



Fault prediction method for nuclear power machinery based on Bayesian PPCA recurrent neural network model

Jun Ling^{1,3} · Gao-Jun Liu¹ · Jia-Liang Li² · Xiao-Cheng Shen² · Dong-Dong You²

Received: 17 April 2020 / Revised: 10 June 2020 / Accepted: 14 June 2020 / Published online: 29 July 2020

© China Science Publishing & Media Ltd. (Science Press), Shanghai Institute of Applied Physics, the Chinese Academy of Sciences, Chinese Nuclear Society and Springer Nature Singapore Pte Ltd. 2020

Abstract Early fault warning for nuclear power machinery is conducive to timely troubleshooting and reductions in safety risks and unnecessary costs. This paper presents a novel intelligent fault prediction method, integrated probabilistic principal component analysis (PPCA), multi-resolution wavelet analysis, Bayesian inference, and RNN model for nuclear power machinery that consider data uncertainty and chaotic time series. After denoising the source data, the Bayesian PPCA method is employed for dimensional reduction to obtain a refined data group. A recurrent neural network (RNN) prediction model is constructed, and a Bayesian statistical inference approach is developed to quantitatively assess the prediction reliability of the model. By modeling and analyzing the data collected on the steam turbine and components of a nuclear power plant, the results of the goodness of fit, mean square error distribution, and Bayesian confidence indicate that the

proposed RNN model can implement early warning in the fault creep period. The accuracy and reliability of the proposed model are quantitatively verified.

Keywords Fault prediction · Nuclear power machinery · Steam turbine · Recurrent neural network · Probabilistic principal component analysis · Bayesian confidence

1 Introduction

Using a real-time monitoring system to collect the operations data of mechanical equipment in nuclear power plants (NPPs) for early warning in the early stage of equipment failure allows troubleshooting, which avoids major safety accidents, reduces unplanned shutdown maintenance of units, and reduces costs [1–3]. The establishment of a data-driven prediction model for mechanical equipment fault prediction has become an important means of predictive maintenance, and research on nuclear power machinery has gradually increased because of the recent and rapid development of artificial intelligence algorithms and big data technology [4–6]. Xie et al. [7] designed an online early warning system to track and predict the critical reaction of a nuclear reactor through two independent online simulation systems. Qian et al. [8] presented a hierarchical multi-dimensional method for fault detection of an NPP main pipeline. Min et al. [9] used a pattern recognition early warning system developed with AAKR technology to demonstrate the effectiveness of a real-time monitoring and early warning system for NPPs. Peng et al. [10] utilized the feature selection ability of association analysis and the depth confidence network (DBN) method to detect faults in nuclear power machinery. Yao et al. [11]

This work was supported by the National Natural Science Foundation of China (No. 51875209), the Guangdong Basic and Applied Basic Research Foundation (No. 2019B1515120060), and the Open Funds of State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment.

✉ Jia-Liang Li
lijialiang_scut@126.com

✉ Dong-Dong You
youdd@scut.edu.cn

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

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

³ Department of Automation, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China

introduced a fault diagnosis method for NPP full range simulators based on state information imaging. By using machine learning and image processing technology, historical data and synthetic grey image data are analyzed, and the system learns to achieve image feature extraction and classification to perform fault diagnosis.

A variety of data-driven approaches are used in the fault prediction of large mechanical equipment. Qin et al. [12] employed an approach based on time series and Bayesian discriminant analyses to solve the problems of type identification and diagnosis of concurrent faults without characteristic parameters in rotating machinery. Mehrdad et al. [13] presented a nonparametric single spline regression approach to construct the power curve model of the generator set. Aye and Heyns [14] proposed an optimal Gaussian process regression through the combination of simple mean value and variance to predict a low error rate for the remaining service life of low-speed bearings. The judgment methods for regression model accuracy are continuously developing. Jiang and Yin [15] proposed and applied a recursive total principle component regression-based design and implementation approach for efficient data-driven fault detection for vehicular cyber-physical systems. Gao et al. [16] established a partial least squares-aided data-driven model predictive control approach to improve prediction accuracy. Herp et al. [17] developed a statistical method of online extraction and prediction of turbine state based on Bayesian inference. The residuals of bearing temperature measurement were inferred online, and the prediction probability is calculated by the sample model and the risk function describing the state transition probability to predict the fault state in advance. Li et al. [18] proposed a method for rolling bearing fault identification based on the multifractal and grey system theories, aiming at the non-equilibrium and nonlinear characteristics of bearing vibration signals and the complexity of the distribution of state indication information in the signal. Liu et al. [19] presented the thermal component of a fault prediction method based on the convolutional neural network (CNN) to address the disastrous consequences caused by thermal component faults of gas turbines. This method shows that the CNN is a feasible method to resolve thermal component fault detection. Liu and Karimi [20] established two machine learning models based on an artificial neural network and a high-dimensional model representation to predict the operation characteristics of steam turbines and air compressors and provided a basis for continuous health monitoring and fault diagnosis.

Deep learning can directly reflect the characteristics through the training sample data to reduce the influence of assumptions and simplification on calculation results and has recently been widely used for mechanical fault identification [21–23]. The RNN model shows good

performance in capturing temporal correlations in data and can store and transmit the sequence information multiple times. Liu et al. [24] developed an RNN-based fault identification approach that uses a denoising auto-encoder based on a gated recursive unit to predict multiple vibration values of rolling bearings in subsequent time series. Hadi and Shahnazari [25] proposed a fault detection and isolation (FDI) method. Based on the RNN, this method models and inverts the nonlinear system, establishes a factory prediction model, and makes use of the residual generated from the model for fault identification. Wang et al. [26] developed an RNN-based algorithm to effectively handle the multi-classification fault diagnosis for wind power systems. Palau et al. [27] employed Weibull time to an event-RNN algorithm for distributed collaborative prognostics. The industrial gas turbine unit data and c-mapps engine degradation dataset are used in the experiment. Wang et al. [28] analyzed the motor vibration signal and multi-scale stator current signal and presented a multi-resolution and multi-sensor fusion network model for motor fault diagnosis based on RNN.

The dimension of the data must be reduced, and more refined information must be used to conduct a comprehensive analysis of the collected data to refine and simplify the research. Principal component analysis (PCA) can effectively reduce the dimensionality of compressed data by retaining the original data feature information and solve the multi-variable correlation problem to reduce the complexity of the problem analysis and is widely applied in big data processing, pattern recognition, and image processing fields. Li et al. [29] established an optimized PCA model to perform fault detection of sensors in NPPs and verified that the model can detect and reconstruct the fault sensors well by simulation. Prusty et al. [30] employed PCA to reduce the dimension of a large number of plant signals transmitted by the Prototype Fast Breeder Reactor (PFBR) in a NPP, improving the decision-making capability of the operator in catastrophic conditions. Wu et al. [31] constructed a fault detection model of a pressurized water reactor in a NPP based on the BN-FDD system framework. PCA, fuzzy theory, and data fusion were used to promote data accuracy, and multiple sensor data were combined into one node data. Sharifi and Langari [32] divided the measurement space into several local linear regions associated with a PPCA model and presented a sensor fault diagnosis method for a nonlinear system by considering the data uncertainty. Xiang et al. [33] utilized the PPCA denoising model for rolling bearing fault prediction. In this model, the subspace of the principal component retains the more useful original information and fault signal, and the noise and related linear information are projected into the remaining subspace.

The above literature introduces the analysis methods of fault prediction for different mechanical equipment. The following problems regarding the application of fault detection and early warning for nuclear power rotating mechanical equipment must be improved: (1) imperfections in the source data and multiple variable redundancies; (2) monitoring and early warning during the creep period of equipment failure; and (3) the quantitative evaluation of the reliability of the prediction model.

This study combines pattern recognition technology and deep learning to present a fault prediction approach for steam turbines, pumps, and other mechanical equipment in NPPs based on Bayesian PPCA RNN to improve the aforementioned issues. After wavelet packet threshold denoising, the signal data are dimensionally reduced by using a Bayesian PPCA method. A fully connected RNN prediction model is established and verified by using the goodness of fit and mean square error. The model reliability is quantified by calculating the Bayesian factor and confidence. Combined with the prior information in the historical dataset, the proposed method calculates the residual between the prediction and hypothetical health values to find the unit failure during the creep period.

2 Data integration analysis

For the monitoring data of rotating machinery in NPPs, the discrete wavelet packet transform (DWPT) method is first used to denoise the data. The number of monitoring data prediction variables must be reduced to simplify the research and reduce the prediction time and calculation cost. Because the variables with a certain correlation cannot be directly ignored, a Bayesian PPCA method [34] is employed to extract the principal component signal with the highest contribution rate for analysis and prediction. Considering the data uncertainty, PPCA estimates the discarded information as Gaussian noise.

There are N samples with dimension d , where d is the number of variables, and $\mathbf{X} = (x_1, x_2, \dots, x_N)$; thus, each variable contains N denoised time series points. PPCA assumes the existence of q ($q \leq d$) dimension hidden variables $\boldsymbol{\beta}$. The hidden variable model relates the relevant matrix $\mathbf{X}$ of the relevant sample data and the matrix $\boldsymbol{\beta}$ of uncorrelated hidden response variables with the formula (1):

$$\mathbf{X} = \mathbf{W}\boldsymbol{\beta} + \boldsymbol{\mu} + \boldsymbol{\varepsilon}, \quad (1)$$

where $\mathbf{W}$ is the $q \times d$ weight matrix that describes the relationship between $\mathbf{X}$ and $\boldsymbol{\beta}$. Here, $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_d)^T$ is the mean vector that describes the mean value of each variable in the data matrix, which is $\mu_i = (1/N) \sum_{j=1}^N x_{ij}$. $\boldsymbol{\beta}$

is a vector $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_N)$ composed of q latent variable factors. Each latent variable factor contains N corresponding points in the hidden space, obeys $\boldsymbol{\beta} \sim N(0, \mathbf{I}_q)$, and is the result of dimension reduction. Here, $\boldsymbol{\varepsilon}$ is the possible error or noise that cannot be eliminated by the DWPT method combined with Bayesian theory, which is the Gaussian variable $N \sim (0, \sigma^2 \mathbf{I}_d)$ ($\mathbf{I}$ is the unit matrix) with an independent distribution. The principal axis of the PPCA is an incremental rule, which is a special factor analysis model. Combining with various conditions and formula (1), we obtain

$$\mathbf{X} \sim N(\boldsymbol{\mu}, \mathbf{W}\mathbf{W}^T + \sigma^2 \mathbf{I}_d). \quad (2)$$

The specific parameters are solved by the maximum likelihood estimation. From formula (2), the prior distribution of the hidden variable $\boldsymbol{\beta}$ is

$$p(\boldsymbol{\beta}) = (2\pi)^{-\frac{q}{2}} \exp\left\{-\frac{1}{2}\boldsymbol{\beta}^T \boldsymbol{\beta}\right\}. \quad (3)$$

Under the condition of an implicit variable $\boldsymbol{\beta}$, from Eq. (1), the prior probability distribution of the sample data $\mathbf{X}$ is

$$p(\mathbf{X}|\boldsymbol{\beta}) = (2\pi\sigma^2)^{-\frac{d}{2}} \exp\left\{-\frac{1}{2\sigma^2}\|\mathbf{X} - \mathbf{W}\boldsymbol{\beta} - \boldsymbol{\mu}\|_2^2\right\}. \quad (4)$$

By substituting formula (4) into the Bayesian formula, the posterior probability density distribution of the hidden variable $\boldsymbol{\beta}$ concerning sample data $\mathbf{X}$ can be obtained as follows:

$$p(\boldsymbol{\beta}|\mathbf{X}) = (2\pi)^{-\frac{q}{2}} |\sigma^2 \mathbf{M}|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \langle \boldsymbol{\beta} \rangle)^T (\sigma^2 \mathbf{M})^{-1} (\boldsymbol{\beta} - \langle \boldsymbol{\beta} \rangle)\right\}, \quad (5)$$

where $\mathbf{M} = (\mathbf{W}^T \mathbf{W} + \sigma^2 \mathbf{I}_q)^{-1}$ and $\langle \boldsymbol{\beta} \rangle = \mathbf{M}^{-1} \mathbf{W}^T (\mathbf{X} - \boldsymbol{\mu})$. The PPCA considers that the q components of the expected $\langle \boldsymbol{\beta} \rangle$ of the $\boldsymbol{\beta}$ posterior probability are the result of $\mathbf{X}$ dimension reduction, namely the q principal components. The maximum likelihood function can be used to estimate $\mathbf{W}$, σ :

$$\hat{\mathbf{W}} = \mathbf{U}_q (\boldsymbol{\Lambda}_q - \sigma^2 \mathbf{I}_q)^{1/2} \mathbf{R}, \quad (6)$$

$$\hat{\sigma}^2 = \frac{1}{d-q} \sum_{j=q+1}^d \lambda_j. \quad (7)$$

In formulas (6) and (7), the covariance matrix of sample data $\mathbf{X}$ is decomposed in accordance with eigenvalues, and λ_j is obtained. Thus, $C\mathbf{v}_j = \lambda_j \mathbf{v}_j$ and $\mathbf{v}_j$ is an eigenvector. $\mathbf{U}_q$ and $\boldsymbol{\Lambda}_q$ are diagonal matrices composed of eigenvector matrices and eigenvalues that correspond to the previous q eigenvalues, respectively, with $\mathbf{U}_q = (v_1, v_2, \dots, v_q)$, $\boldsymbol{\Lambda}_q = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_q)$. When the cumulative variance rate

of the first k principal components in q principal components reaches a certain contribution rate (as determined by the actual situation), only the k -dimensional data after the dimension reduction can be used for the subsequent data processing.

3 Fully connected RNN prediction model

3.1 RNN model construction

In this study, an RNN suitable for sequence data modeling is used to predict a time series. The neurons with a cyclic structure retain and apply the state information of the previous moment as memory to current output calculation; thus, the nodes between the same hidden layers are connected. The RNN can transmit information laterally among neurons and partially express correlations within the data. This information transmission mode matches well with the state process of operational nuclear power machinery. The running state at a given moment will have a certain impact on the running state at the next moment, and the collected data also correlate.

According to the embedding dimension m and time delay τ , an RNN structure prediction model is constructed (Fig. 1), where $x_t(x_1, x_2, \dots, x_\tau)$ is the network input, the corresponding hidden layer is $s_t(s_1, s_2, \dots, s_\tau)$, and the output layer is $\hat{y}_{t+(n+1)\tau}(y_1, y_2, \dots, y_\tau)$. $\mathbf{V}$ is the weight matrix from the hidden layer to the output layer. $\mathbf{W}$ is a weight matrix that represents the value of the last time on the hidden layer as the input weight. $\mathbf{U}$ is the weight matrix from the input layer to the hidden layer.

In the model, the same weight parameters are used at different times, and the activation function uses the rectified linear unit (ReLU) function uniformly. Units with a

certain amount of m are connected, and the last unit provides the output value. The hidden layer state s_0 of the initial input is a random value, and the prediction value of the next τ time point outputted by the last unit is $\hat{y}_{t+(n+1)\tau}$. For time $t + n\tau$, formula (8) is used to calculate the forward propagation from input to output:

$$\begin{aligned} c_{t+n\tau} &= Ws_{t+(n-1)\tau} + Ux_{t+n\tau} + a, \\ s_{t+n\tau} &= \text{ReLU}(c_{t+n\tau}), \\ o_{t+n\tau} &= Vs_{t+n\tau} + b, \\ \hat{y}_{t+(n+1)\tau} &= \text{ReLU}(o_{t+n\tau}). \end{aligned} \quad (8)$$

In formula (8), c_t and o_t are used as intermediate variables to participate in the backpropagation calculation; a and b represent the bias terms of the hidden layer neurons and output layer neurons, respectively.

After constructing the model, a backpropagation through time (BPTT) algorithm [35] is employed to train the RNN model. Based on formula (8), the loss function is established, and the minimum value of the loss function is calculated. The negative log likelihood function is used to establish the loss function:

$$L_t = - \sum_i^n y_t^{(i)} \log(\hat{y}_t^{(i)}), \quad (9)$$

where $y_t^{(i)}$ is the i th element of the output y_t , $\hat{y}$ is the predicted data, and n is the number of data points in each group. After the loss function is determined, the partial derivative values of each time step, such as Eq. (10), are accumulated by using optimization strategies such as random gradient descent, to update the weights and bias:

$$\left[\frac{\partial L_t}{\partial V}, \frac{\partial L_t}{\partial b}, \frac{\partial L_t}{\partial W}, \frac{\partial L_t}{\partial U}, \frac{\partial L_t}{\partial a} \right]. \quad (10)$$

3.2 Model reliability verification

The reliability represents the ability of the model to accurately reflect the characteristics of the dataset and to predict the data information of future time nodes. To verify the reliability of the model, three methods are introduced: (1) goodness of fit R^2 ; (2) mean square error (MSE); and (3) Bayesian confidence. The model has high precision and good reliability when $R^2 \in (0 \sim 1)$ (R^2 is closer to 1, MSE smaller). The Bayesian hypothesis test method [36–38] considers the data uncertainty and intuitively verifies the model reliability.

The model prediction value is assumed as $\hat{y} = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)$, and the monitoring value is $y = (y_1, y_2, \dots, y_n)$. The residual e is expressed by formula (11), where e_j is the residual between the j th monitoring and predictive value, obeying a normal distribution

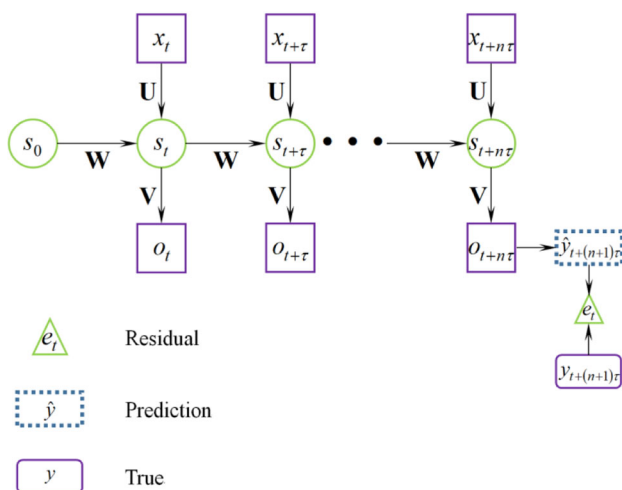


Fig. 1 Prediction model with RNN structure

$N(\mu, \sigma_1^2)$, and ε is the mean value of n residuals, obeying a normal distribution $N(\mu, \sigma_1^2/n)$:

$$e_j = y_j - \hat{y}_j, j \in 1, 2, \dots, n$$

$$\varepsilon = \frac{1}{n} \sum_{j=1}^n e_j. \quad (11)$$

The null hypothesis is defined as $H_0 : \mu = 0$, and the alternative hypothesis is $H_1 : \mu_1 \neq 0$. The prior probability density of μ is assumed as $N(0, \sigma_0^2)$. The mean value of residuals is computed in sections, and the mean variance is taken as σ_0^2 . Thus, the probability density function of μ is:

$$g(\mu) = \sqrt{\frac{1}{2\pi\sigma_0^2}} \exp\left\{-\frac{\mu^2}{2\sigma_0^2}\right\}. \quad (12)$$

The Bayesian factor is the primary evaluation index of the Bayesian hypothesis test, which is the ratio of prior and posterior probabilities. If a Bayesian factor is significantly greater than 1, the sample information supports the null hypothesis H_0 and is expressed as:

$$B^\pi(\bar{\varepsilon}) = \sqrt{1 + \frac{n\sigma_0^2}{\sigma_1^2}} \exp\left\{\frac{n\bar{\varepsilon}^2}{2} \left(-\frac{1}{\sigma_1^2} + \frac{1}{n\sigma_0^2 + \sigma_1^2}\right)\right\}. \quad (13)$$

The posterior probability of the mean value μ can be further obtained by formula (14):

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

where λ reflects the confidence degree of the prediction model. When $\lambda \rightarrow 0$, the confidence degree of the support model is 0% and the model reliability is low. When $\lambda \rightarrow \infty$, the confidence degree of the support model is 100%, and the reliability is high.

3.3 Fault prediction

In this study, part of the data is used as a training set to construct the RNN prediction model. The abnormal signal is identified by setting the threshold in advance. The threshold $e_{\max}$ is the value of the maximum residual in the training set. In the testing set, residuals between the monitoring value of each time point and the predicted value are expressed as follows:

$$\Theta_{\text{test}} = \{e_j \mid e_j = y_j - \hat{y}_j, j \in 1, 2, \dots, n\}, \quad (15)$$

and when $\Theta_{\text{test}} > e_{\max}$, an alarm will be given. In the real-time condition monitoring of the steam turbine unit, when the residual exceeds the set threshold for a long time, the unit is considered a failure.

4 Illustration

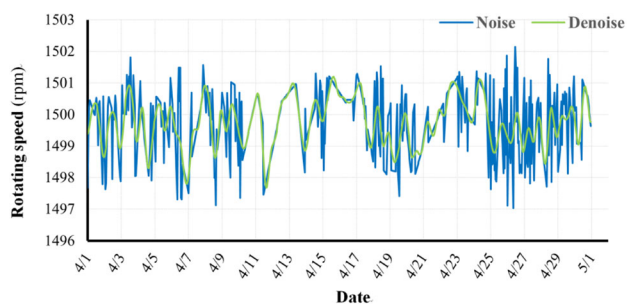
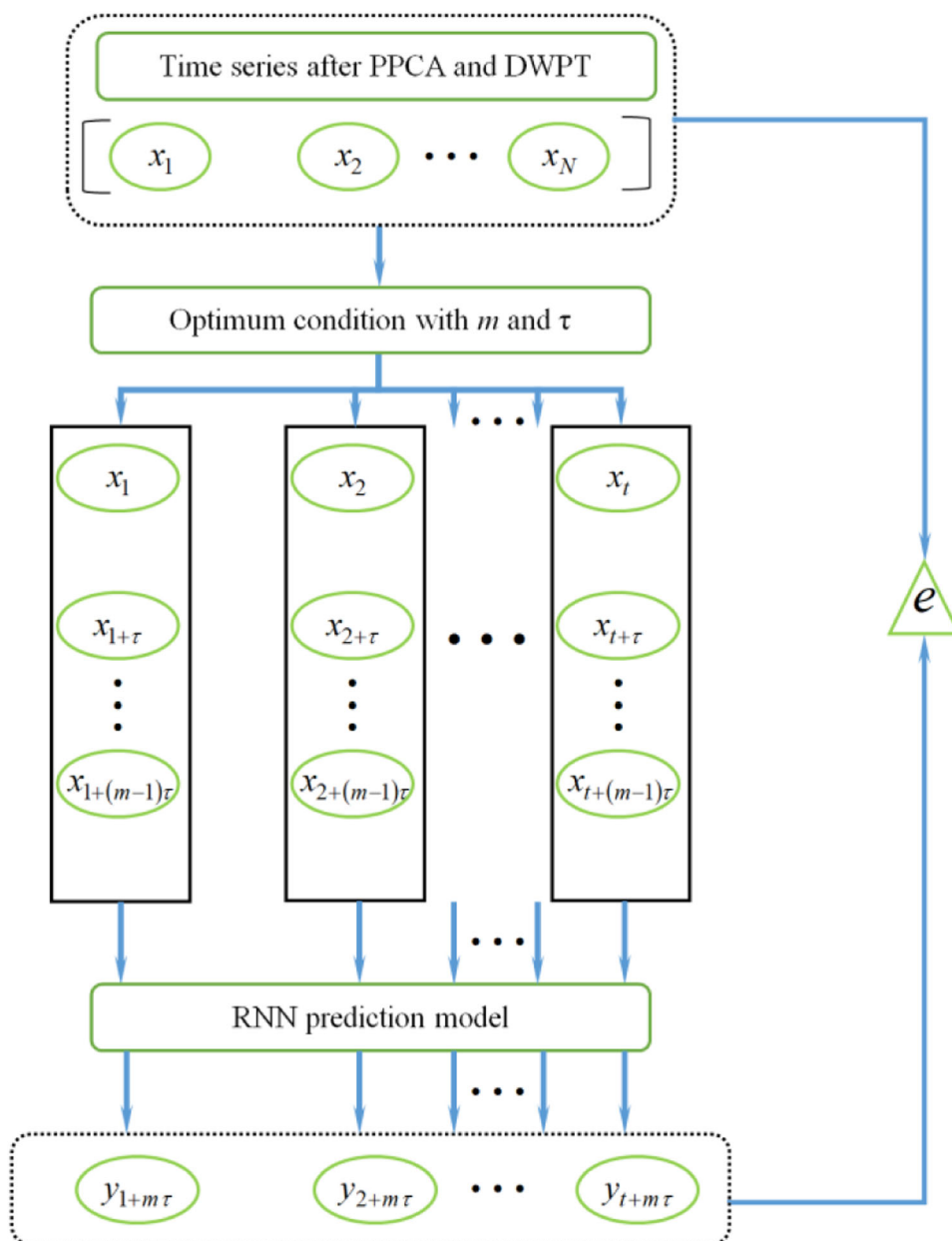
This study uses rotating speed signal data of a pressure cylinder of a nuclear power turbine in April 2019 to explain the algorithm flow and model building. The rated rotating speed of the sampled turbine is 1500 rpm. The dataset consists of two rotor speeds, one bearing group speed, and 720 time points. The data before April 20 are used as the training set to build the model and train the weights, and the data after April 20 are used as the verification set to verify the model reliability.

Figure 2 shows a flowchart of fault prediction by the RNN model. After the original three-dimensional rotational speed signal is denoised by DWPT and reduced by PPCA, a one-dimensional time series signal is obtained. The signal data with the delay time and embedding dimension optimized by the enumeration method are set as a training dataset. The RNN model after training characteristic parameters is used to predict the value of the next time point, and the residual between the prediction and monitoring values is calculated for early warning.

4.1 Data denoising and dimensional reduction

Three speed signals are denoised by the DWPT. The time series signal is decomposed into three levels by using the db8 wavelet packet, and the wavelet coefficients of each point are obtained. The wavelet coefficients are filtered according to the Bayesian threshold approach, and the signal is reconstructed. Figure 3 shows the noise and denoising data of the bearing group speed signal. The denoising signal is very similar to the original signal, and the trend is consistent. The feature information of the original signal is retained.

The dimensions of each type of PPCA data are reduced to more than 70% of the cumulative variance contribution rate of the retained dimension. Thus, the proportion of information after dimensional reduction is more than 70%. Three speed signals are dimensionally reduced in this case. Table 1 shows the results of dimensional reduction, where w_i is the weight value of each component signal to the principal component after Bayesian PPCA of the rotating speed signals. The component signal contributes more to the result analysis for greater absolute values of the weight parameter. The cost of training calculation can be decreased, and one-dimensional data can retain most of the information from the original data without any signal distortion by reducing the three-dimensional rotating speed signal to the one-dimensional signal corresponding to PC1 (Table 1).

Fig. 2 Fault prediction by the RNN model**Fig. 3** Noise reduction effect of the bearing group signal (Color online)**Table 1** PCA weight and contribution rate of rotating speeds

Weight	PC1	PC2	PC3
w_1	0.364	- 0.931	- 0.005
w_2	0.658	0.261	- 0.706
w_3	0.659	0.254	0.708
Contribution rate	0.717	0.275	0.008
Cumulative contribution rate	0.717	0.992	1.000

4.2 Determining the embedding dimension and time delay of the input layer

The time series data after dimensional reduction by PPCA are prone to chaos; therefore, the input layer must be determined by phase space reconstruction for the prediction model. By using the enumeration method with various time delay and embedding dimension combinations, the changes in the R^2 and MSE parameters of the RNN model training are analyzed. Figures 4 and 5 show the changing trends of R^2 and MSE with different time delays and embedding dimensions, respectively. R^2 and MSE slowly decrease and increase, respectively, with an increase in

time delay in the training set; in the verification set, R^2 descends steeply, and MSE increases (Fig. 4). MSE and R^2 change in waves as the embedding dimension increases (Fig. 5). Additionally, R^2 reaches a maximum at $m = 4$ in the training and verification sets, and MSE reaches a minimum at $m = 4, 7, 9$ and $m = 4, 14$ in the training and verification sets, respectively. Thereafter, to ensure the maximum value of R^2 and the minimum value of MSE, the optimal embedding dimension $m = 4$ is chosen. For the optimal time delays, $t = 1$ and $t > 1$ indicate single-step and multi-step training, respectively. Larger t values predict data a later times, but a t of 2 or 3 provides optimum accuracy (Fig. 4).

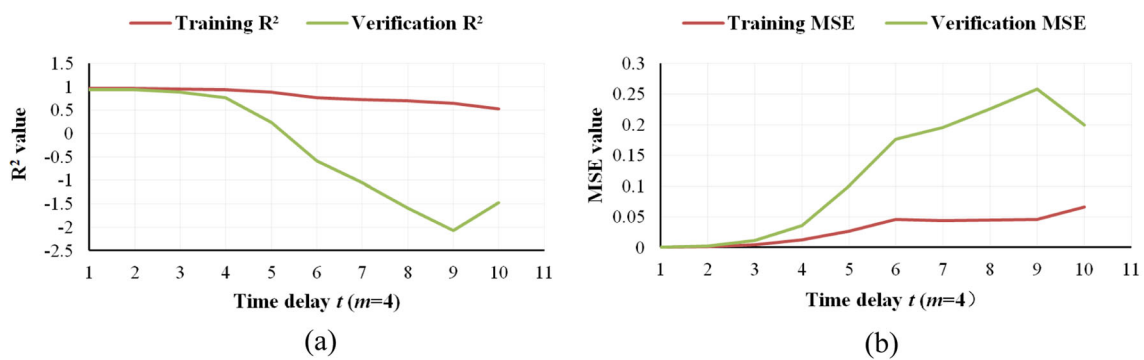


Fig. 4 R^2 (a) and MSE (b) under different time delays (Color online)

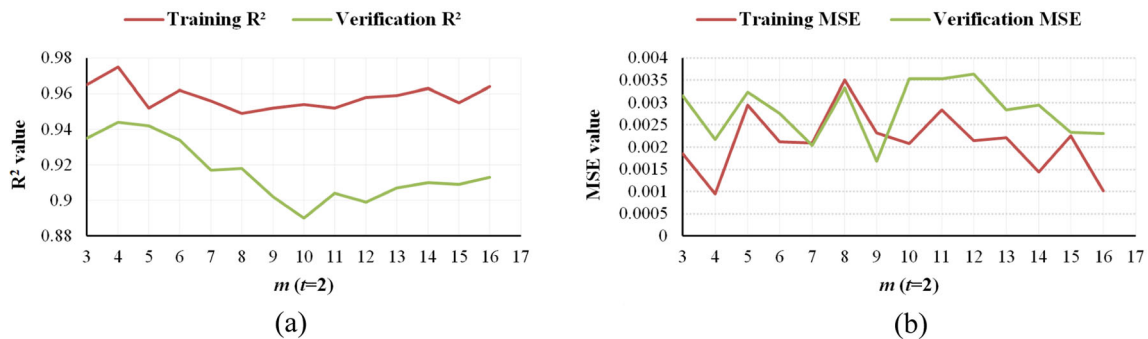


Fig. 5 R^2 (a) and MSE (b) under different embedding dimensions (Color online)

Table 2 Values of R^2 , MSE, and λ

Time delay	Embedding dimension	Training set		Verification set		
		R^2	MSE	R^2	MSE	λ
$\tau_d = 2$	$m = 3$	0.965	0.001855	0.935	0.003154	93.55%
	$m = 4$	0.975	0.000951	0.944	0.002177	93.64%
	$m = 5$	0.952	0.002946	0.942	0.003242	92.37%
	$m = 6$	0.962	0.002127	0.934	0.002755	93.27%
	$m = 7$	0.956	0.002099	0.917	0.002046	93.45%
	$m = 8$	0.949	0.003509	0.918	0.003326	91.93%

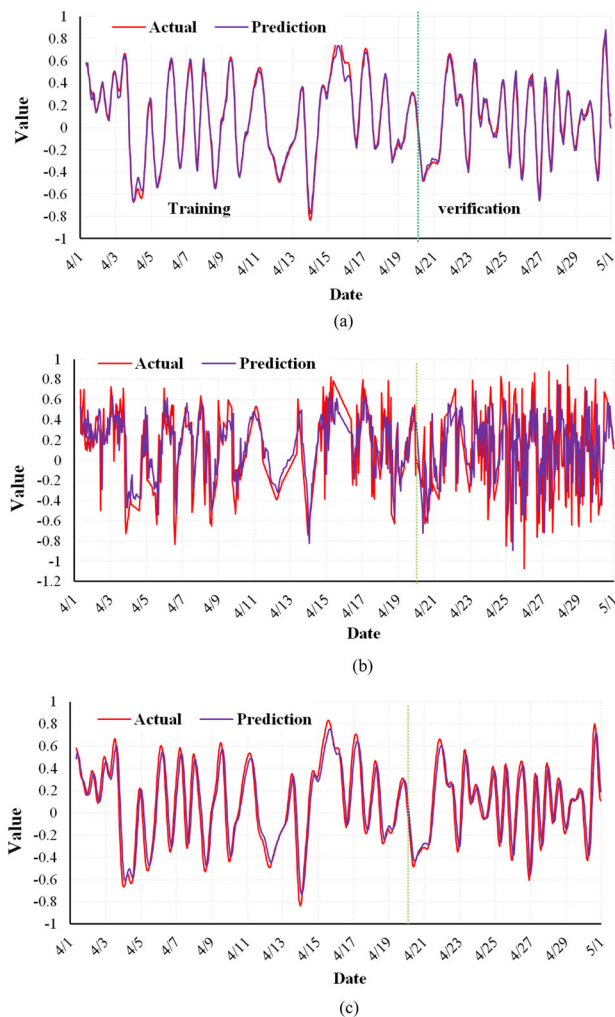


Fig. 6 Prediction and monitoring values of rotating speed signal. (a) Denoise data in RNN model; (b) noise data in RNN model; (c) denoise data in ANN model (Color online)

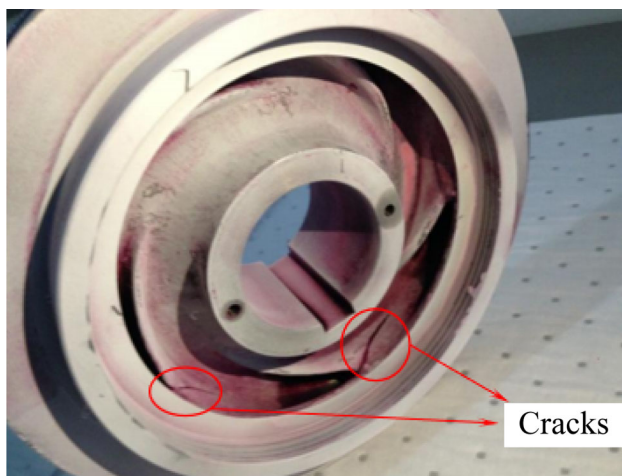


Fig. 7 Cracks in a two-stage impeller (Color online)

4.3 RNN model prediction and verification

The fully connected RNN model is constructed according to the optimized time delay and embedding dimension. For the rotating speed principal component, the time delay is the interval time between two adjacent input data, and the embedding dimension is the number of RNN input units. After model training, the R^2 , MSE, and Bayesian confidence λ indexes are employed for training and verification datasets to validate the accuracy and reliability of the model. When R^2 is closer to 1, MSE is smaller, λ is higher, and the model has high accuracy and good reliability. Table 2 lists the values of R^2 , MSE, and λ in the case of time delay $\tau_d = 2$ and embedding dimension $m = 3, 4, \dots, 8$, respectively. Under the parameters $\tau_d = 2$, $m = 4$, the R^2 values of the training set and verification set are both above 0.94, MSE values are maintained at the order of 0.001 magnitudes, and the λ values of the verification set are stable at 93% (Table 2, Figs. 4 and 5).

Figure 6 shows the comparison between the monitoring and predictive values of the rotating speed signal under three different conditions. Figure 6a illustrates the results of the proposed model and shows that the two curves coincide well. To analyze the effect of noise reduction, the rotating speed signals without DWPT are trained and predicted (Fig. 6b). The difference between two curves in the figure indicates that noise reduction of the source data is necessary. Furthermore, the traditional artificial neural network (ANN) model is applied for comparative analysis, and the predictive results are described in Fig. 6c. The degree of agreement between the two curves is lower than that of the RNN model in Fig. 6a. The R^2 values of the training and verification sets of the ANN model are 0.963 and 0.933, respectively, and the MSE values of the training and verification sets of the ANN model are 0.017104 and 0.022082, respectively. The results of these two parameters of the RNN model under $m = 4$, $\tau_d = 2$ in Table 2 are better than those of the ANN model. Thus, the accuracy and reliability of the proposed model are validated.

4.4 Fault prediction of a two-stage impeller

A dataset with fault points is applied to further verify the model reliability and test the early warning function for faults with the model. Figure 7 shows cracks (found in late May 2017 during major maintenance) in a two-stage impeller of a turbine unit in an NPP. The flaw detection results show that five blades have cracks, of which the shortest and longest measure approximately 29 and 45 mm, respectively.

Table 3 RNN model parameters and fault early warning parameters

Fault dataset ($m = 4, \tau_d = 2$)	$R^2_{(training)}$	MSE (training)	Positive threshold	Negative threshold	System alarm time	RNN model alarm time
Rotating speed	0.967	0.001248	0.043382	− 0.173924	3/7 17:00	3/5 05:00
Vibration	0.944	0.001619	0.139959	− 0.205049	3/7 17:00	3/5 21:00

Fig. 8 Fault early warning effect of rotating speed signal (Color online)

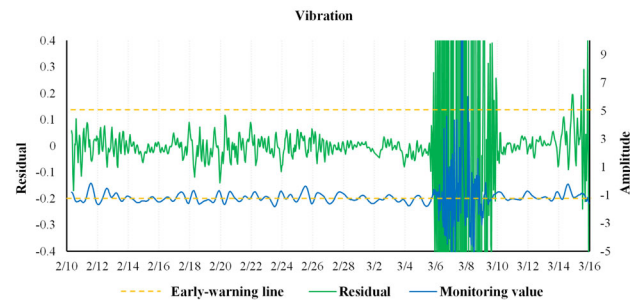


Fig. 9 Fault early warning effect of vibration signal (Color online)

The rotating speed and vibration signals from February 10 to March 16, 2017, are extracted from the monitoring system as the experimental dataset. Both the speed and vibration datasets consist of 34 days of monitoring data with 24 time points per day and a total of 816 time points. The signal data are divided into training (February 10 to March 2), verification (March 2–4), and testing (after March 5) datasets for modeling and fault prediction. Similar to the previous case, the two datasets are employed for the RNN model prediction after noise reduction and reconstruction of the embedding dimension and time delay. Table 3 lists some model prediction parameters and fault early warning parameters. The results of R^2 and MSE in the training set indicate good RNN model reliability. The positive threshold, which is the maximum positive residual allowed in the early warning system, is set using 1.2 times the maximum residual in the training set. The negative

threshold, which is the maximum negative residual allowed in the early warning system, is set using 1.2 times the minimum residual in the training set. Table 3 shows the values of the positive and negative thresholds. Two warning lines are drawn in orange (Figs. 8 and 9). When the green residual curve exceeds the warning line, the blue monitoring signal curve still fluctuates normally (Fig. 8). The monitoring system, which sets fixed alarm thresholds for specific signal values instead of residuals, was recorded to give an alarm at 17:00 on March 7, whereas the alarm time of the RNN model is 5:00 on March 5 (60 h in advance). Similarly, Fig. 9 shows the early warning effect of another vibration signal. Table 3 lists the corresponding alarm times of the RNN model and monitoring system. The RNN model produces an alarm 44 h in advance. The experimental results indicate that the cracks in the two-stage impeller cause the abnormal signals of turbine speed and vibration; the RNN model can predict the anomaly well. The model is further validated as reliable.

5 Conclusion

Because of multiple variable redundancies and turbine data imperfections, the Bayesian PPCA method is used to preprocess the DWPT denoising data and obtain a dataset with a high signal-to-noise ratio and low dimension. The rotating speed signal is reduced from a three- to one-

dimensional principal component, and the contribution rate is more than 70%.

A fully connected RNN prediction model is established. The goodness of fit of each signal data is calculated to be higher than 0.93, and MSE fluctuates on the order of 0.001, which verifies the model reliability. Furthermore, a Bayesian hypothesis testing method, which considers the data uncertainty and prior information of the training set, is employed to quantify the model confidence. The Bayesian confidence values of the verification set under different embedding dimensions are calculated at more than 90%. In the comparison case study of the two-stage impeller cracking and the monitoring system, the RNN prediction model produces alarms 60 and 44 h in advance for the rotating speed and vibration signals, respectively. The prediction results indicate that the RNN model can effectively identify faults during the creep period.

References

1. J.P. Ma, J. Jiang, Applications of fault detection and diagnosis methods in nuclear power plants: a review. *Prog. Nucl. Energy* **53**, 255–266 (2011). <https://doi.org/10.1016/j.pnucene.2010.12.001>
2. S. Mandal, B. Santhi, S. Sridhar et al., Nuclear power plant thermocouple sensor-fault detection and classification using deep learning and generalized likelihood ratio test. *IEEE Trans. Nucl. Sci.* **64**, 1526–1534 (2017). <https://doi.org/10.1109/TNS.2017.2697919>
3. Y.J. Gong, X.Y. Su, H. Qian et al., Research on fault diagnosis methods for the reactor coolant system of nuclear power plant based on D–S evidence theory. *Ann. Nucl. Energy* **112**, 395–399 (2018). <https://doi.org/10.1016/j.anucene.2017.10.026>
4. D. Wu, J. Wang, S.J. Du et al., Validation of the discharge force analysis method and key influence factors study. *Nucl. Tech.* **43**(4), 040007 (2020). <https://doi.org/10.11889/j.0253-3219.2020.hjs.43.040007>. (in Chinese)
5. W.F. Lyu, M.L. Chen, Q.Q. Huang et al., Development on calculation code CPAP for radioactive activation product of pressurized water reactor nuclear power plant. *Nucl. Tech.* **43**(4), 040005 (2020). <https://doi.org/10.11889/j.0253-3219.2020.hjs.43.040005>. (in Chinese)
6. J.T. Mo, A multi-field coupling simulation method for rod dropping behavior based on overset mesh. *Nucl. Tech.* **43**(5), 050605 (2020). <https://doi.org/10.11889/j.0253-3219.2020.hjs.43.050605>. (in Chinese)
7. H.Y. Xie, J.X. Li, Z.Y. Yan, et al., Research of critical reactivity control online early warning technology in nuclear power plant. In: *Proceedings of the 2nd International Symposium on Software Reliability, Industrial Safety, Cyber Security and Physical Protection of Nuclear Power Plant*, Aug (2017), pp. 102–108. https://doi.org/10.1007/978-981-10-7416-5_13
8. H. Qian, Si-Yun Lin, Miao Zheng, et al. Research on intelligent early-warning system of main pipeline in nuclear power plants based on hierarchical and multidimensional fault identification method. In: *Proceedings of the International Conference on Intelligent Computing for Sustainable Energy and Environment*, Sep (2017), pp. 195–205. https://doi.org/10.1007/978-981-10-6364-0_20
9. J.H. Min, D.W. Kim, C.Y. Park, Demonstration of the validity of the early warning in online monitoring system for nuclear power plants. *Nucl. Eng. Des.* **349**, 56–62 (2019). <https://doi.org/10.1016/j.nucengdes.2019.04.028>
10. B.S. Peng, H. Xia, Y.K. Liu et al., Research on intelligent fault diagnosis method for nuclear power plant based on correlation analysis and deep belief network. *Prog. Nucl. Energy* **108**, 419–427 (2018). <https://doi.org/10.1016/j.pnucene.2018.06.003>
11. Y.T. Yao, J. Wang, M. Xie, L.Q. Hu et al., A new approach for fault diagnosis with full-scope simulator based on state information imaging in nuclear power plant. *Ann. Nucl. Energy* **141**, 107274 (2020). <https://doi.org/10.1016/j.anucene.2019.107274>
12. A.S. Qin, H. Qin, Y.R. Lv et al., Concurrent fault diagnosis based on Bayesian discriminating analysis and time series analysis with dimensionless parameters. *IEEE Sens. J.* **19**, 2254–2265 (2019). <https://doi.org/10.1109/JSEN.2018.2885377>
13. M. Mehrjoo, M.J. Jozani, M. Pawlak, Wind turbine power curve modeling for reliable power prediction using monotonic regression. *Renew. Energy* **147**, 214–222 (2020). <https://doi.org/10.1016/j.renene.2019.08.060>
14. S.A. Aye, P.S. Heyns, An integrated Gaussian process regression for prediction of remaining useful life of slow speed bearings based on acoustic emission. *Mech. Syst. Signal Process.* **84**, 485–498 (2017). <https://doi.org/10.1016/j.ymssp.2016.07.039>
15. Y.C. Jiang, S. Yin, Recursive total principle component regression based fault detection and its application to vehicular cyber-physical systems. *IEEE Trans. Ind. Inform.* **14**, 1415–1423 (2018). <https://doi.org/10.1109/TII.2017.2752709>
16. T.Y. Gao, S. Yin, J.B. Qiu et al., A partial least squares aided intelligent model predictive control approach. *IEEE Trans. Syst. Man. Cybern-Syst.* **48**, 2013–2021 (2018). <https://doi.org/10.1109/TSMC.2017.2723017>
17. J. Herp, M.H. Ramezani, M. Bach-Andersen et al., Bayesian state prediction of wind turbine bearing failure. *Renew. Energy* **116**, 164–172 (2018). <https://doi.org/10.1016/j.renene.2017.02.069>
18. J.C. Li, Y.P. Cao, Y.L. Ying et al., A rolling element bearing fault diagnosis approach based on multifractal theory and gray relation theory. *PLoS ONE* **11**, e0167587 (2016). <https://doi.org/10.1371/journal.pone.0167587>
19. J. Liu, J.F. Liu, D.R. Yu et al., Fault detection for gas turbine hot components based on a convolutional neural network. *Energies* **11**, 2149 (2018). <https://doi.org/10.3390/en11082149>
20. Z.M. Liu, I.A. Karimi, Gas turbine performance prediction via machine learning. *Energy* **192**, 116627 (2020). <https://doi.org/10.1016/j.energy.2019.116627>
21. M. He, D. He, Deep Learning based approach for bearing fault diagnosis. *IEEE Trans. Ind. Appl.* **53**, 3057–3065 (2017). <https://doi.org/10.1109/TIA.2017.2661250>
22. H.S. Zhao, H.H. Liu, W.J. Hu et al., Anomaly detection and fault analysis of wind turbine components based on deep learning network. *Renew. Energy* **127**, 825–834 (2018). <https://doi.org/10.1109/TIA.2017.2661250>
23. L. Wen, L. Gao, X.Y. Li, A new deep transfer learning based on sparse auto-encoder for fault diagnosis. *IEEE Trans. Syst. Man Cybern-Syst.* **49**, 136–144 (2019). <https://doi.org/10.1109/TSMC.2017.2754287>
24. H. Liu, J.Z. Zhou, Y. Zheng et al., Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Trans.* **77**, 167–178 (2018). <https://doi.org/10.1016/j.isatra.2018.04.005>
25. H. Shahnazari, Fault diagnosis of nonlinear systems using recurrent neural networks. *Chem. Eng. Res. Des.* **153**, 233–245 (2020). <https://doi.org/10.1016/j.cherd.2019.09.026>
26. J.N. Wang, Y. Dou, Z.H. Wang et al., Multi-fault diagnosis method for wind power generation system based on recurrent

- neural network. *Proc. Inst. Mech. Eng. A-J Pow.* **233**, 604–615 (2019). <https://doi.org/10.1177/0957650919844065>
27. A.S. Palau, M.H. Dhada, K. Bakliwal et al., An industrial multi agent system for real-time distributed collaborative prognostics. *Eng. Appl. Artif. Intell.* **85**, 590–606 (2019). <https://doi.org/10.1016/j.engappai.2019.07.013>
 28. J.J. Wang, P.L. Fu, L.B. Zhang et al., Multilevel information fusion for induction motor fault diagnosis. *IEEE-ASME Trans. Mech.* **24**, 2139–2150 (2019). <https://doi.org/10.1109/TMECH.2019.2928967>
 29. W. Li, M.J. Peng, Y.K. Liu et al., Fault detection, identification and reconstruction of sensors in nuclear power plant with optimized PCA method. *Ann. Nucl. Energy* **113**, 105–117 (2018). <https://doi.org/10.1016/j.anucene.2017.11.009>
 30. M.R. Prusty, T. Jayanthi, J. Chakraborty et al., Feasibility of ANFIS towards multiclass event classification in PFBR considering dimensionality reduction using PCA. *Ann. Nucl. Energy* **99**, 311–320 (2017). <https://doi.org/10.1016/j.anucene.2016.09.015>
 31. G.H. Wu, J.J. Tong, L.G. Zhang et al., Framework for fault diagnosis with multi-source sensor nodes in nuclear power plants based on a Bayesian network. *Ann. Nucl. Energy* **122**, 297–308 (2018). <https://doi.org/10.1016/j.anucene.2018.08.050>
 32. R. Sharifi, R. Langari, Nonlinear sensor fault diagnosis using mixture of probabilistic PCA models. *Mech. Syst. Signal Process.* **85**, 638–650 (2017). <https://doi.org/10.1016/j.ymssp.2016.08.028>
 33. J.W. Xiang, Y.T. Zhong, H.F. Gao, Rolling element bearing fault detection using PPCA and spectral kurtosis. *Measurement* **75**, 180–191 (2015). <https://doi.org/10.1016/j.measurement.2015.07.045>
 34. M.E. Tipping, C.M. Bishop, Mixtures of probabilistic principal component analyzers. *Neural Comput.* **11**, 443–482 (1999). <https://doi.org/10.1162/089976699300016728>
 35. J.T. Liu, C.H. Wu, J.W. Wang, Gated recurrent units based neural network for time heterogeneous feedback recommendation. *Inform. Sci.* **423**, 50–65 (2018). <https://doi.org/10.1016/j.ins.2017.09.048>
 36. D. You, X. Shen, Y. Zhu et al., A quantitative validation method of kriging metamodel for injection mechanism based on Bayesian statistical inference. *Metals* **9**, 493 (2019). <https://doi.org/10.3390/met9050493>
 37. B. Cai, X. Kong, Y. Liu et al., Application of Bayesian networks in reliability evaluation. *IEEE Trans. Ind. Inform.* **15**, 2146–2157 (2019). <https://doi.org/10.1109/TII.2018.2858281>
 38. W.J. Cui, B. Cao, Y.X. Chen, Uncertainty analysis of Gaussian plume model based on Bayesian MCMC method. *Nucl. Tech.* **43**(4), 040009 (2020). <https://doi.org/10.11889/j.0253-3219.2020.hjs.43.040009>. (in Chinese)