machines

MDPI

Article

A Novel Principal Component Analysis-Informer Model for Fault Prediction of Nuclear Valves

Zhao An [1,2], Lan Cheng [2], Yuanjun Guo [3,*], Mifeng Ren [2], Wei Feng [3], Bo Sun [4], Jun Ling [5], Huanlin Chen [1], Weihua Chen [1], Yalin Luo [1] and Zhile Yang [3]

1 State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment, Shenzhen 518172, China; zhao.an@siat.ac.cn (Z.A.); chenhuanlin@cgnpc.com.cn (H.C.); chenweihua@cgnpc.com.cn (W.C.); luoyalin@cgnpc.com.cn (Y.L.)
2 School of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China; chenglan@tyut.edu.cn (L.C.); renmifeng@tyut.edu.cn (M.R.)
3 Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; wei.feng@siat.ac.cn (W.F.); zl.yang@siat.ac.cn (Z.Y.)
4 School of Automation Engineering, Shanghai University of Electric Power, Shanghai 200090, China; bo.sun@siat.ac.cn
5 School of Automation, Shanghai Jiaotong University, Shanghai 200240, China; lingjun_315@sjtu.edu.cn
* Correspondence: yj.guo@siat.ac.cn

Abstract: In this paper, a deep learning fault detection and prediction framework combining principal component analysis (PCA) and Informer is proposed to solve the problem of online monitoring of nuclear power valves which is hard to implement. More specifically, PCA plays the role of dimensionality reduction and fault feature extraction. It maps data with multi-dimensional space to low-dimensional space and extracts the main features. At the same time, the T-square and Q statistic thresholds are also provided to realize abnormal status monitoring. Meanwhile, Informer is a long-term series prediction method. It encrypts and decrypts data through the encoder and decoder to train a prediction model. Through the training of fault data, fault prediction can be realized. Experiments based on the sound waves collected from real valves can be continued, which also illustrates the effectiveness of the PCA–Informer model for fault diagnosis and fault prediction of nuclear power valves. Therefore, the online monitoring and maintenance of nuclear valves and other important equipment, without shutting down the nuclear station, can be achieved.

Keywords: fault diagnosis; fault prediction; PCA; deep learning; nuclear valves

check for updates

Citation: An, Z.; Cheng, L.; Guo, Y.; Ren, M.; Feng, W.; Sun, B.; Ling, J.; Chen, H.; Chen, W.; Luo, Y.; et al. A Novel Principal Component Analysis-Informer Model for Fault Prediction of Nuclear Valves. *Machines* **2022**, *10*, 240. https://doi.org/10.3390/machines10040240

Received: 22 February 2022
Accepted: 17 March 2022
Published: 29 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The aging and wear of nuclear power valves has become a very common problem in the field of nuclear power fault diagnosis, which is attributed to the continuous opening and closing of the valves. Therefore, online monitoring of the failure status of nuclear power valves and failure prediction is very necessary, especially in such a special environment with a high radiation of nuclear power. By monitoring early indicators of failure prospectively, emerging problems can be identified before they become more widespread, thus playing an import role for industrial systems [1,2].

In addition to that, as many industrial managers are paying more attention to predictive maintenance, which is an important way to increase productivity while keeping the cost at a lower level, the fault prediction method is another rising hot topic [3]. Especially in the circumstance of the nuclear power unit, regular maintenance can only be carried out when lockout is taken, which is very time consuming and causes a waste of nuclear equipment. Fault diagnosis and prediction with high accuracy can be an efficient precaution.

In general, fault diagnosis methods have four classifications which are model-based, signal-based, knowledge-based, and hybrid/active methods [4]. Due to the rapid developments in network technologies, many engineering systems, especially the valve in the

hydraulic system of safety-critical manufacturing systems, have created a large volume of historic data. Therefore, the fault model is easily established by the knowledge-based methods without the primary known models or signal patterns [5,6]. The advantage of this method is that it does not require any initial hypotheses about the relationship of input variables with the output; in addition, it is a method that does not rely on explicit models or signal symptoms. Generally, the methods of dealing with valve fault diagnosis problems are model-based approaches and sensor-based approaches [7–9]. In particular, internal leakage and friction in the valves are the most studied, widely expressed, and most represented valve faults [10–12]. In the beginning, the benchmark system was proposed, based on an in-depth study of the phenomena that can lead to potential faults in valve actuator systems. However, data-driven methods require less hypotheses [13,14]. Machine learning is one of the import methods for data-driven fault diagnosis, where information entropy is flexible and tolerant to the non-linearity problem, and is applied to analyze the characteristics of the signals.

Artificial neural network (ANN) is one of the most effective fault diagnosis methods. Two ANNs were used to classify eleven faulty types, such as the closed manual pressure valve, and normal models in seven or eight categories at the best vigilance values [15]. Later, a combination method for the fault detection of compressor valves based on basis pursuit (BP), wave matching, and Support vector machines (SVM) is presented [16]. A feedforward neural network was applied in the laser welding process monitoring and defects diagnosis [17]. However, ANNs are unable to generate discriminative features of faults and should be combined with feature extraction [18]. The feature extraction process is a conundrum and greatly impacts the result. In some of the recent literature, machine learning methods are unable to generate discriminative features of raw data and are always combined with the signal features extraction process [19,20].

Furthermore, prediction methods based on a deep learning structure is becoming a popular solution. In addition, it is a fully data-driven learning method that does not rely on rules-based programming. Except for these conventional machine learning algorithms, deep learning, a branch of machine learning, has been applied to many industrial applications, such as object identification and process monitoring, among others [21]. The proposed system adopted the machine learning approach, which learns the collected data and improves the prediction performance. New sensor technologies integrate machine learning, deep learning [22], and artificial intelligence [23] for structural health monitoring and sensor fault detection [24,25].

Models based on attention mechanisms are becoming popular in fault prediction. For example, the transformer is a new kind of neural architecture which encodes the input data as powerful features, and then calculates both representations and their relationship [26,27]. Latterly, Informer is a newly developed prediction model based on the transformer, which is quite suitable for long sequence time-series forecasting [28]. Many other variations are presented, such as the Reformer [29], Sparse Transformer [30], Longformer [31], and Performer [32], among others [33–35]. Many of these aforementioned models have been applied in natural language processing and prediction, sequence functions, image processing, and many other applications.

This paper proposes a PCA–Informer model for operation fault detection and prediction of nuclear valves to solve the problem of monitoring long time series. PCA is, first, applied to extract the fault features of sound and vibration signals, and the T^2 and Q statistics are calculated. Then, a deep learning model, based on Informer, is further applied to predict the likely fault of the variable worth paying attention to. The Informer mechanism, with a temporal attention encoder and a decoder, is used to realize fault prediction. In our early work, PCA and its variations, such as recursive PCA and model-based PCA, were successfully applied in electric power grid status monitoring and islanding detection [36–38]. However, without the deep learning method, the traditional method and its extensions are not exceptionally accurate, especially in situations of noisy data and small samples.

2. Theories and Methods

2.1. Valve Leakage Acoustic Signals

Acoustic signals, also called acoustic emission (AE), is defined as the transient elastic waves generated by the rapid release of energy from localized sources. The AE technique is a nondestructive testing and monitoring method of mechanical parts through the detecting and analyzing of signals. It has been found that AE signals can be produced by fluid leakage and other working status, such as cracking, rubbing, polishing, impacting, and cavitations. Therefore, abnormal working conditions can also be revealed by AE signals. The frequency varies from 100 kHz to 1 MHz, and it can be easily identified by environmental noise from the factory. In the case of internal leakage in liquid valves, the generation of pseudo source produced AE signals, which were largely attributed to the turbulence induced by the internal leakage.

Popular processing techniques for AE signals are wavelet transform, spectrum analysis, and neural networks. In this paper, a data-driven method is proposed to combine PCA and deep learning models for clustering and the prediction of fault conditions. Before this, for the purpose of characteristic extraction, time domain and frequency domain analysis are needed. When check valves work under normal conditions, the signals detected are mainly background noise, which has small magnitude and behaves stably. When leakage occurs, a sudden change in signal amplitude will appear, which is attributed to the turbulence induced by the internal leakage. The root-mean-square (RMS) value is often used to calculate the average energy contained in the raw AE signals. Assuming an AE signal as a total number of K samples, its RMS value is provided by:

$$\mathrm{AE}_{\mathrm{RMS}} = \left(\frac{1}{K} \sum_{n=0}^{K-1} x[n]^2 \right)^{\frac{1}{2}} \tag{1}$$

In order to effectively deal with noisy data and retain the overall and local important features of the original data, it is necessary to extract features from the original time-series acoustic signal before applying PCA. Frequency domain analysis can convert the original intractable time-domain signal into an easily analyzed frequency-domain signal. For frequency-domain analysis, the Fourier transform is always the first choice, through which the peak frequency, magnitude of the dominant frequency component, and the energy contained within the frequency band and other parameters can all be revealed. Usually, the power spectral density represents the distribution of the signal power over frequencies, which can be calculated as follows:

$$P[m] = \frac{T}{N} |X[m]|^2 = \frac{T}{N} \left| \sum_{n=0}^{N-1} x[n] \exp(-j2\pi mn/N) \right|^2 ; 0 \le m \le N-1 \tag{2}$$

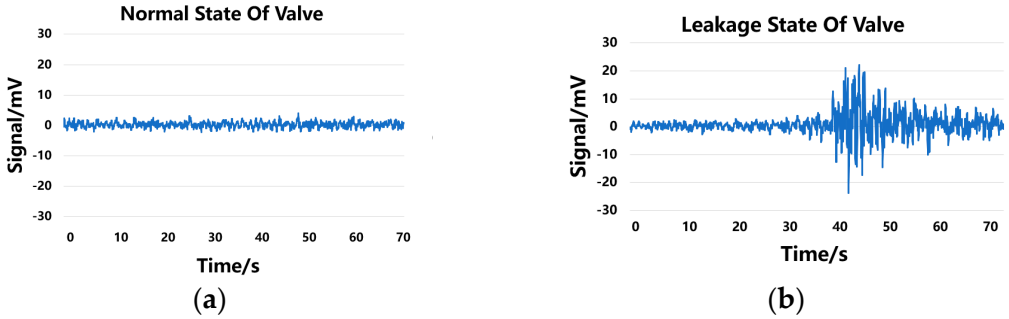
where $P[m]$ is the power spectral density, $X[m]$ is the discreate Fourier transform of an AE signal, and T is the sampling period. According to Parseval's theorem, an equation can be rearranged and yielded:

$$\mathrm{AE}_{\mathrm{RMS}}^2 = \sum_{m=0}^{N-1} P[m] \Delta F \tag{3}$$

$\Delta F = 1/(NT)$ represents the discrete frequency sample interval. $P[m]$ is the average power of the signal measured in the time domain.

2.2. Valve Leakage Represented by AE Signals

The valve internal leakage fault can be simulated by the manual valve rotation method, which is commonly used by many researchers. The AE signals of working valves under well-sealed and leakage conditions are shown as Figure 1.

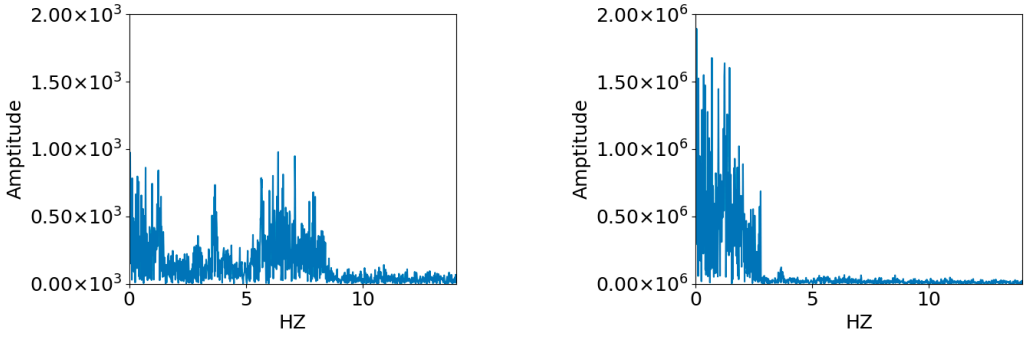


(**a**) (**b**)

Figure 1. AE signal waveform under valve well−sealed normal conditions and leakage: (**a**) normal state and (**b**) leaked state.

As evident from the time domain waveform in Figure 1, the horizontal axis represents time in s and the vertical axis represents the amplitude of the acoustic signal in mV. When the valve is closed and well−sealed, the acoustic signal is randomly distributed; when valve leakage occurs, the amplitude of acoustic signal has obvious mutations.

According to the power spectrum in the frequency domain (see Figure 2), the unit of the abscissa is HZ, and the ordinate is the amplitude. The acoustic signal energy of leakage in the valve is mainly concentrated within 1.5 Hz, while the acoustic signal energy distribution is relatively high when the valve is a normal width, distributed in the range of 0 to 8 Hz in Figure 2a. Therefore, valve leakage and normal state can be completely distinguished from the distribution of signal energy in the frequency domain. In addition, since the amplitude of the ordinate in Figure 2a is much smaller than that in Figure 2b, and we want to observe the difference in the energy distribution of the acoustic signal of the valve in two different states, that is, the distribution of the abscissa, we adjusted the appropriate ordinate. The range of the ordinate in Figure 2a is $[0, 2.0 \times 10^3]$, and the range of the ordinate in Figure 2b is $[0, 2.0 \times 10^6]$.



(**a**) (**b**)

Figure 2. AE signal spectrum in valve well−sealed normal conditions and leakage: (**a**) normal state and (**b**) leaked state.

2.3. Fault Diagnosis Method

In order to obtain features of data, PCA deposes the data matrix into a principal component subspace and a residual subspace. Among these, the principal component subspace reflects the change of standard data, and the residual subspace represents the noise of abnormal data. In other words, by projecting data into a smaller space, we reduce the dimensionality of our feature space. Consider an np data matrix $Y \in R^{n \times m}$, consisting of m variables and n samples; the Y is decomposed into the following form:

$$Y = t_1 s_1^{\mathrm{T}} + t_2 s_2^{\mathrm{T}} + \cdots + t_k s_k^{\mathrm{T}} + E = TS^{\mathrm{T}} + E \quad (4)$$

$T \in R^{n \times k}$ is the score matrix ($k \leq m$, k is the number of principal components), t_i is the score vectors, s_i is the loading vectors, and E is the residual matrix. The key statistics monitored by PCA are given by Hoteling's T^2:

$$T_i^2 = t_i \lambda^{-1} {t_i}^{\mathrm{T}} = x_i s \lambda^{-1} s^{\mathrm{T}} {x_i}^{\mathrm{T}} \tag{5}$$

Hoteling's T^2 statistic is a measure of the major variation in the variables and detects the fault if the variation in the latent variables is greater than it in the normal condition. The statistic, Q, is defined by:

$$Q_i = e_i^{\mathrm{T}} e_i = x_i^{\mathrm{T}} (I - s s^{\mathrm{T}} x_i) \tag{6}$$

The Q statistic, also named SPE, is the squared prediction error. It is a measure of the goodness of the fit of the new sample to the model. Since the T^2 statistic conforms to the F-distribution with a certain degree of confidence α (usually 99%), the statistical T_α^2 is:

$$T_\alpha^2 = \frac{n(m^2 - 1)}{m(m - n)} F(n, m - n, \alpha) \tag{7}$$

In the formula, $F(n, m - n, \alpha)$ is an F-distribution with $m - n$ degrees of freedom and the chosen level of significance α, $m - n$, α. The $100(1 - \alpha)\%$ control upper limit of Q is:

$$Q_\alpha = \theta_1 \left[\frac{\eta_\alpha \sqrt{2\theta_2 l_0^2}}{\theta_1} + 1 + \frac{\theta_2 l_0 (l_0 - 1)}{\theta_1} \right]^{\frac{1}{l_0}} \tag{8}$$

where θ_k is the sum of 1, 2, and 3 powers of the residual eigenvalues:

$$\theta_k = \sum_{i=s+1}^{m} \lambda_i^k, (k = 1, 2, 3) \tag{9}$$

l_0 is the intermediate result variable,

$$l_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2} \tag{10}$$

To monitor multi-source data in production operation, two statistical indicators, T_α^2 and Q_α, can be used. A visual interpretation can better illustrate how PCA works, as shown in Figure 3.

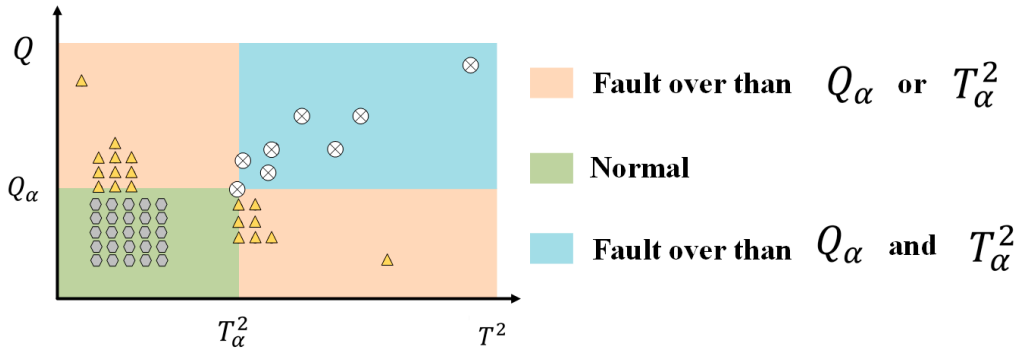


Figure 3. A geometric interpretation for fault detection using two statistics.

If some sudden abnormal deviations occur on the variables, such as a mismatch between the actual variable and base variable, the T^2 statistic is able to detect such violations. On the other hand, a change from the new data of the variables might violate the Q

threshold. In other words, for fault detection, regions in Figure 3 with data points in the red area indicate abnormal events. In order to detect faults using T^2 or SPE, the T^2-contribution and Q-contribution plot represents the importance of each variable. The Q-contribution plot is defined as:

$$\mathrm{SPE} = \left\|\widetilde{C}x\right\|^2 = \sum_{i=1}^{m} Cont_i^{\mathrm{SPE}} \tag{11}$$

$$Cont_i^{\mathrm{SPE}} = (\gamma_i^{\mathrm{T}}\widetilde{C}x)^2 \tag{12}$$

where $Cont_i^{\mathrm{SPE}}$ is the contribution of each variable to the statistic SPE, $\widetilde{C} = I - SS^T$, and γ_i^{T} is the i-th column of the identity matrix H_m. The T^2-contribution plot is defined as:

$$T^2 = (x^T R x) = \left\|R^{\frac{1}{2}}x\right\|^2 = \sum_{i=1}^{m} Cont_i^{T^2} \tag{13}$$

$$Cont_i^{T^2} = \left(\gamma_i^{\mathrm{T}} R^{\frac{1}{2}} x\right)^2 = x^T R^{\frac{1}{2}} \gamma_i \gamma_i^{\mathrm{T}} R^{\frac{1}{2}} x \tag{14}$$

where $R = P^T \Lambda^{-1} P$, γ_i^{T} is the i-th column of the identity matrix H_m, and $\|\cdot\|$ represents the norm in phase space.

When the T^2 or SPE exceeds the threshold, the contribution of the individual variables to the statistics can be identified, and the variable with the largest contribution is considered to be the potential fault source, which will be the input of the following fault prediction model.

2.4. Fault Prediction Models Using Deep Learning Method

In this section, an Informer model used for fault prediction, without ignoring relevant features across all time steps, is designed. Input $\chi^t = \left\{x_1^t, \ldots, x_L^t \middle| x_i^t \in R^{d_x}\right\}$ at time t, and the output is the predicted value $y^t = \left\{y_1^t, \ldots, y_L^t \middle| x_i^t \in R^{d_x}\right\}$. When processing a sequence with a large time span and multiple inputs, the encoding and decoding structure is used to enhance the connection of the time sequence context for prediction.

The encoder inputs x^t to hidden $H^t = \{h_1^t, \ldots, h_L^t\}$, and decodes from H^t to output ${y_{k+1}}^t$. The full scheme of the deep learning fault clustering model is shown in Figure 4.

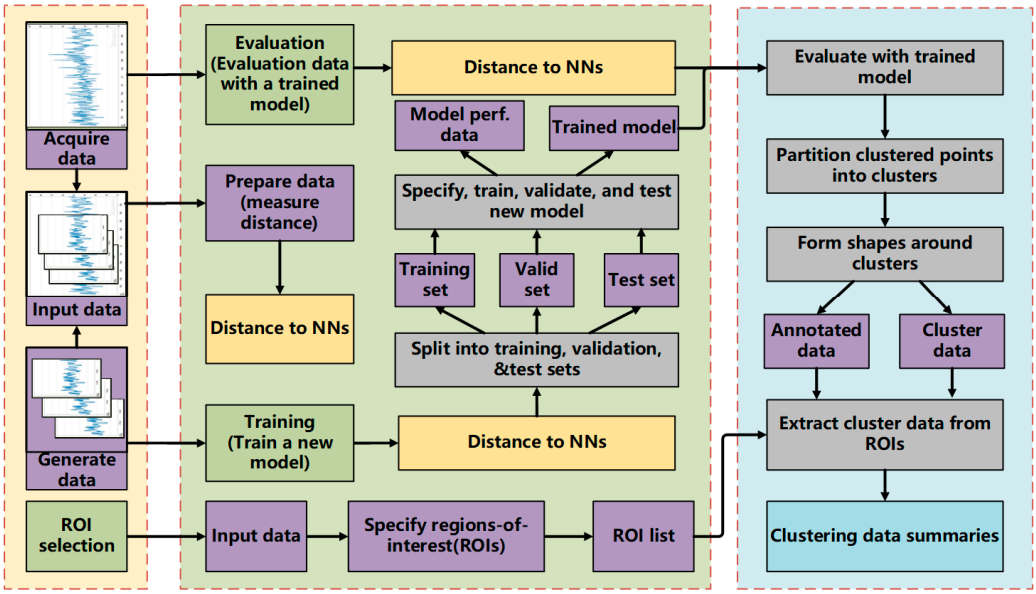


Figure 4. Deep learning fault clustering scheme.

2.4.1. Efficient Self-Attention Mechanism

The self-attention mechanism uses queries, keys, and values to perform attention calculations. The specific calculation method is as follows:

$$Attention(Q, K, V) = \text{softmax}(QK^T/\sqrt{d})V \tag{15}$$

where $Q \in R^{L_Q \times d}$, $K \in R^{L_K \times d}$, $V \in R^{L_V \times d}$, d is the input dimension. The attention of the i-th query is defined as the probabilistic form:

$$Attention(Q, K, V) = \sum_j \frac{k(q_i, k_j)}{\sum_l k(q_i, k_l)} v_j = E_{m(k_j|q)}[v_j] \tag{16}$$

To adaptively extract the driving sequence with attention weight:

$$\widetilde{x}_i = (\alpha_i^1 x_i^1, \alpha_i^2 x_i^2, \ldots, \alpha_i^n x_i^n)^{\mathrm{T}} \tag{17}$$

The hidden status at the time instant of i can be updated to:

$$h_i = f_1(h_{i-1}, \widetilde{x}_i) \tag{18}$$

f_1 is a function of LSTM units, calculated according to Equations (3)–(7). h_i is replaced with the newly calculated $\widetilde{x}_i$. Using the input attention mechanism, the encoder can selectively focus on certain driving sequences rather than treat all input driving series equally. The structure of the self-attention mechanism is shown in Figure 5.

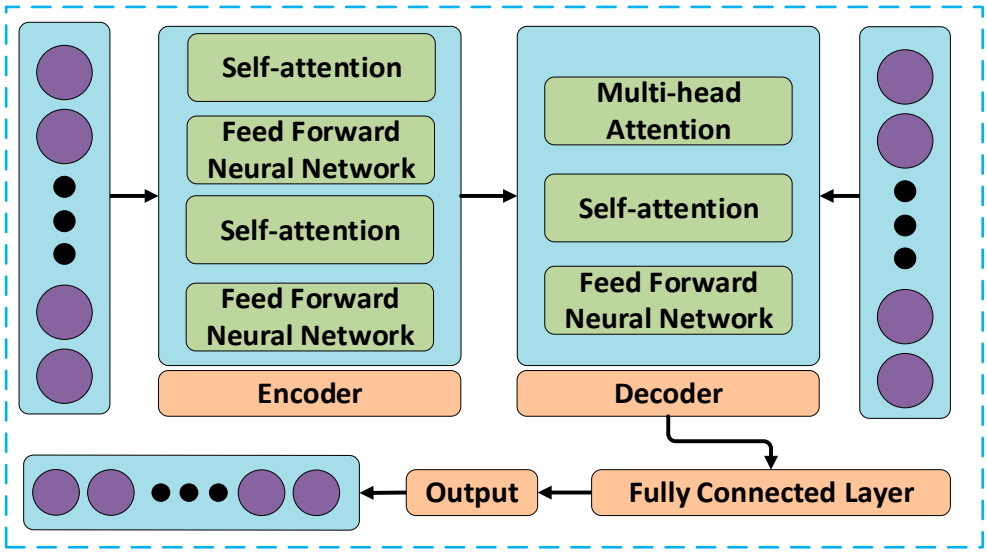


Figure 5. The architecture of encoder and decoder.

2.4.2. Decoder with Temporal Attention

In order to predict the output $\widetilde{y}$, another recurrent neural network based on LSTM is used to decode the encoded input information. The longer the input sequence, the worse the performance of the encoder and decoder. Therefore, a temporal attention mechanism is used in the decoder to adaptively select the relevant encoder hidden state across all time steps. Specifically, this is based on the previous hidden state of the decoder $d_{i-1} \in R^p$ and the $s'_{i-1} \in R^p$ of the LSTM unit to calculate the attention weight of each encoder hidden state at time i.

$$l_j^t = v_d^{\mathrm{T}} \tanh(W_d[d_{i-1}; s'_{i-1}] + U_d h_j) \quad 1 \le j \le \mathrm{T} \tag{19}$$

$$\beta_i^j = \frac{\exp(l_i^j)}{\sum_{k=1}^{T} \exp(l_i^k)} \tag{20}$$

where $[d_{i-1}; s'_{i-1}] \in R^{2p}$ is a concatenation of the previous hidden state of the LSTM unit and the unit state. $v_d \in R^m, W_d \in R^{m\times 2p}$ and $U_d \in R^{m\times m}$ are the parameters used for learning. Similarly, the offset term is also omitted here. The attention weight β_i^j represents the importance of the i-th hidden state for prediction. Since each encoder hidden state h_j is mapped to the temporal component of the input, the attention mechanism calculates the context vector c_i as all encoder hidden states $\{h_1, h_2, \ldots, h_{\mathrm{I}}\}$:

$$c_i = \sum_{j=1}^{I} \beta_i^j h_j \tag{21}$$

The context vector ci will be updated at each time step. Afterwards, the context vector is combined with the given target series $(y_1, y_2, \ldots, y_{\mathrm{I}-1})$:

$$\widetilde{y}_{i-1} = \widetilde{w}^{\mathrm{T}}[y_{i-1}, c_{i-1}] + \widetilde{b} \tag{22}$$

where $[y_{i-1}, c_{i-1}] \in R^{\mathrm{m}+1}$ is a concatenation of the decoder input $\widetilde{y}_{i-1}$ and the vector c_{i-1}. The parameters $\widetilde{w} \in R^{\mathrm{m}+1}$ and $\widetilde{b} \in R$ map the concatenation to the size of the decoder input. The newly calculated $\widetilde{y}_{i-1}$ can be used to update the hidden state of the decoder at time i:

$$d_i = f_2(d_{i-1}, \widetilde{y}_{i-1}) \tag{23}$$

The non-linear function f_2 is selected as the unit. This function is widely used in long-term dependencies. d_i can be updated as:

$$f'_t = \sigma(W'_f[d_{i-1}; \widetilde{y}_{i-1}] + b'_f) \tag{24}$$

$$K'_i = \sigma(W'_j[d_{i-1}; \widetilde{y}_{i-1}] + b'_j) \tag{25}$$

$$\mathrm{T}\,\prime_i = \sigma(W'_o[d_{i-1}; \widetilde{y}_{i-1}] + b'_o) \tag{26}$$

$$s_i' = f_i' \odot s'_{i-1} + K'_i \odot \tanh(W'_s[d_{i-1}; \widetilde{y}_{i-1}] + b'_s) \tag{27}$$

$$d_i = T'_i \odot \tanh(s'_i) \tag{28}$$

where $[d_{i-1}; \widetilde{y}_{i-1}] \in R^{p+1}$ is the concatenation of the previous hidden state d_{i-1} and decoder $\widetilde{y}_{i-1}$, $W'_f, W'_K, W'_T, W'_s \in R^{p\times(p+1)}$, $b'_f, b'_K, b'_T, b'_s \in R^p$ are the parameters to be learned. σ is a logical S-type function and element-wise multiplication. For the prediction of steam demand in paper mills, the model is used to approximate the function F to obtain an estimate of the current output $\widetilde{y}_i$, in the case of combining the input and the previous output. Specifically, it can be obtained with:

$$\widetilde{y}_I = F(y_1, \cdots, y_{I-1}, x_1, x_2, \cdots, x_I) = v_y^{\mathrm{T}}(W_y[d_I; c_I] + b_w) + b_v \tag{29}$$

where $[d_I; c_I] \in R^{p+m}$ is the concatenation of the hidden state of the decoder and the context vector. Parameters $W_y \in R^{p\times(p+m)}$ and $b_w \in R^{p+m}$ map the cascade to the size of the hidden state of the decoder. A linear function with weight $v_y \in R^{p+m}$ and bias $b_v \in R$ is used to predict the final result.

Informer uses an encoder to receive a large number of long sequences of input. In the encoder, ProbSparse is used to replace attention. We replaced the standard self-attention with the proposed ProbSparse. The attention is drawn to reduce the features extracted as the number of network layers increases to avoid the explosion of the network scale. The decoder receives a long sequence of inputs, calculates the attention of weighted features, and outputs a predicted value.

3. Experiments

3.1. Experiment Setup

In order to verify the validity and advancement of the algorithm, a group of nuclear power valve detection devices are used to collect acoustic signals. We used four sensors to acquire acoustic signals simultaneously, so each group of acquired acoustic signals had a four-dimensional signal. Thus, each group of collected acoustic signals had four-dimensions. The acquisition of data was done by the oscilloscope. First, the sensor was connected to the valve, and the output was connected to the oscilloscope. After the circuit of the valve was opened, the acoustic signal collected by the sensor was converted into an electrical signal by the analog-to-digital converter (ADC) of the oscilloscope and displayed on the screen. The acquired data was displayed in a scrolling manner, starting on the right side of the display and continuing to the left side (while acquisition was in progress). In addition, the oscilloscope has mixed-signal acquisition memory, so all acoustic data would be stored in memory while the acquisition was in progress. As shown in Figure 6, Figure 6a is a nuclear power valve, Figure 6b is a signal generator, which can input and output acoustic signals, and Figure 6c is the control circuit of the nuclear power valve, which can control the start and stop of the valve.

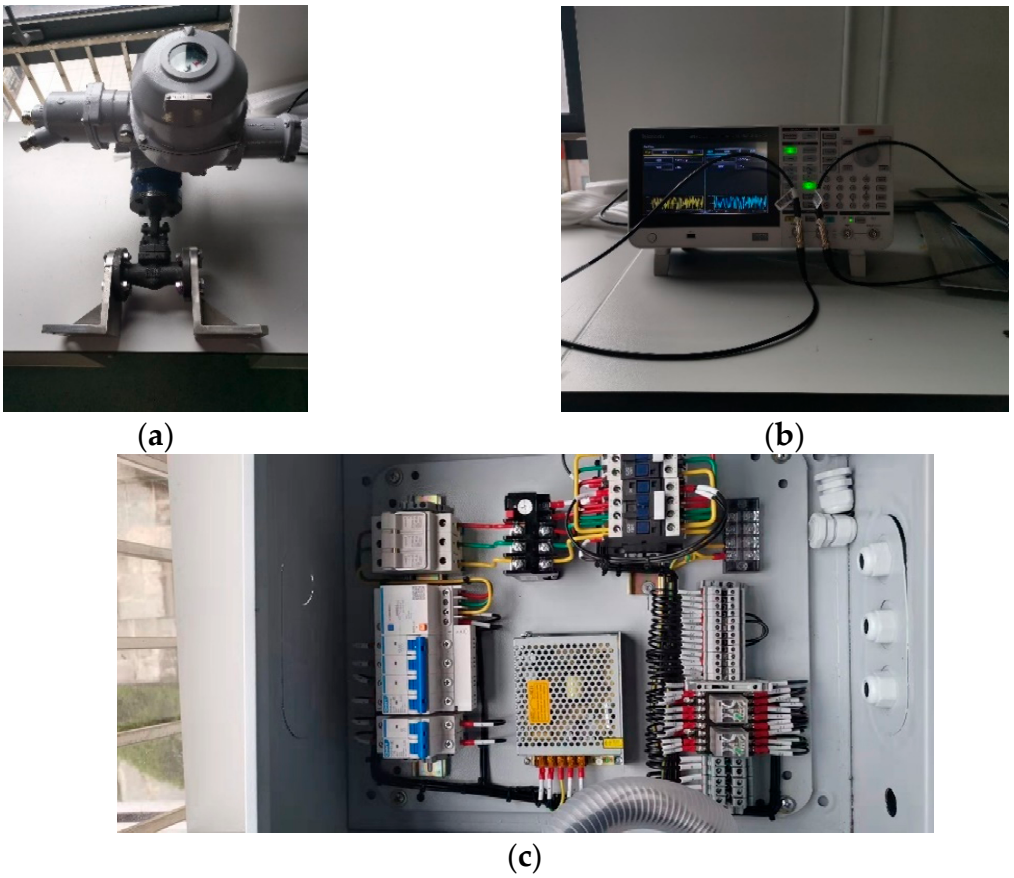


Figure 6. Nuclear power valve testing device: (**a**) nuclear power valve, (**b**) signal generator, and (**c**) control circuit of the nuclear power valve.

3.2. Data Preprocessing

For long-term sequences of acoustic data containing many noisy signals, training with raw data is feasible but direct training is usually inefficient and, thus, requires feature extraction. Fourier transform and wavelet transform are common means of signal processing, but specific problems need to be analyzed in detail. If it is a stationary periodic signal, the Fourier transform is better, and if it is a non-stationary aperiodic signal, the wavelet is better. In addition, using wavelet transform to extract features and convert

them into frequency domain images for subsequent fault classification can improve the accuracy of fault classification. Therefore, in this paper, the wavelet transform was chosen. Figure 7a–c are wavelet transform time–frequency diagrams of different types of fault signals, respectively, and Figure 7d is the wavelet transform time-frequency diagram of normal signals. After that, some unknown fault signals were input into the classifier. Since there are three types of faults, the classifier divides these input fault signals into three categories, A, B, and C, through clustering, which correspond to the Figure 7a–c three fault categories in Figure 7. Since there are three fault categories, we cluster the three different faults into three categories A, B, and C, corresponding to the three categories A, B, and C in Figure 8, respectively.

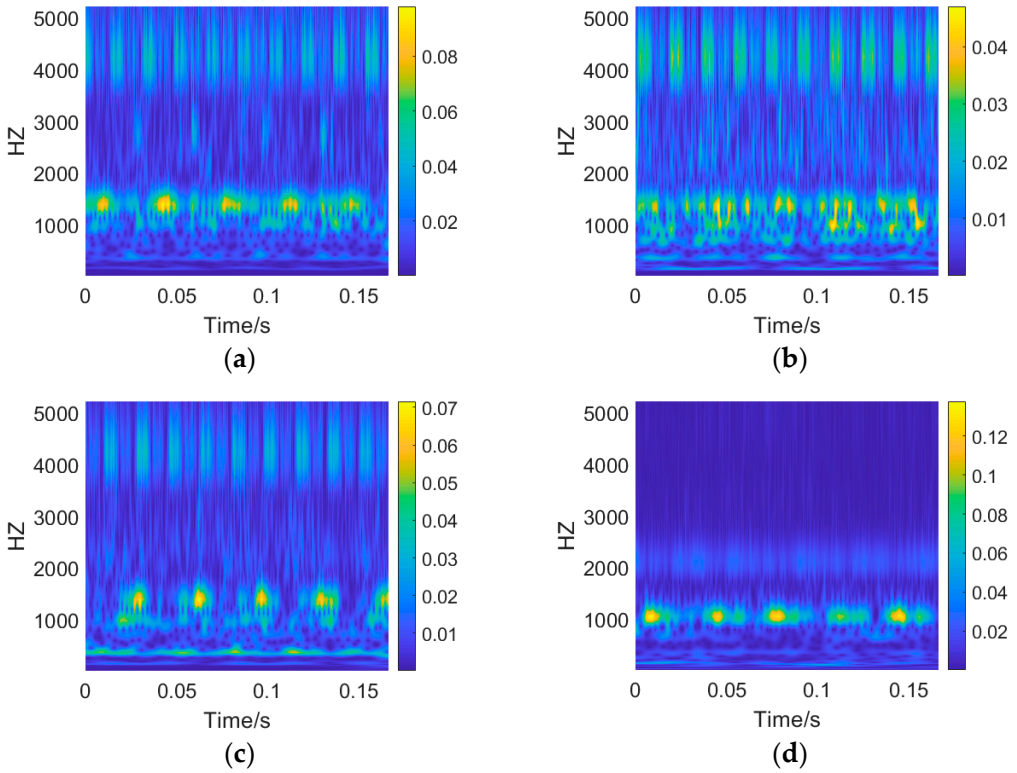


Figure 7. Wavelet transform time–frequency diagram of fault and normal state time series for: (**a**) valve body leakage fault, (**b**) sealing surface leakage fault, (**c**) electric device fault and (**d**) normal signals.

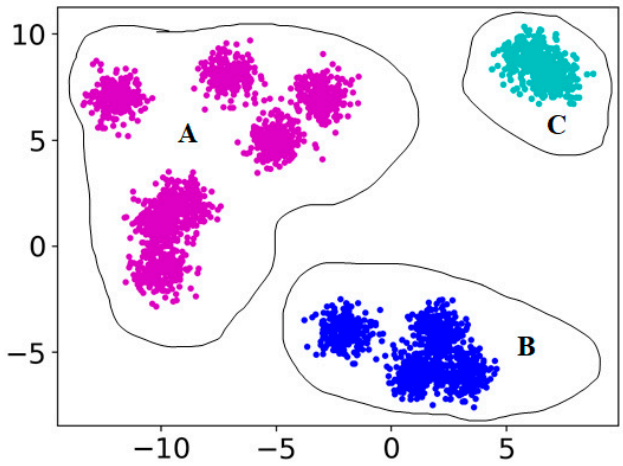


Figure 8. Fault feature visualization.

3.3. Fault Detection Results for Electric Valves

Since the valve acoustic signal of each fault state was collected simultaneously by four sensors during the experiment, each group of acoustic signals had four dimensions. For different kinds of acoustic signals, a set of 4D acoustic wave data, were selected as the training data for training the PCA model, and another set of 4D acoustic wave data were selected as test data to test the training model. For multi-dimensional time-series signals, it is difficult for us to distinguish small differences through human eyes. PCA can analyze the main components in the training data and reveal its main features. By computing the thresholds of the T^2 and SPE statistics of the training data, if other classes of test data are fed into PCA, the thresholds will be exceeded and failure detection will succeed. Therefore, in this paper, we applied PCA for fault detection, and the train data and test data are shown in Figure 9a,b.

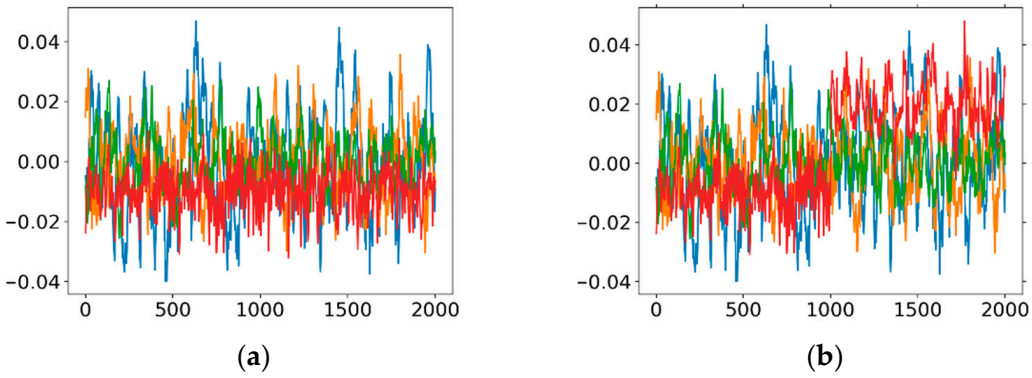


Figure 9. Fault detection results: (**a**) training data set and (**b**) testing data.

Figure 10 shows the results of sorting the eigenvalues and outputting them in descending order during the PCA modelling. It can be seen from the figure that the gap between the first three eigenvalues and the last eigenvalue is too large. Therefore, it can be considered that the last eigenvalue is less important, and, thus, the first three eigenvalues are taken as the principal components. This process can be understood as dimensionality reduction, which can compress data while minimizing information loss. For fault detection, it can compress data and extract the features at the same time.

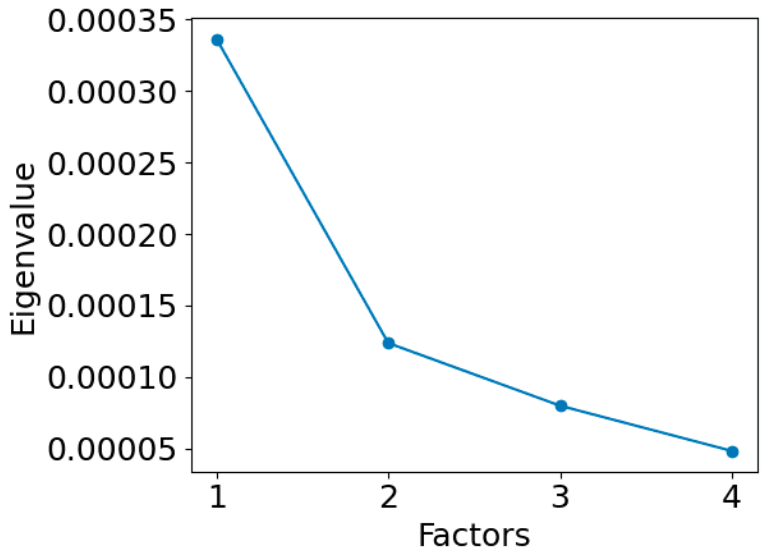


Figure 10. Sorted eigenvalues with descending order.

We selected three principal components, namely three eigenvalues. From the perspective of vector space, because the eigenvectors corresponding to different eigenvalues are linearly independent, each eigenvector is regarded as a coordinate axis, and the eigenvalue is the coordinate value of the corresponding coordinate axis (i.e., the eigenvector). In simple terms, the original matrix is represented by eigenvalues (coordinates) and eigenvectors (coordinate axes). Since we cannot plot the four-dimensional space, to better understand the relationship between the eigenvectors, Figure 11 shows a heatmap formed by the matrix of eigenvectors corresponding to the selected principal components.

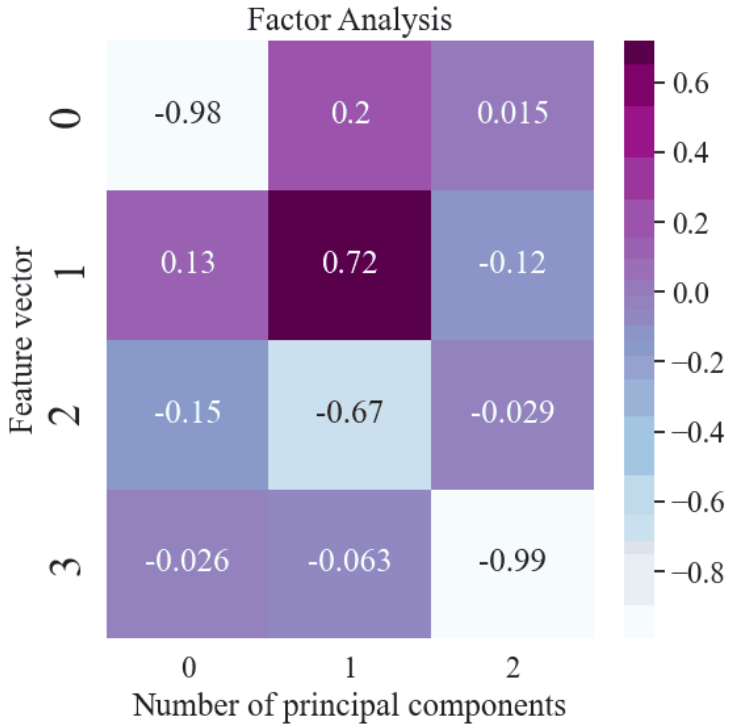


Figure 11. Eigenvector matrix corresponding to principal components.

The thresholds for the T^2 and SPE statistics can be obtained from the training data by PCA, and T^2 and SPE statistics can be obtained from test data by PCA. The T^2 statistic reflects the degree to which each principal component deviates from the model, in terms of variation trend and amplitude. It is a measure of the internalization of the model, and it can be used to monitor multiple principal components at the same time. The SPE statistic describes the degree of deviation of the measured value of the input variable from the principal component model, which is a measure of the external changes of the model. Figure 12 shows the results of PCA fault detection, where the red straight line represents the thresholds for the T^2 and SPE statistics. If the T^2 or SPE statistics exceed the thresholds, it means a fault has occurred and the red dotted rectangular box area is marked as the fault state.

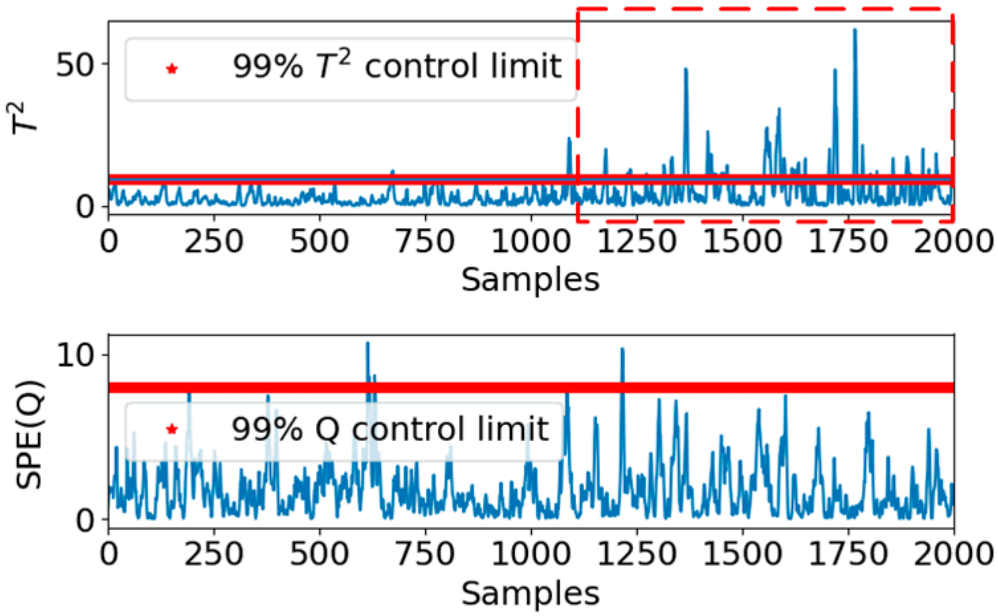


Figure 12. The results of principal component analysis.

3.4. Fault Prediction Results for Electric Valves

In Figure 13, in order to verify the validity of the Informer model for fault prediction, six different fault categories were used for the experiments, and the results of the actual and predicted values are shown as well. The red dotted line represents the predicted value, and the black solid line represents the true value. In addition, the performance evaluation indicators of each group of experimental results have been listed in Table 1. It can be seen from Figure 13 that the change trend of the predicted value is consistent with the actual value, and the change trend of the time-series signal can be effectively predicted. However, from the evaluation index between the predicted value and the actual value in Table 1, it can be seen that there are still some differences between the predicted value and the actual value. This is because we used acoustic wave data, and the sampling noise was relatively obvious. The literature [28] also mentioned that, on different data sets, the prediction effect of the Informer model has certain differences. In the future, we will start by changing the sampling rate, reducing noise, and adjusting the parameters of the prediction model to increase the success rate of fault prediction.

Table 1. Fault prediction result evaluation index.

True & Pred	MAE	MSE	RMSE	MAPE
a	0.5464955	0.49104416	0.7007454	2.1672266
b	0.5304586	0.45938802	0.6777817	2.4918344
c	0.6368491	0.70024675	0.83680749	5.0315237
d	0.5420266	0.46397582	0.6811577	8.153066
e	0.53628075	0.47729263	0.69086367	3.9367547
f	0.56860536	0.5392031	0.7343045	5.1112056

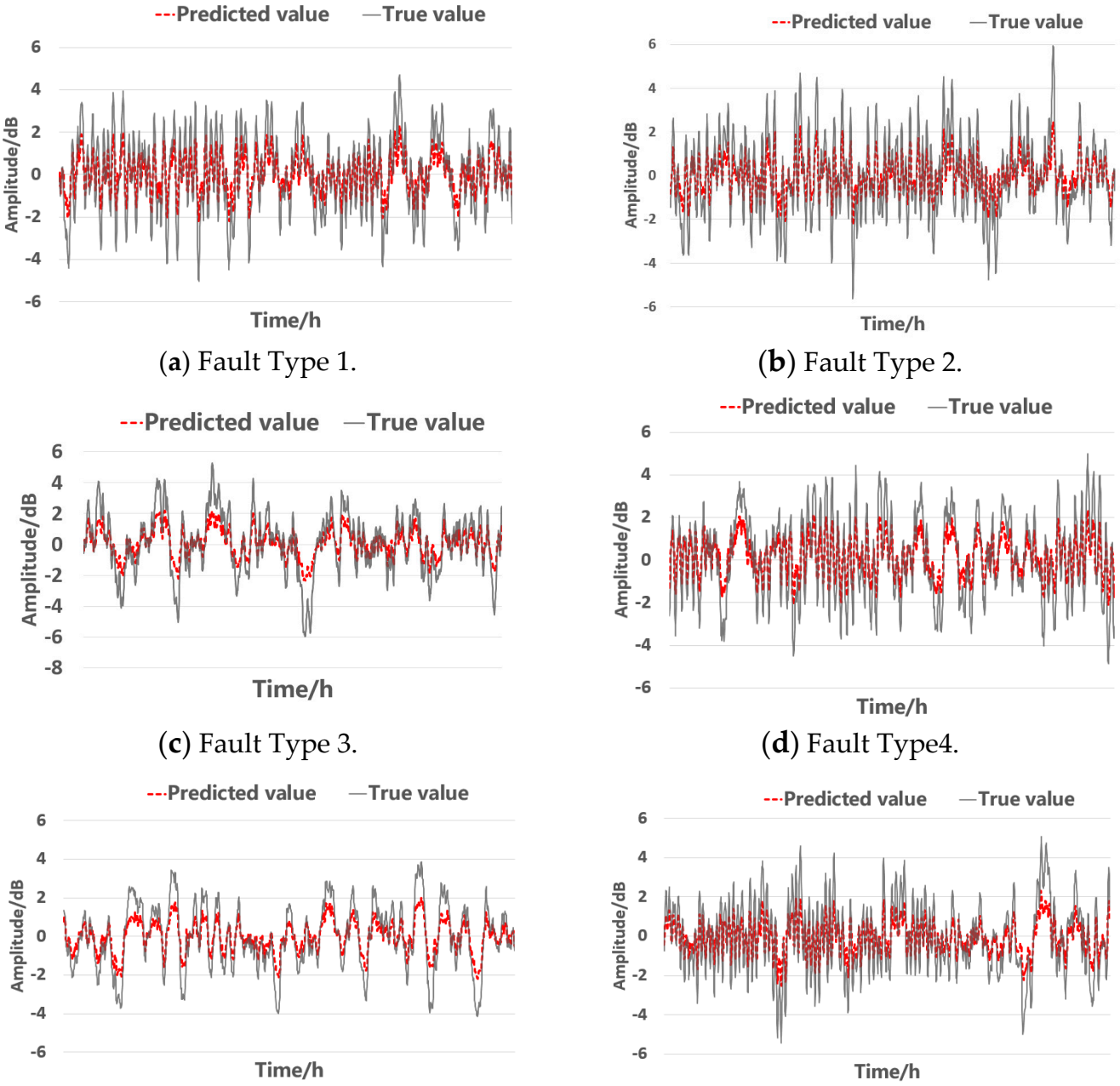


Figure 13. Long time–series fault prediction.

4. Conclusions

This paper proposes a PCA–Informer model, aiming at the fault detection and prediction of nuclear valves. The first stage selects the features that need attention, and the second stage uses an Informer attention mechanism to decode the hidden state for the prediction. Since it not only uses an input attention mechanism to relevant driving series but also employs an attention mechanism to select relevant hidden features across all time steps, the experimental results demonstrate that the PCA–Informer model can successfully achieve an accurate result in fault detection and prediction. Therefore, a safe online monitoring of key elements in nuclear stations, such as electrical valves, can be realized on a regular basis, and maintenance time and cost can be reduced. Meanwhile, fault diagnoses and predictions can be carried out with high accuracy to prevent potential hazards to operators and a waste of nuclear equipment.

Author Contributions: Conceptualization, Z.A., Y.G., B.S., W.C., Y.L. and Z.Y.; data curation, Z.A., Y.G., H.C. and W.C.; formal analysis, L.C., Y.G., B.S. and H.C.; funding acquisition, J.L., H.C., W.C., Y.L. and Z.Y.; investigation, W.F., B.S., J.L., Y.L. and Z.Y.; methodology, B.S.; project administration, M.R., W.F. and Z.Y.; resources, M.R., W.F. and J.L.; software, Z.A. and B.S.; supervision, Y.G., M.R., W.F., W.C. and Z.Y.; validation, M.R. and H.C.; visualization, Z.A.; writing—original draft, Z.A., L.C., Y.G. and W.F.; writing—review and editing, Z.A., L.C., Y.G., M.R., J.L. and Z.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment [K-A2020.407], National Science Foundation of China [52077213, 62073232 and 62003332], Joint Research Fund between the National Natural Science Foundation of China and Shenzhen [U1813222, U20A20283], The Natural Science Foundation of Shanxi Province, China [202103021231], and Guangdong special branch plans young talent with scientific and technological innovation [2019TQ05Z654].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank State Key Laboratory of Nuclear Power Safety Monitoring Technology and Equipment for their support of this research. The authors would like to thank the all the reviewers for their constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gao, Z.; Cecati, C.; Ding, S.X. A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis with Model-Based and Signal-Based Approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3757–3767. [CrossRef]
2. Lei, Y.; Yang, B.; Jiang, X.; Jia, F.; Li, N.; Nandi, A.K. Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mech. Syst. Signal Process.* **2020**, *138*, 106587. [CrossRef]
3. Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. *Mech. Syst. Signal Process.* **2018**, *108*, 33–47. [CrossRef]
4. Jiao, J.; Zhao, M.; Lin, J.; Liang, K. A comprehensive review on convolutional neural network in machine fault diagnosis. *Neurocomputing* **2020**, *417*, 36–63. [CrossRef]
5. Kordestani, M.; Zanj, A.; Orchard, M.E.; Saif, M. A modular fault diagnosis and prognosis method for hydro-control valve system based on redundancy in multi-sensor data information. *IEEE Trans. Reliab.* **2018**, *68*, 330–341. [CrossRef]
6. Liu, C.; Wang, Y.; Pan, T.; Zheng, G. Fault diagnosis of electro-hydraulic servo valve using extreme learning machine. *Int. Trans. Electr. Energy Syst.* **2020**, *30*, e12419. [CrossRef]
7. Yucesan, Y.A.; Dourado, A.; Viana, F. A survey of modeling for prognosis and health management of industrial equipment. *Adv. Eng. Inform.* **2021**, *50*, 101404. [CrossRef]
8. Xu, B.; Shen, J.; Liu, S.; Su, Q.; Zhang, J. Research and Development of Electro-hydraulic Control Valves Oriented to Industry 4.0: A Review. *Chin. J. Mech. Eng.* **2020**, *33*, 29. [CrossRef]
9. Djeziri, M.; Djedidi, O.; Benmoussa, S.; Bendahan, M.; Seguin, J.L. Failure Prognosis Based on Relevant Measurements Identification and Data-Driven Trend-Modeling: Application to a Fuel Cell System. *Processes* **2021**, *9*, 328. [CrossRef]
10. Li, X.; Ren, P.; Zhang, Z.; Jia, X.; Peng, X. A p−V Diagram Based Fault Identification for Compressor Valve by Means of Linear Discrimination Analysis. *Machines* **2022**, *10*, 53. [CrossRef]
11. Chen, H.; Xiong, Y.; Li, S.; Song, Z.; Hu, Z.; Liu, F. Multi-Sensor Data Driven with PARAFAC-IPSO-PNN for Identification of Mechanical Nonstationary Multi-Fault Mode. *Machines* **2022**, *10*, 155. [CrossRef]
12. Shi, S.; Li, G.; Chen, H.; Hu, Y.; Wang, X.; Guo, Y.; Sun, S. An efficient VRF system fault diagnosis strategy for refrigerant charge amount based on PCA and dual neural network model. *Appl. Therm. Eng.* **2018**, *129*, 1252–1262. [CrossRef]
13. Liu, Z.; Liu, J.; Huang, Y.; Li, T.; Nie, C.; Xia, Y.; Zhan, L.; Tang, Z.; Zhang, L. Fault Critical Point Prediction Method of Nuclear Gate Valve with Small Samples Based on Characteristic Analysis of Operation. *Materials* **2022**, *15*, 757. [CrossRef]
14. Xu, R.; Peng, M.; Wang, H. Study on the Condition Monitoring Technology of Electric Valve Based on Principal Component Analysis. In Proceedings of the International Congress and Workshop on Industrial AI 2021, Luleå, Sweden, 5–7 October 2021; Springer: Cham, Switzerland, 2021; pp. 141–151.
15. Han, X.; Jiang, J.; Xu, A.; Huang, X.; Pei, C.; Sun, Y. Fault Detection of Pneumatic Control Valves Based on Canonical Variate Analysis. *IEEE Sens. J.* **2021**, *21*, 13603–13615. [CrossRef]
16. Lv, Q.; Yu, X.; Ma, H.; Ye, J.; Wu, W.; Wang, X. Applications of Machine Learning to Reciprocating Compressor Fault Diagnosis: A Review. *Processes* **2021**, *9*, 909. [CrossRef]

17. Zhang, Z.; Li, B.; Zhang, W.; Lu, R.; Wada, S.; Zhang, Y. Real-time penetration state monitoring using convolutional neural network for laser welding of tailor rolled blanks. *J. Manuf. Syst.* **2020**, *54*, 348–360. [CrossRef]
18. Kumar, P.; Hati, A.S. Review on machine learning algorithm based fault detection in induction motors. *Arch. Comput. Methods Eng.* **2021**, *28*, 1929–1940. [CrossRef]
19. Jan, S.U.; Lee, Y.D.; Koo, I.S. A distributed sensor-fault detection and diagnosis framework using machine learning. *Inf. Sci.* **2021**, *547*, 777–796. [CrossRef]
20. Bode, G.; Thul, S.; Baranski, M.; Müller, D. Real-world application of machine-learning-based fault detection trained with experimental data. *Energy* **2020**, *198*, 117323. [CrossRef]
21. Yuan, H.; Wu, N.; Chen, X.; Wang, Y. Fault Diagnosis of Rolling Bearing Based on Shift Invariant Sparse Feature and Optimized Support Vector Machine. *Machines* **2021**, *9*, 98. [CrossRef]
22. Ali, Y.H. Artificial intelligence application in machine condition monitoring and fault diagnosis. *Artif. Intell. Emerg. Trends Appl.* **2018**, *275*, 464. [CrossRef]
23. Nath, A.G.; Udmale, S.S.; Singh, S.K. Role of artificial intelligence in rotor fault diagnosis: A comprehensive review. *Artif. Intell. Rev.* **2021**, *54*, 2609–2668. [CrossRef]
24. Sony, S.; LaVenture, S.; Sadhu, A. A literature review of next-generation smart sensing technology in structural health monitoring. *Struct. Control Health Monit.* **2019**, *26*, e2321. [CrossRef]
25. Ye, X.W.; Jin, T.; Yun, C.B. A review on deep learning-based structural health monitoring of civil infrastructures. *Smart Struct. Syst.* **2019**, *24*, 567–585.
26. Han, K.; Xiao, A.; Wu, E.; Guo, J.; Xu, C.; Wang, Y. Transformer in transformer. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 1–12.
27. Wu, N.; Green, B.; Ben, X.; O'Banion, S. Deep transformer models for time series forecasting: The influenza prevalence case. *arXiv* **2020**, arXiv:2001.08317.
28. Zhou, H.; Zhang, S.; Peng, J.; Zhang, S.; Li, J.; Xiong, H.; Zhang, W. Informer: Beyond efficient transformer for long sequence time-series forecasting. In Proceedings of the AAAI virtually, New York, NY, USA, 2–9 February 2021.
29. Kitaev, N.; Kaiser, Ł.; Levskaya, A. Reformer: The efficient transformer. *arXiv* **2020**, arXiv:2001.04451.
30. Correia, G.M.; Niculae, V.; Martins, A.F. Adaptively sparse transformers. *arXiv* **2019**, arXiv:1909.00015.
31. Beltagy, I.; Peters, M.E.; Cohan, A. Longformer: The long-document transformer. *arXiv* **2020**, arXiv:2004.05150.
32. Pandey, U.K.; Pal, S. Data Mining: A prediction of performer or underperformer using classification. *arXiv* **2011**, arXiv:1104.4163.
33. Child, R.; Gray, S.; Radford, A.; Sutskever, I. Generating long sequences with sparse transformers. *arXiv* **2019**, arXiv:1904.10509.
34. Dai, Z.; Yang, Z.; Yang, Y.; Carbonell, J.; Le, Q.; Salakhutdinov, R. Transformer-xl: Attentive language models beyond a fixed-length context. *arXiv* **2019**, arXiv:1901.02860.
35. Lee, E.R.; Cho, J.; Yu, K. A systematic review on model selection in high-dimensional regression. *J. Korean Stat. Soc.* **2018**, *48*, 1–12. [CrossRef]
36. Guo, Y.; Li, K.; Laverty, D.; Xue, Y. Synchrophasor-Based Islanding Detection for Distributed Generation Systems Using Systematic Principal Component Analysis Approaches. *IEEE Trans. Power Deliv.* **2015**, *30*, 2544–2552. [CrossRef]
37. Zhu, J.; Yang, Z.; Mourshed, M.; Guo, Y.; Zhou, Y.; Chang, Y.; Wei, Y.; Feng, S. Electric Vehicle Charging Load Forecasting: A Comparative Study of Deep Learning Approaches. *Energies* **2019**, *12*, 2692. [CrossRef]
38. Qian, H.; Sun, B.; Guo, Y.; Yang, Z.; Ling, J.; Feng, W. A parallel deep learning algorithm with applications in process monitoring and fault prediction. *Comput. Electr. Eng.* **2022**, *99*, 107724. [CrossRef]