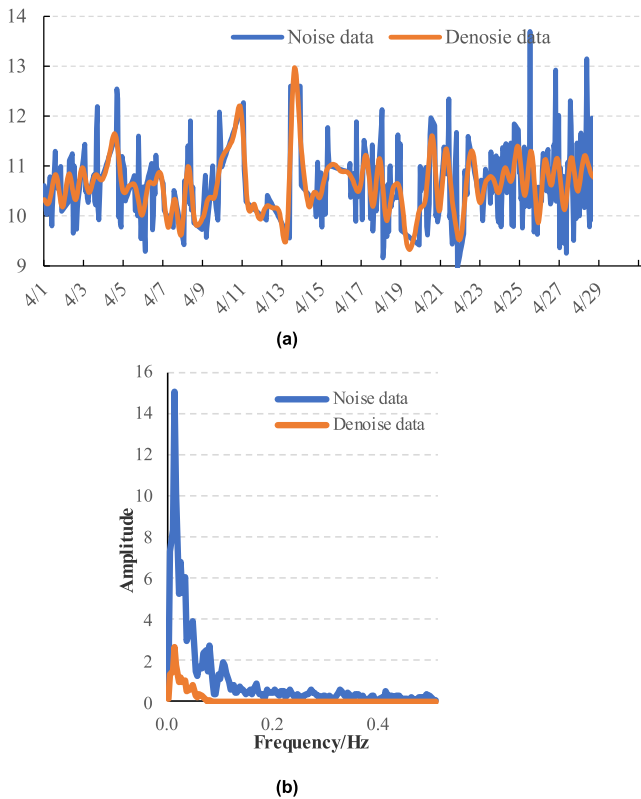


FIGURE 7. Noise reduction effect of vibration signals (a) time domain (b) frequency domain.

C. DIMENSION REDUCTION

The Bayesian PPCA analysis method is employed to reduce the data dimension. Fig. 8 shows the results of PCA analysis on 29 dimensional signals. It is observed that the first princi-

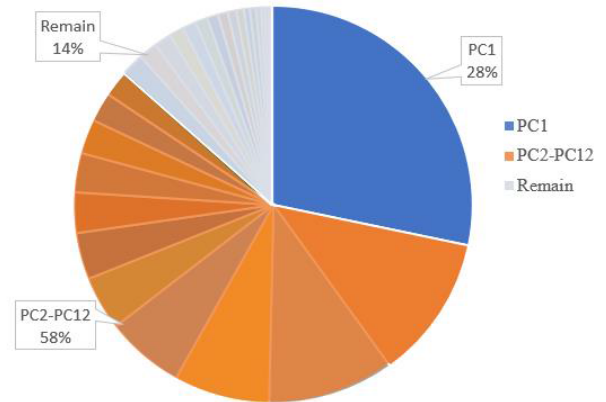


FIGURE 8. PCA contribution rate of 29 dimensional signals.

pal component only accounts for 28% of the information in the data set. The first 12 principal components are required to retain 80% of the information. Therefore, the traditional PCA is not effective in dimension reduction for this example.

To improve the model prediction accuracy with more data information, principal components are combined with actual signals in the modeling. The differential expansion and axial displacement are directly used in the model. PCA is applied for vibration, temperature and rotation speed variables. The dimension reduction effect is shown in Table 3, where w_i represents the contribution rate of each signal to the first principal component after PCA analysis. The values in shaded cells represent the factors that make more contribution to the main components. These factors would play a key role on the condition assessment of the system in the fault diagnosis, which will be subsequently investigated by using the principal component inverse search method.

TABLE 3. PCA weights and contribution rates of parameters.

Weight	Vibration	Temperature	Rotation speed
w_1	0.05	0.20	0.31
w_2	-0.34	0.18	0.69
w_3	-0.44	0.25	0.65
w_4	-0.61	0.31	0.08
w_5	0.28	0.36	
w_6	-0.20	0.23	
w_7	0.36	0.20	
w_8	0.25	0.14	
w_9		0.28	
w_{10}		0.15	
w_{11}		0.37	
w_{12}		0.31	
w_{13}		0.35	
w_{14}		0.01	
w_{15}		0.27	
contribution rate to first principal component	0.74	0.62	0.69

The phase space reconstruction of principal components is employed to determine the input layer of the prediction model. Fig. 9 shows the response value of the discrimination

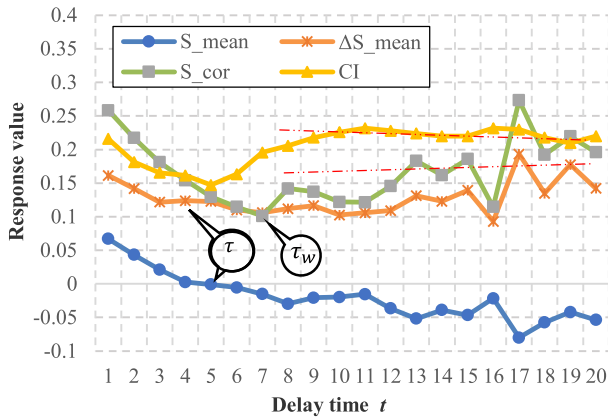


FIGURE 9. Phase space reconstruction of vibration signal principal component by C-C algorithm.

formula for the principal component using the C-C algorithm at different delay times. From the figure, the delay time $\tau = 4$ or $\tau = 3$, $\tau_w = 7$ are obtained, resulting in the embedding dimension $m = 4$ or $m = 3$, and 4 input units of the prediction model.

In Fig. 9, the Correlation Integral (CI) tends to be stable with the time delay. The Lyapunov index, $\lambda = 0.034 > 0$, calculated by using CI, determines that the principal component time series is chaotic. The process is applied for phase space reconstruction of other signals and principal components. The results are shown in Table 4. It is found that except for the differential expansion other time series produce the positive Lyapunov index, indicating that the phase space reconstruction is needed for those time series.

TABLE 4. Lyapunov index of various signals.

Signal category	Embedding dimension	Delay time	Lyapunov index
Differential expansion	4	2	-0.0043
Displacement	3	4	0.0921
Rotation speed	6	3	0.0659
Vibration	4	3	0.034
Temperature	5	2	0.2008

D. DATA RECONSTRUCTION

Fig. 10 explains the construction process of the prediction model. For the vibration principal component, take the embedding dimension m as the number of input units and the delay time τ_d as the distance between the input points.

Various combinations of delay time and embedding dimension are used to establish the model and the model accuracy is justified through R^2 (goodness of fit). The range of R^2 value is $0 \sim 1$ with one indicating the best model. Table 5 shows the goodness of fit value and mean square error of each delay time in different embedding dimensions. It is observed that the results of different combinations are close. The R^2 values for all models are larger than 0.95, and the mean square

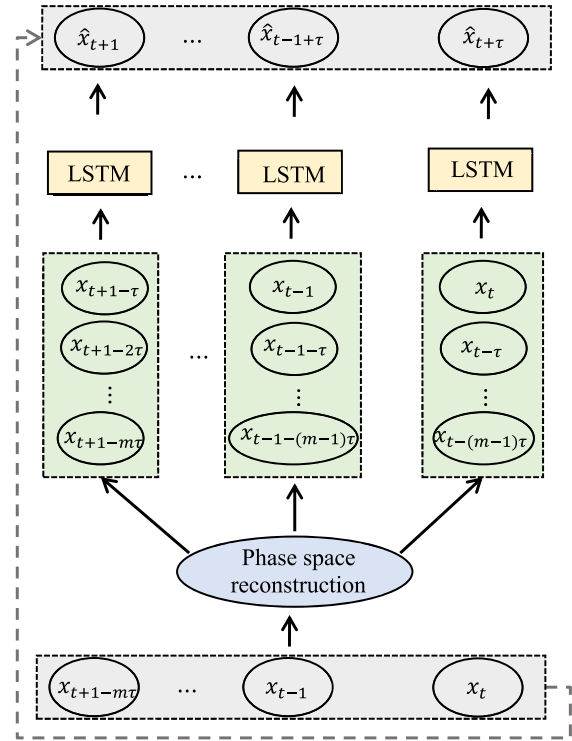


FIGURE 10. Construction process of prediction model.

TABLE 5. Prediction model accuracy under different combinations of embedded dimension and delay time.

Embedding dimension m and delay time t	Training		validation	
	R^2	MSE	R^2	MSE
$m=3, t=2$	0.971	0.011292	0.967	0.00887
$m=3, t=3$	0.982	0.00922	0.939	0.01374
$m=4, t=2$	0.971	0.01117	0.965	0.0139
$m=4, t=3$	0.976	0.00893	0.950	0.01045

error is also around 0.01, which indicate that the model can accurately predict the future signal value, thus generating a reference value to evaluate the turbine health by comparing with the actual monitoring value.

The combination of the embedding dimension 4 and delay time 3 is chosen to construct the prediction model for turbine condition monitoring. In the model validation, the R^2 value is 0.976 and the mean square error is 0.00893. In the model testing, the R^2 value is 0.950 and the mean square error is 0.0104. The comparison of predicted and measured time series is shown in Fig. 11 in terms of model training, validation and testing. The prior probability of model accuracy is assumed to be 50%, i.e., $\pi_0 = 0.5$. The proposed Bayesian reliability assessment method is applied to solve the λ value of the prediction data in the model validation. As shown in Fig. 12, with the accumulation of time and the increase of the sample size, the confidence level tends to stabilize 92%, indicating that the prediction model is acceptable. The similar results are obtained in analysis of the other four types of signal in Table 1.

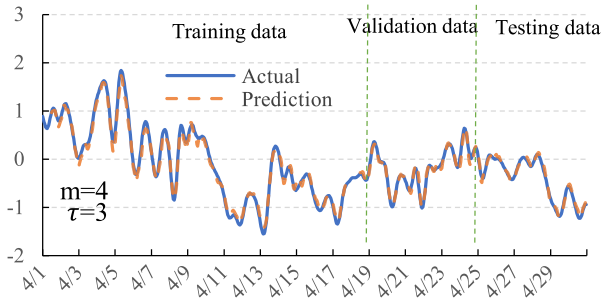


FIGURE 11. Prediction of vibration principal component signal.

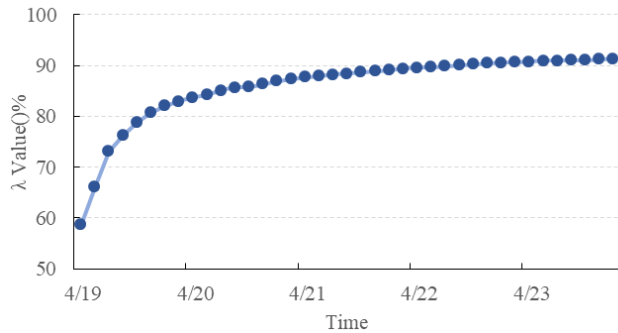


FIGURE 12. Bayesian confidence of prediction model.

In addition, the prediction accuracy of LSTM and RNN model is compared and Fig. 13 shows the comparison results for the speed in validation data series. It is observed that the error of RNN model is obviously larger than that of the LSTM model with the maximum value of 0.4 for the former while only 0.2 for the latter. The LSTM model shows better prediction effect.

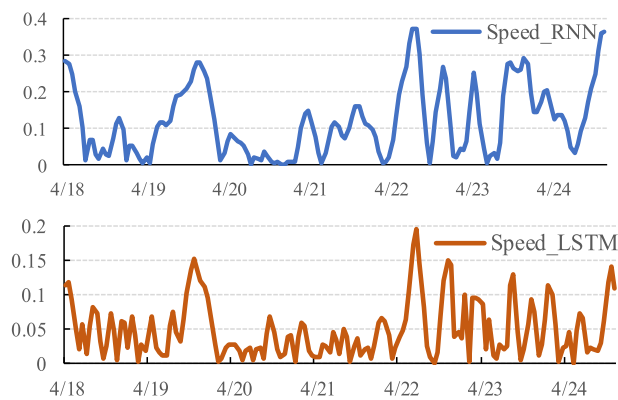


FIGURE 13. Prediction results of LSTM and RNN models.

E. FAULT EARLY WARNING

In this example, the error size is used as an indicator to evaluate the condition of the turbine under investigation. Fig. 14 shows the error values between the model prediction and the actual data of various signals in validation and testing data sets. The maximum absolute value of the validation data

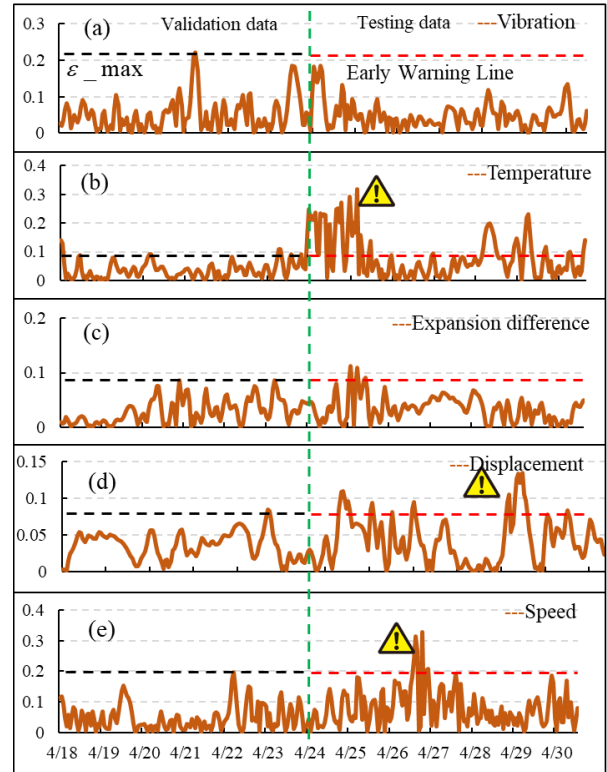


FIGURE 14. Error curves of each signal validation set and testing set.

set error e_{max} is used as the standard deviation to define the warning threshold in the testing data set, as shown as the red line in Fig. 14. From Fig. 14, the error of the temperature principal component signals from April 24 to April 26 between measurement and model prediction obviously exceeds the warning line for a long time, indicating that the turbine has an operation anomaly or equipment fault during this period. In addition, Fig. 14 presents that the rotation speed and displacement also have a short error overflow threshold in a certain period of time, and then return to normal.

The fluctuation of the error value shown in Figure 14 generally results from three possible sources: 1) variation of normal operation, 2) data uncertainties and 3) the possible fault in components. The data collected from the healthy status of the system usually shows the fluctuation resulted from the first two sources. Therefore, the prediction model trained by the healthy data produces the output with the similar error fluctuation as expected. The corresponding threshold determined by the healthy data has taken into account the fluctuation due to the first two sources. The fault will be identified if the error exceeds the predefined threshold. In future research, more robust metric based on Bayesian hypothesis testing will be developed to judge the system status by considering the variation and data uncertainties.

According to the weight information of PCA, the factors with larger influence on principal component can be checked first. For instance, in fault diagnosis on the abnormal situation of the principal component of the rotation speed

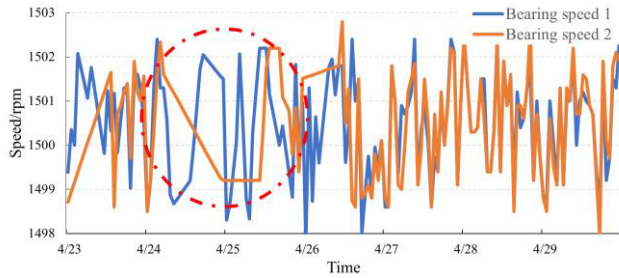


FIGURE 15. Actual value of sub signal of rotation speed in testing set.

from April 26 to April 27, the second and third rotation speed signals with great influence on principal component are judged first by the principal component inverse search method. These two signals correspond to the bearing rotation speed. Fig. 15 illustrates the curves of the two kinds of signals corresponding to the actual monitoring values of the testing data set. The second bearing speed has obvious sampling abnormality from 24th to 26th, but its detection value is still within the normal threshold range of the monitoring system and will not give an alarm. However, the proposed early warning method successfully identifies the signal defect, which can inform the crew to analyze the signal in details and carry out deeper fault diagnosis.

V. MODEL RELIABILITY VALIDATION

To further validate the reliability of the model, the vibration signal is extracted to analyze the crack generation and propagation on the secondary impeller of pneumatic pump in the turbine. The blue curve in Fig. 16 shows the signal change in about four months. Six alarms from the monitoring system indicate that the vibration signal exceeded the system threshold 6 times. In addition, the signal value from May 6 to May 7 has a small abnormality caused by the sensor, but it does not

TABLE 6. Early warning statistics of Bayesian LSTM model.

Fault No.	Monitoring system alarm time	Bayesian LSTM alarm time	Alarm time in advance (hour)
1	3/7 17:00	3/5 21:00	44
2	3/16 19:00	3/15 13:00	30
3	4/6 18:00	4/5 2:00	40
4	4/14 20:00	4/14 4:00	16
5	None (5/6 10:00)	5/5 14:00	20
6	5/11 18:00	5/10 14:00	28
7	5/18 11:00	5/16 19:00	40
Average			31

exceed the threshold, and the monitoring system does not give an alarm.

The signal is processed in the order of wavelet packet denoising, phase space reconstruction and LSTM model prediction. The Bayesian LSTM model is established by using health data before February 10 as training set and health data from February 10 to March 1 as validation set. The prediction error of the model is shown in the orange curve of Fig. 16, and the threshold range is determined by the maximum and minimum value of the validation set error. As shown in the figure, the model correctly predicted all 6 system alarms, with a missed diagnosis rate of 0%. On May 6, the prediction model produced an alarm for the abnormality not detected by the monitoring system. Sorting out these seven faults, the alarm time of monitoring system and prediction model are listed in table 6. The LSTM model predicts all faults and given an alarm 31 hours in advance on average. In addition, during the normal period, the prediction model produced four alarms on March 22, April 9, April 18 and April 26, respectively. Despite a small amplitude and short duration, these four alarms indicate that the signal change law is abnormal compared with the training set, which is also validated by the continuous and serious crack fault observed later. The above analysis illustrates that the model can predict the alarms of the monitoring system and sensor failures in advance.

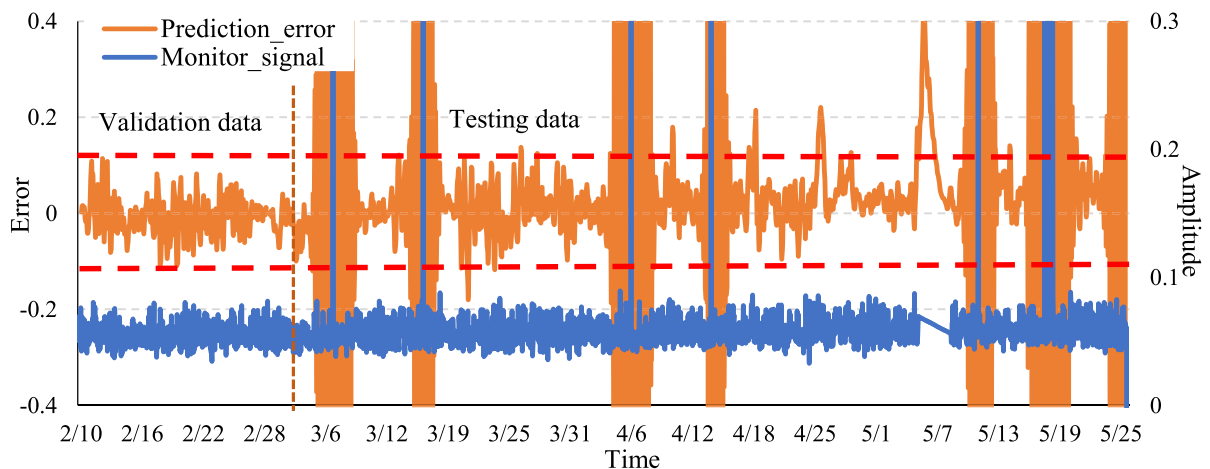


FIGURE 16. The prediction result of Bayesian LSTM model on crack failure.

VI. CONCLUSION

This paper presents a Bayesian LSTM model for fault early warning of steam turbine in nuclear power plant. The main conclusions are drawn as follows:

(1) The proposed early warning model focuses on solving the problems that are difficult to find during the creep period of the steam turbine unit failure. Through using 29 critical factors in steam turbine like vibration, temperature, displacement and speed, the Bayesian LSTM predictive model has been validated with 92% confidence and demonstrated to accurately identify defects.

(2) In the PPCA dimension reduction, the signals are regrouped according to the type in order to improve the effectiveness of the method in terms of contribution rate from 28% to 60% for the first principal component. In addition, the principal component inverse search method is studied. The weights of each signal to the principal component are investigated by the PPCA weight inverse matrix so that the PPCA provides a means to fault diagnosis after early warning.

(3) The error of the validation data set is used as the prior information to define the warning threshold. When the abnormality is detected, by using the principal component inverse search method to detect the signal with high contribution rate, specific abnormal signals can be identified. In the case study, it is successfully found that the abnormality of the principle component of rotation speed is caused by the abnormal sampling of the bearing signal. In the crack fault prediction, the proposed model successfully predicted all alarms sent by the equipment monitoring system.

In future research we will further investigate the proposed methodology by using the simulated data.

REFERENCES

- [1] J. Ma and J. Jiang, "Applications of fault detection and diagnosis methods in nuclear power plants: A review," *Progr. Nucl. Energy*, vol. 53, no. 3, pp. 255–266, Apr. 2011, doi: [10.1016/j.pnucene.2010.12.001](https://doi.org/10.1016/j.pnucene.2010.12.001).
- [2] Y. Gong, X. Su, H. Qian, and N. Yang, "Research on fault diagnosis methods for the reactor coolant system of nuclear power plant based on D-S evidence theory," *Ann. Nucl. Energy*, vol. 112, pp. 395–399, Feb. 2018, doi: [10.1016/j.anucene.2017.10.026](https://doi.org/10.1016/j.anucene.2017.10.026).
- [3] B.-S. Peng, H. Xia, Y.-K. Liu, B. Yang, D. Guo, and S.-M. Zhu, "Research on intelligent fault diagnosis method for nuclear power plant based on correlation analysis and deep belief network," *Progr. Nucl. Energy*, vol. 108, pp. 419–427, Sep. 2018, doi: [10.1016/j.pnucene.2018.06.003](https://doi.org/10.1016/j.pnucene.2018.06.003).
- [4] Z. Dong, "Boolean network-based sensor selection with application to the fault diagnosis of a nuclear plant," *Energies*, vol. 10, no. 12, p. 2125, Dec. 2017, doi: [10.3390/en10122125](https://doi.org/10.3390/en10122125).
- [5] G. Wu, J. Tong, L. Zhang, Y. Zhao, and Z. Duan, "Framework for fault diagnosis with multi-source sensor nodes in nuclear power plants based on a Bayesian network," *Ann. Nucl. Energy*, vol. 122, pp. 297–308, Dec. 2018, doi: [10.1016/j.anucene.2018.08.050](https://doi.org/10.1016/j.anucene.2018.08.050).
- [6] Z. Welz, J. Coble, B. Upadhyaya, and W. Hines, "Maintenance-based prognostics of nuclear plant equipment for long-term operation," *Nucl. Eng. Technol.*, vol. 49, no. 5, pp. 914–919, Aug. 2017, doi: [10.1016/j.net.2017.06.001](https://doi.org/10.1016/j.net.2017.06.001).
- [7] J. H. Min, D.-W. Kim, and C.-Y. Park, "Demonstration of the validity of the early warning in online monitoring system for nuclear power plants," *Nucl. Eng. Design*, vol. 349, pp. 56–62, Aug. 2019, doi: [10.1016/j.nucengdes.2019.04.028](https://doi.org/10.1016/j.nucengdes.2019.04.028).
- [8] Y. Li, Y. Yang, G. Li, M. Xu, and W. Huang, "A fault diagnosis scheme for planetary gearboxes using modified multi-scale symbolic dynamic entropy and mRMR feature selection," *Mech. Syst. Signal Process.*, vol. 91, pp. 295–312, Jul. 2017, doi: [10.1016/j.ymssp.2016.12.040](https://doi.org/10.1016/j.ymssp.2016.12.040).
- [9] X. Zhen, M. Yu, F. Zheng, I. B. Nachum, M. Bhaduri, D. Laidley, and S. Li, "Multitarget sparse latent regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1575–1586, May 2018, doi: [10.1109/TNNLS.2017.2651068](https://doi.org/10.1109/TNNLS.2017.2651068).
- [10] D. L. J. Alexander, A. Tropsha, and D. A. Winkler, "Beware of R(2): Simple, unambiguous assessment of the prediction accuracy of QSAR and QSPR models," *J. Chem. Inf. Model.*, vol. 55, no. 7, pp. 1316–1322, Jul. 2015, doi: [10.1021/acs.jcim.5b00206](https://doi.org/10.1021/acs.jcim.5b00206).
- [11] Y. Ying, J. Li, Z. Chen, and J. Guo, "Study on rolling bearing on-line reliability analysis based on vibration information processing," *Comput. Electr. Eng.*, vol. 69, pp. 842–851, Jul. 2018, doi: [10.1016/j.compeleceng.2017.11.029](https://doi.org/10.1016/j.compeleceng.2017.11.029).
- [12] X. Jiang and C. Foster, "Remote thermal performance monitoring and diagnostics: Turning data into knowledge," in *Proc. ASME Power Conf.*, Boston, MA, USA, 2013, Art. no. V002T13A004, doi: [10.1115/POWER2013-98246](https://doi.org/10.1115/POWER2013-98246).
- [13] X. Jiang and C. Foster, "Plant performance monitoring and diagnostics: Remote, real-time and automation," in *Proc. ASME Turbine Tech. Conf. Expo.*, Düsseldorf, Germany, Jun. 2014, Art. no. V006T06A034, doi: [10.1115/GT2014-27314](https://doi.org/10.1115/GT2014-27314).
- [14] S. Xu, X. Jiang, J. Huang, S. Yang, and X. Wang, "Bayesian wavelet PCA methodology for turbomachinery damage diagnosis under uncertainty," *Mech. Syst. Signal Process.*, vol. 80, pp. 1–18, Dec. 2016, doi: [10.1016/j.ymssp.2016.04.031](https://doi.org/10.1016/j.ymssp.2016.04.031).
- [15] X. Jiang, S. Mahadevan, and Y. Yuan, "Fuzzy stochastic neural network model for structural system identification," *Mech. Syst. Signal Process.*, vol. 82, pp. 394–411, Jan. 2017, doi: [10.1016/j.ymssp.2016.05.030](https://doi.org/10.1016/j.ymssp.2016.05.030).
- [16] V. Martínez-Martínez, F. J. Gomez-Gil, J. Gomez-Gil, and R. Ruiz-Gonzalez, "An artificial neural network based expert system fitted with genetic algorithms for detecting the status of several rotary components in agro-industrial machines using a single vibration signal," *Expert Syst. Appl.*, vol. 42, nos. 17–18, pp. 6433–6441, Oct. 2015, doi: [10.1016/j.eswa.2015.04.018](https://doi.org/10.1016/j.eswa.2015.04.018).
- [17] A. Ayodeji, Y.-K. Liu, and H. Xia, "Knowledge base operator support system for nuclear power plant fault diagnosis," *Progr. Nucl. Energy*, vol. 105, pp. 42–50, May 2018, doi: [10.1016/j.pnucene.2017.12.013](https://doi.org/10.1016/j.pnucene.2017.12.013).
- [18] D. Cabrera, F. Sancho, C. Li, M. Cerrada, R.-V. Sánchez, F. Pacheco, and J. V. de Oliveira, "Automatic feature extraction of time-series applied to fault severity assessment of helical gearbox in stationary and non-stationary speed operation," *Appl. Soft Comput.*, vol. 58, pp. 53–64, Sep. 2017, doi: [10.1016/j.asoc.2017.04.016](https://doi.org/10.1016/j.asoc.2017.04.016).
- [19] D. Dou and S. Zhou, "Comparison of four direct classification methods for intelligent fault diagnosis of rotating machinery," *Appl. Soft Comput.*, vol. 46, pp. 459–468, Sep. 2016, doi: [10.1016/j.asoc.2016.05.015](https://doi.org/10.1016/j.asoc.2016.05.015).
- [20] D. F. Amare, T. B. Aklilu, and S. I. Gilani, "Gas path fault diagnostics using a hybrid intelligent method for industrial gas turbine engines," *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 40, no. 12, Dec. 2018, doi: [10.1007/s40430-018-1497-6](https://doi.org/10.1007/s40430-018-1497-6).
- [21] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mech. Syst. Signal Process.*, vol. 107, pp. 241–265, Jul. 2018, doi: [10.1016/j.ymssp.2017.11.024](https://doi.org/10.1016/j.ymssp.2017.11.024).
- [22] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019, doi: [10.1016/j.ymssp.2018.05.050](https://doi.org/10.1016/j.ymssp.2018.05.050).
- [23] D. Lee, P. H. Seong, and J. Kim, "Autonomous operation algorithm for safety systems of nuclear power plants by using long-short term memory and function-based hierarchical framework," *Ann. Nucl. Energy*, vol. 119, pp. 287–299, Sep. 2018, doi: [10.1016/j.anucene.2018.05.020](https://doi.org/10.1016/j.anucene.2018.05.020).
- [24] H.-X. Tian, D.-X. Ren, and K. Li, "A hybrid vibration signal prediction model using autocorrelation local characteristic-scale decomposition and improved long short term memory," *IEEE Access*, vol. 7, pp. 60995–61007, 2019, doi: [10.1109/ACCESS.2019.2916000](https://doi.org/10.1109/ACCESS.2019.2916000).
- [25] B. Zhang, S. Zhang, and W. Li, "Bearing performance degradation assessment using long short-term memory recurrent network," *Comput. Ind.*, vol. 106, pp. 14–29, Apr. 2019, doi: [10.1016/j.compind.2018.12.016](https://doi.org/10.1016/j.compind.2018.12.016).
- [26] J. Yang, Y. Guo, and W. Zhao, "Long short-term memory neural network based fault detection and isolation for electro-mechanical actuators," *Neurocomputing*, vol. 360, pp. 85–96, Sep. 2019, doi: [10.1016/j.neucom.2019.06.029](https://doi.org/10.1016/j.neucom.2019.06.029).
- [27] M. Li, D. Yu, Z. Chen, K. Xiahou, T. Ji, and Q. H. Wu, "A data-driven residual-based method for fault diagnosis and isolation in wind turbines," *IEEE Trans. Sustain. Energy*, vol. 10, no. 2, pp. 895–904, Apr. 2019, doi: [10.1109/TSTE.2018.2853990](https://doi.org/10.1109/TSTE.2018.2853990).

- [28] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications," *Mech. Syst. Signal Process.*, vol. 42, nos. 1–2, pp. 314–334, Jan. 2014, doi: [10.1016/j.ymssp.2013.06.004](https://doi.org/10.1016/j.ymssp.2013.06.004).
- [29] C. Montalvo and A. García-Berrocal, "Improving the *in situ* measurement of RTD response times through discrete wavelet transform in NPP," *Ann. Nucl. Energy*, vol. 80, pp. 114–122, Jun. 2015, doi: [10.1016/j.anucene.2015.02.004](https://doi.org/10.1016/j.anucene.2015.02.004).
- [30] B. R. Upadhyaya, C. Mehta, and D. Bayram, "Integration of time series modeling and wavelet transform for monitoring nuclear plant sensors," *IEEE Trans. Nucl. Sci.*, vol. 61, no. 5, pp. 2628–2635, Oct. 2014, doi: [10.1109/TNS.2014.2341035](https://doi.org/10.1109/TNS.2014.2341035).
- [31] P. Baraldi, F. Di Maio, and E. Zio, "Unsupervised clustering for fault diagnosis in nuclear power plant components," *Int. J. Comput. Intell. Syst.*, vol. 6, no. 4, pp. 764–777, Aug. 2013, doi: [10.1080/18756891.2013.804145](https://doi.org/10.1080/18756891.2013.804145).
- [32] L. Yong-Kuo, A. Abiodun, W. Zhi-Bin, W. Mao-Pu, P. Min-Jun, and Y. Wei-Feng, "A cascade intelligent fault diagnostic technique for nuclear power plants," *J. Nucl. Sci. Technol.*, vol. 55, no. 3, pp. 254–266, Mar. 2018, doi: [10.1080/00223131.2017.1394228](https://doi.org/10.1080/00223131.2017.1394228).
- [33] S. Park, J. Park, and G. Heo, "Transient diagnosis and prognosis for secondary system in nuclear power plants," *Nucl. Eng. Technol.*, vol. 48, no. 5, pp. 1184–1191, Oct. 2016, doi: [10.1016/j.net.2016.03.009](https://doi.org/10.1016/j.net.2016.03.009).
- [34] W. Li, M. Peng, and Q. Wang, "False alarm reducing in PCA method for sensor fault detection in a nuclear power plant," *Ann. Nucl. Energy*, vol. 118, pp. 131–139, Aug. 2018, doi: [10.1016/j.anucene.2018.04.012](https://doi.org/10.1016/j.anucene.2018.04.012).
- [35] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, Sep. 1994, doi: [10.1093/biomet/81.3.425](https://doi.org/10.1093/biomet/81.3.425).
- [36] R. R. Coifman and M. V. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 713–718, Mar. 1992, doi: [10.1109/18.119732](https://doi.org/10.1109/18.119732).
- [37] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000, doi: [10.1109/83.862633](https://doi.org/10.1109/83.862633).
- [38] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 61, no. 3, pp. 611–622, Aug. 1999, doi: [10.1111/1467-9868.00196](https://doi.org/10.1111/1467-9868.00196).
- [39] F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence, Warwick (Lecture Notes in Mathematics)*, vol. 898, D. A. Rand and L.-S. Young, Eds. Coventry, U.K.: Springer, 1981, pp. 366–381, doi: [10.1007/BFb0091924](https://doi.org/10.1007/BFb0091924).
- [40] H. S. Kim, R. Eykholt, and J. D. Salas, "Nonlinear dynamics, delay times, and embedding windows," *Physica D*, vol. 127, nos. 1–2, pp. 48–60, Mar. 1999, doi: [10.1016/S0167-2789\(98\)00240-1](https://doi.org/10.1016/S0167-2789(98)00240-1).
- [41] Y. Cheng, H. Zhu, J. Wu, and X. Shao, "Machine health monitoring using adaptive kernel spectral clustering and deep long short-term memory recurrent neural networks," *IEEE Trans. Ind. Informat.*, vol. 15, no. 2, pp. 987–997, Feb. 2019, doi: [10.1109/TII.2018.2866549](https://doi.org/10.1109/TII.2018.2866549).
- [42] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [43] D. You, X. Shen, Y. Zhu, J. Deng, and F. Li, "A quantitative validation method of kriging metamodel for injection mechanism based on Bayesian statistical inference," *Metals*, vol. 9, no. 5, p. 493, Apr. 2019, doi: [10.3390/met9050493](https://doi.org/10.3390/met9050493).
- [44] B. Cai, X. Kong, Y. Liu, J. Lin, X. Yuan, H. Xu, and R. Ji, "Application of Bayesian networks in reliability evaluation," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2146–2157, Apr. 2019, doi: [10.1109/TII.2018.2858281](https://doi.org/10.1109/TII.2018.2858281).
- [45] B. Cai, X. Shao, Y. Liu, X. Kong, H. Wang, H. Xu, and W. Ge, "Remaining useful life estimation of structure systems under the influence of multiple causes: Subsea pipelines as a case study," *IEEE Trans. Ind. Electron.*, to be published, doi: [10.1109/TIE.2019.2931491](https://doi.org/10.1109/TIE.2019.2931491).



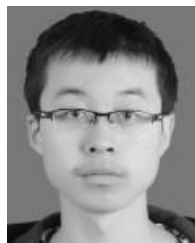
GAOJUN LIU received the B.S. degree in computer science and applications from Harbin Engineering University.

He is currently a Senior Executive Designer with China Nuclear Power Design Company Ltd. His research interests include design and implementation of real-time information monitoring systems in nuclear power plants.



HAIXIA GU received the M.S. degree in energy and power engineering from Dongnan University, China.

She became a Senior Engineer with China Nuclear Power Engineering Company Ltd., in 2016. Her research interests include condition monitoring, big data mining, and applications in nuclear power plants.



XIAOCHENG SHEN was born in Jiangsu, China, in 1996. He received the B.Eng. degree in mechanical engineering from the Nanjing Institute of Technology, Nanjing, China, in 2018. He is currently pursuing the M.Eng. degree in mechanical engineering with the South China University of Technology, Guangzhou.

His current research interests include health monitoring and system development.



DONGDONG YOU received the M.S. degree in computer science and technology, and the Ph.D. degree in mechanical engineering from the South China University of Technology, Guangzhou, China, in 2003 and 2007, respectively.

He is currently an Associate Professor with the School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou. His research interests include intelligent maintenance of large industrial systems, manufacturing process optimization design, and numerical modeling and simulation.

• • •