

## Article

# A PCA-LSTM-Based Method for Fault Diagnosis and Data Recovery of Dry-Type Transformer Temperature Monitoring Sensor

Mingze Zheng<sup>1,2</sup>, Kun Yang<sup>1,3</sup>, Chunxue Shang<sup>1,4</sup> and Yi Luo<sup>1,3,\*</sup> 

- <sup>1</sup> GIS Technology Research Center of Resource and Environment in Western China, Ministry of Education, Yunnan Normal University, Kunming 650500, China; zhengmingze@user.ynnu.edu.cn (M.Z.); yangkun@ynnu.edu.cn (K.Y.); shangchunxue@ynnu.edu.cn (C.S.)
- <sup>2</sup> School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China
- <sup>3</sup> Faculty of Geography, Yunnan Normal University, Kunming 650500, China
- <sup>4</sup> Dean's Office, Yunnan Normal University, Kunming 650500, China
- \* Correspondence: lysisist@ynnu.edu.cn

**Abstract:** The failure that occurs during the dry-type transformer temperature monitoring sensor working will result in wrong data output, which may cause the monitor and monitoring background to respond incorrectly. To solve this problem, a fault diagnosis and data recovery algorithm based on principal component analysis (PCA), long short-term memory neural network (LSTM), and decision tree is proposed. It can realize the fault sensor location, fault diagnosis, and data recovery under dynamic processes. First, a set of temperature monitors was designed to collect the temperature inside the dry-type transformer in real-time by using the collected temperature data to build a PCA-based fault diagnosis model and a LSTM-based data recovery model. A fault location model based on a decision tree was constructed for five typical sensor fault types. Finally, the three models were constructed to obtain the sensor fault diagnosis and recovery algorithm. We then transplanted the algorithm to the temperature monitor. The experimental results showed that the recognition rate of the algorithm for different fault diagnoses of single- or multiple-sensors was above 96%. The diagnosis time was less than 1 ms. The recovery error was within 0.1 °C. The field experiments verified that the algorithm could significantly improve the stability of the monitor. Even if the sensor fails, it can also ensure that the dry-type transformer works within the normal range.

**Keywords:** fault diagnosis; data recovery; principal component analysis; long short-term memory neural network



**Citation:** Zheng, M.; Yang, K.; Shang, C.; Luo, Y. A PCA-LSTM-Based Method for Fault Diagnosis and Data Recovery of Dry-Type Transformer Temperature Monitoring Sensor. *Appl. Sci.* **2022**, *12*, 5624. <https://doi.org/10.3390/app12115624>

Academic Editor:  
Mohamed Benbouzid

Received: 16 May 2022  
Accepted: 27 May 2022  
Published: 1 June 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

With the rapid development of the world's economy and the continuous improvement in people's living standards, the electricity load in urban and rural areas has also increased. This has brought more and more applications for dry-type transformers. Whether the transformer works normally or not directly affects the entire power system. The insulation performance of dry-type transformers is closely related to their heat generation and heat dissipation performance. If the winding temperature is too high, it will cause accelerated aging of the insulating material and shorten the service life of the transformer. This will result in its economic benefits not being maximized [1]. In extreme cases, it may even cause serious accidents such as transformer fires and explosions [2]. This will not only cause equipment damage, power outages in local areas, and cause massive economic losses to the society, but may also threaten the personal safety of the relevant personnel. Therefore, it is necessary to monitor the temperature status of the transformer in real-time when the transformer is working, and to respond to the abnormality in time [3–5].

However, the sensor used in the existing dry-type transformer temperature detection equipment hardly performs any self-confirmation on its working state, that is, the sensor is always considered to be working normally. In this way, once the sensor fails, its output results will seriously deviate from reality, which may cause false alarms, affect the work of the system, and even cause catastrophic consequences [6].

In sensor fault diagnosis, the difference between the signal of the model system and that of the actual system is defined as “the residual”. Dynamic process models are applied to analyze the input and output signals. This method was first proposed by Willsky [7], but they did not give the details of the processing. There is abundant fault information in the residual signal, and the fault can be diagnosed based on an appropriate decision function or decision rule. The method is often used together with the fault estimation method to construct the sensor fault diagnosis process. By establishing a model and choosing a decision function (or rule), the evaluation function of the residual can be compared with the selected threshold function. As a result, a change is detected and a failure of the sensor system is judged [8]. Chen et al. effectively detected the initial fault of the sensor in the high-speed railway electric drive device through SPE statistics [9]. Li et al. also used SPE statistics to diagnose the faults of sensors in the actual nuclear power plant facility system to reduce the system’s false alarm rate [10]. Hanen et al. studied the fault detection and isolation of the electric-drive sensor based by improving the parity space method. A fast and simple algorithm for sensor fault detection was designed for second-order systems. The simplicity of the final algorithm led to a shorter execution time and less resource consumption in the conduction [11]. Hamed et al. proposed a sensor fault detection method based on nonlinear parity technology, which can be used in a pH neutralization system. The nonlinear fault detection and recognition algorithm can effectively detect and isolate the sensor fault on the pH channel as well as immediately and accurately detect the time of the fault occurrence [12].

In sensor data recovery, Wang B. et al. proposed a pressure sensor data recovery model based on the correlation vector machine using the normal output data before the fault [13]. Zhu T. et al. proposed a recovery method for aircraft engine sensor failure based on the least squares support vector machine (LS-SVM) [14]. Oh, B.K. et al. proposed a structure response recovery method based on the convolutional neural network. Using the strain monitoring data stably measured before data loss, a convolutional neural network (CNN) model for data recovery was constructed. In the case of sensor failure, a trained convolutional neural network was used to recover the missing strain response using functional sensors alone [15].

The application of sensor state self-confirmation technology to dry-type transformer temperature monitoring equipment has application requirements. However, most of the existing research on the fault diagnosis of sensors is only for a single sensor, and there is a lack of research and differential diagnosis for multi-sensor faults. Additionally, the research and application of fault sensor location are insufficient and the accuracy of the fault data recovery is not high, which limits the correctness of the data diagnosis.

Tamás Orosz et al. highlighted the importance of the no free lunch theorem of mathematical optimization [16]. The selection of the model must be aimed at the specific learning problem. The model can only work best if the characteristics of the model match the characteristics of the problem. Sensor data recovery is a typical time series forecasting problem. Its characteristic is that it needs to make trend predictions based on sufficient historical data, fully considering the statistical characteristics and random characteristics. Among the various data prediction models, the RNN neural network can process time-series data. Its variant LSTM further overcomes the problem of gradient disappearance or explosion, which is prone to occur when the RNN processes long sequences. Experiments show that its performance is overall better than the traditional RNN [17].

In summary, this research constructed an algorithm for sensor fault diagnosis and recovery based on the PCA, LSTM neural network, and decision tree and then transplanted the algorithm to the developed platinum resistance dry-type transformer temperature monitor. Using the designed temperature monitor to collect the temperature data, a PCA-based fault diagnosis model and a LSTM-based data recovery model were constructed. A decision tree-based fault location model was constructed based on five typical sensor fault types. Finally, the three models were constructed to obtain the sensor fault diagnosis and recovery algorithm, and the algorithm was transplanted into the developed platinum resistance dry-type transformer temperature monitor. The results of the laboratory simulation and field experiments showed that the algorithm had good effects on the multi-sensor fault diagnosis and data recovery of dry-type transformers. Therefore, it provides a reliable method for ensuring the normal operation of the dry-type transformer temperature monitor.

## 2. Materials and Methods

### 2.1. Monitor Principle and Design

In order to obtain the real-time temperature data inside the dry-type transformer, in this study, a platinum resistance dry-type transformer temperature monitor was developed. A field experiment was carried out at a transformer manufacturer in Yunnan Province, China. The monitor could acquire the dry-type transformer temperature data in real-time and make timely and correct responses to abnormal temperatures, ensuring that the transformer worked within the rated temperature range. The dry-type transformer monitor had a total of six modules, and its physical diagram and architecture are shown in Figure 1.

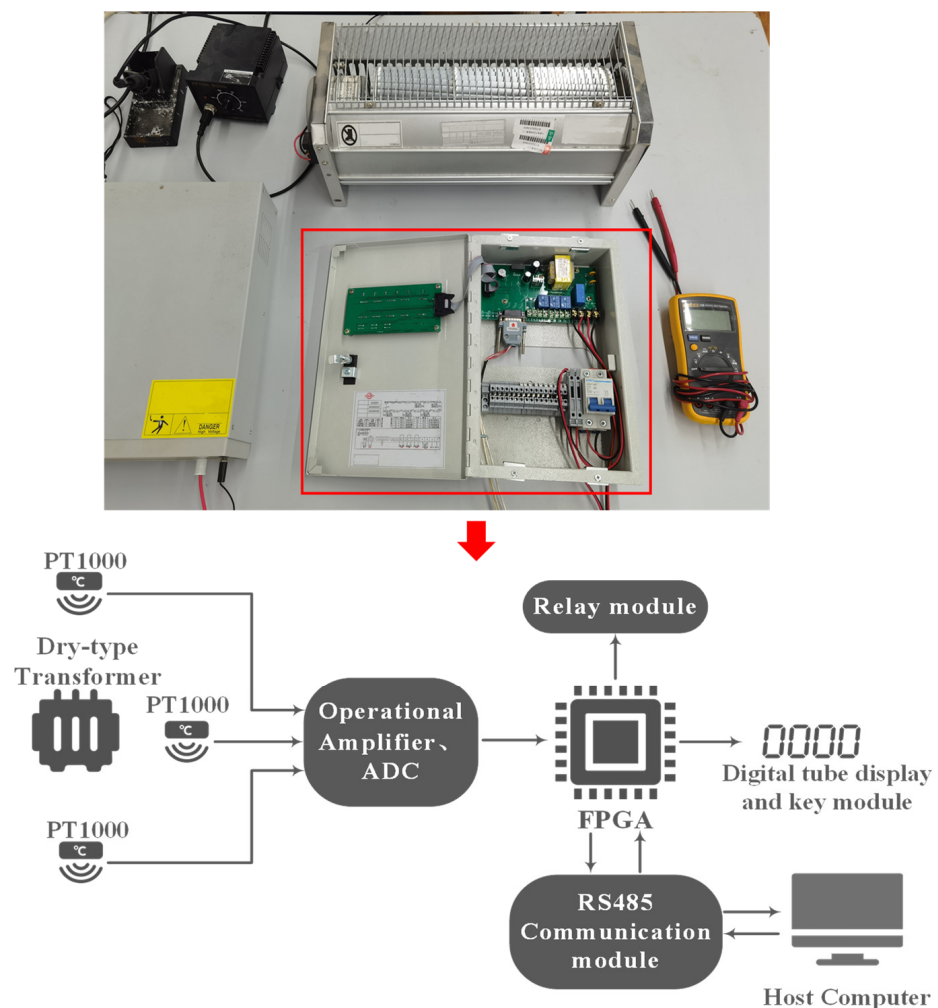


Figure 1. The dry-type transformer temperature monitor and its architecture.

The dry-type transformer used in the experiment was cooled by forced air cooling. That is, the fan was arranged under the high-voltage winding. This method could keep the temperature of the transformer stable, especially under overload conditions, making it more reliable. However, it also led to the obvious step-by-step distribution of the internal temperature of the transformer, and the temperature distribution of the lower part showed an asymmetrical phenomenon. Therefore, the monitor needed to use multiple sensors for multi-point sampling and real-time monitoring. At the same time, according to the industry standard of China JB/T 7631-2005 [18], the accuracy of the thermostat must meet 0.1 °C, the range is 0–270 °C, it has a black-box function, and it could communicate with the host computer.

Based on the above job requirements, the temperature controller designed in this study used three-way PT1000 platinum resistance as the temperature sensor. The resistance of the thermal resistance was 1000 Ω at 0 °C, and its resistance changed linearly with temperature. The four-wire structure was used to eliminate the lead resistance. These were arranged at different positions at the lower end of the high-voltage winding, and the signals were amplified and transmitted to the single-chip microcomputer through three-way operational amplifiers. Its resistance value was converted by the operational amplifier and ADC, then finally converted into specific temperature data by FPGA operation. At the same time, the monitor used five digital tubes to scan and display the monitoring temperature of each channel and could communicate with the host computer through RS485.

Because the national standard of China GB/T 1094.11-2007 [19] stipulates that the maximum temperature rise of the transformer shall not exceed 150 °C. To ensure that the transformer worked in the normal temperature range, the FPGA needed to select the maximum value of the three-temperature data as the reference value to determine whether the reference value exceeded the default monitoring threshold. The relay module realized the functions of turning off the fan below 80 °C, starting to cool down the fan when the temperature exceeded 100 °C, a high-temperature alarm by buzzer when the temperature exceeded 130 °C, a high-temperature trip when the temperature exceeded 150 °C, fault tripping, and fault alarm when the temperature exceeded the range from –30 °C to 240 °C.

## 2.2. Implementation of Fault Diagnosis

PCA is a multivariate statistical process control method. It can effectively reduce the dimension of high-dimensional data. Dimensionality reduction is very important to find the inherent laws of high-dimensional data so that more variable indicators can be represented by fewer comprehensive indicators. In geometry, the coordinate system formed by the samples is projected into a new coordinate space through linear combination, and the new coordinate axis represents the direction with the largest variance [20]. The fault diagnosis of the algorithm is realized based on PCA. The three-way sensor of the monitor can be considered as three characteristics.

The basic theory is that  $S$  is assumed to represent a measurement sample containing  $m$  sensors. Each sensor has  $n$  independent sampling data to construct a measurement data matrix, where each column represents a measured variable, and each row represents a sample. Perform covariance decomposition on the data matrix and choose the number of pivots. The following formula is obtained:

$$S \approx \frac{X^T \cdot X}{n-1} = V \cdot \Lambda \cdot V^T = [P \cdot \bar{P}] \cdot \Lambda \cdot [P \cdot \bar{P}]^T \quad (1)$$

where  $\Lambda$  is a diagonal matrix, which is also the eigenvalue matrix of  $S$ , and the elements on its diagonal satisfy  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ .  $V$  is the eigenvector matrix of  $S$  with the dimension  $m \times m$ .  $P$  is the first column  $A$  of  $V$ , containing the information about all of the pivots.  $\bar{P}$  is the remaining  $m-A$  columns of  $V$ , containing the non-pivot information.

Decompose the original data to obtain the principal subspace and the residual subspace. Therefore, the eigenvalue decomposition of  $X$  can be decomposed as follows:

$$X = X + E = T \cdot P + E \tag{2}$$

Among them,  $T_{n \times A} = X_{n \times m} \cdot P_{m \times A}$  is the main subspace;  $E = X - \hat{X}$  is the residual subspace;  $\hat{X} = T \cdot P^T$  is the score matrix;  $P_{m \times A}$  is the load matrix, which is composed of the first  $A$  eigenvectors of  $S$ .

$Q$  statistic: Prediction squared error. That is, the SPE statistic, which indicates the square of the Euclidean distance of the residual space projection vector  $e$  on this space. The calculation method is:

$$Q = \|e\|^2 \tag{3}$$

When the sensor is normal, the value of the  $Q$  statistic should be in a fixed range. Once the sensor fails, the projection of the temperature data at this time in the residual space must be enlarged, causing the calculated value of the Euclidean distance to be higher than the limited range. Its fixed range threshold is  $Q_a$ , which can be calculated from the last  $n - 1$  eigenvalues:

$$Q_a = \theta_1 \left[ \frac{c_a \sqrt{2\theta_2 h_0^2}}{\theta_2} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \tag{4}$$

where

$$h_0 = 1 - \frac{2\theta_1 \theta_3}{3\theta_2^2} \tag{5}$$

$$\theta_i = \sum_{j=k+1}^n \lambda_j^i, i = 1, 2, 3 \tag{6}$$

where  $k$  is the main element and  $\lambda_i$  is the  $i$ th eigenvalue of the covariance matrix  $R$ .

By setting the confidence level to 90%, analyze whether the projection change of the data in the residual space exceeds the threshold. That is, whether  $Q$  is higher than  $Q_a$  in diagnosing whether the sensor is faulty.

### 2.3. Realization of Sensor Fault Location

According to the national standard of China JB/T 7631-2005 [18], the displayed temperature difference between the sensors shall not exceed 0.5 °C. Exceeding this temperature range can be regarded as the occurrence of failure. To locate the faulty sensor. The maximum value of the selected three-channel sensor data after diagnosis is defined as the reference value  $X$ , and the remaining two sensor data are defined as  $Y$  and  $Z$ , respectively. We combined the five typical fault types of impact fault, drift fault, brownout fault, constant value output, and deviation fault [21], and then summarized them into seven scenarios [22]. The fault location decision tree was constructed as shown in Figure 2.

Scenario 1: The temperature difference between the temperature values of the three sensors is less than 0.5 °C, indicating that there is no sensor fault. This scenario is to prevent misdiagnosis by the PCA troubleshooting model. The secondary diagnosis of the sensor data is performed here to improve the accuracy of the data diagnosis.

Scenario 2: The maximum temperature difference between  $X$  and  $Y$  and between  $Y$  and  $Z$  is less than 0.5 °C, but that between  $X$  and  $Z$  is more than 0.5 °C. It is considered that the three-channel temperature value is  $X > Y > Z$ . Then, the LSTM predictor is started at this point. If the temperature difference between the predicted result  $P$  and  $X$  is less than 0.5 °C, sensor  $Z$  is faulty. If the temperature difference between the predicted results of  $P$  and  $Y$  is less than 0.5 °C, sensor  $X$  is faulty. Such failures are often deviation failures.

Scenario 3: The temperature difference between the maximum value  $X$  and  $Y$  is less than 0.5 °C, however, between  $X$  and  $Z$ ,  $Y$  and  $Z$  are all more than 0.5 °C. Then, start the LSTM predictor at this point. If the temperature difference between the predicted results  $P$

and Z is less than 0.5 °C, it is considered that the sensors X and Y are faulty. Otherwise, sensor Z is faulty. Such faults are often impulse faults or open circuit faults.

Scenario 4: The temperature difference between the maximum value X and Y is more than 0.5 °C, however, between X and Z, Y and Z are all less than 0.5 °C, the same as Scenario 2.

Scenario 5: The temperature difference between the maximum value X and Z is less than 0.5 °C, but between X and Y, Y and Z are all more than 0.5 °C, the same as Scenario 3.

Scenario 6: The temperature difference between the maximum X and Y, X and Z are all more than 0.5 °C, but between Y and Z it is less than 0.5 °C. Then, start the LSTM predictor. If the temperature difference between the predicted results P and X is less than 0.5 °C, the sensors Y and Z are faulty. Otherwise, sensor X is faulty.

Scenario 7: The difference between the temperature values of the three sensors is greater than 0.5 °C. All three sensors are faulty.

Therefore, to meet the positioning requirements, the prediction accuracy of the predictor needs to be high enough, and the difference between the predicted value P and the actual value should not exceed 0.5 °C to ensure the accuracy of the fault location. Otherwise, there will be no way to isolate the faulty data.

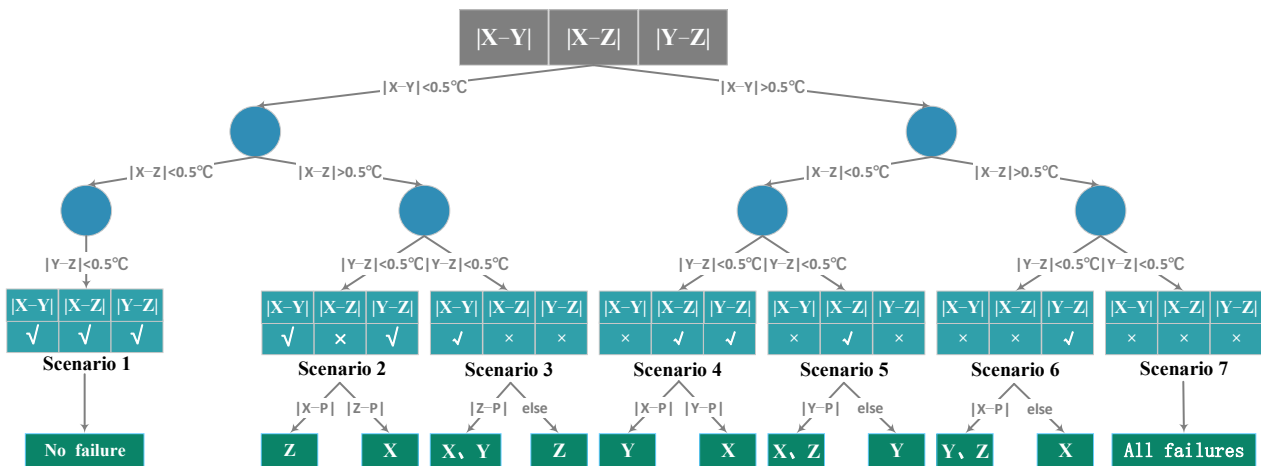


Figure 2. The sensor fault localization decision tree.

2.4. Implementation of Data Recovery

In this paper, the LSTM neural network was used as the data prediction model, and its structure is shown in Figure 3.

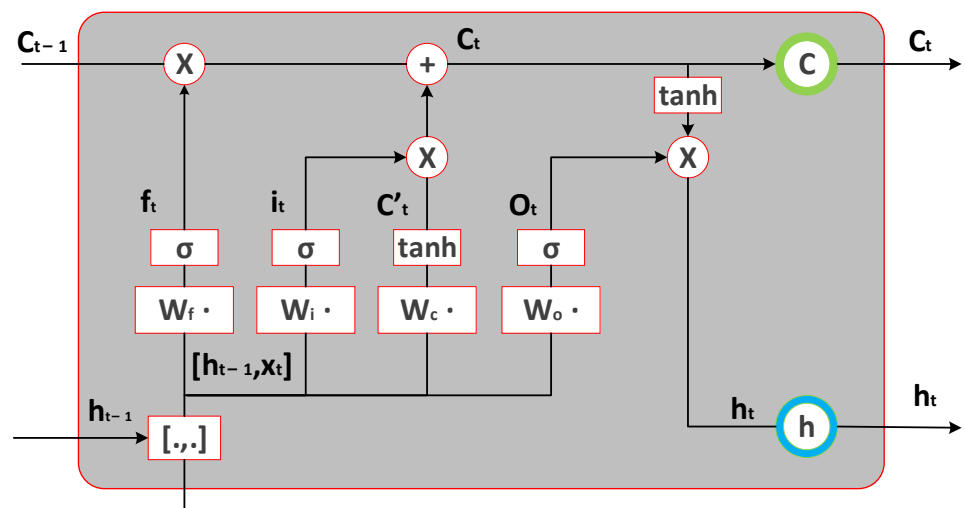


Figure 3. The structure of the LSTM.

LSTMs use gates to control the input and output of data. A gate is a fully connected layer whose input is a vector and its output is a real vector between 0 and 1. Assuming  $W$  is the weight vector of the gate and  $b$  is the bias term, then the gate can be expressed as:

$$g(x) = \sigma(Wx + b) \quad (7)$$

The use of the gate is to multiply the output vector of the gate by the vector, so we need to control it element by element because the output of the gate is a real vector between 0 and 1. When the gate output is 0, any vector multiplied by it will result in a 0 vector, which is equivalent to nothing passing through. When the output is 1, any vector multiplied by it will not change anything, which is equivalent to passing anything. Because the range of  $\sigma$  (sigmoid function) is (0,1), so the state of the door is half-open and half-closed. The LSTM uses two gates to control the content of the cell state  $c$ . One is the forget gate, which determines how much of the cell state  $c_{t-1}$  at the previous moment is retained in the current moment  $c_t$ . The other is the input gate, which determines how much of the network's input  $x_t$  is saved to the cell state  $c_t$  at the current moment. The LSTM uses an output gate to control how much of the cell state is output to the current output value  $h_t$  of the LSTM. The detailed formula is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (8)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (9)$$

where  $f_t$  is the forget gate;  $i_t$  is the input gate;  $W_f$  is the weight matrix of the forget gate;  $W_i$  is the weight matrix of the input gate;  $h_{t-1}$  is the hidden state at time  $t - 1$ ;  $x_t$  is the input vector at time  $t$ ;  $b_f$  is the bias term of the forget gate;  $b_i$  is the bias term of the input gate.

Multiply the last unit state  $c_{t-1}$  by the forget gate  $f_t$  element by element, then multiply the current input unit state  $\tilde{c}_t$  by the input gate  $i_t$  element by element, and then add the two products to obtain the current unit state  $\tilde{c}_t$ . The unit state used to describe the current input is calculated from the previous output and the current input. The detailed formula is as follows:

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (10)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \quad (11)$$

where  $\tanh$  is the activation function;  $W_c$  is the weight matrix of the current input cell state;  $b_c$  is the bias term of the memory update.

The output gate controls the effect of long-term memory on the current output. The final output of the LSTM is determined by the output gate and the unit state:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (12)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \tilde{c}_t \quad (13)$$

where  $o_t$  is the output gate;  $W_o$  is the weight matrix of the output gate;  $h_t$  is the current hidden state, which is the final output of this unit;  $b_o$  is the bias term of the output gate.

At the same time, to make the model state the best, it is necessary to optimize the hyperparameters of the LSTM network to make the training model meet the requirements. K-fold cross-validation (CV) is used to calculate the error by dividing the sample into K parts, using K-1 parts as the training set each time, and the remaining part as the validation set. This process is repeated K times. Finally, the average of the K errors is used as the CV estimate of the outer sample error [23]. Grid search optimization is an exhaustive method that enumerates or separates the hyperparameters that need to be optimized. The optimal solution is obtained by traversing all combinations of the hyperparameters and comparing the evaluation indicators during calculation. When the number of hyperparameters grows, the computational complexity of the grid search increases exponentially. If the sample size is large, it is often impractical to use a grid search

for too many hyperparameters at once. At this time, the fast-tuning method of coordinate descent can be used. According to the influence weight of the hyperparameters on the model, the hyperparameters with the greatest influence are optimized first, and then the other parameters are optimized in turn. Combining the grid search method and the K-fold cross-validation to optimize the hyperparameters can avoid the optimization process from falling into the local optimal solution to a certain extent.

2.5. The Construction of Fault Diagnosis and Recovery Algorithm

We combined the characteristics and specific implementation process of the four functional models of temperature data acquisition, fault diagnosis, fault location, and data recovery. In this study, each function was connected in series according to the data flow to form an overall algorithm for fault diagnosis and recovery, the structure of which is shown in Figure 4.

The algorithm first sorts the temperature data obtained by the three-way sensors, selects the maximum value in the three-way, and defines it as the temperature X. The remaining temperature values are defined as temperature Y and temperature Z. The temperature data at this time cannot be judged whether it is correct or not. It can only be divided into two situations: the steady state signal and sudden change signal. The PCA model diagnoses the three-way data, and distinguishes three situations: the steady state signal, the normal sudden change signal, and the fault signal. The steady-state signal and the normal mutation signal are the normal conditions, and the fault signal is the abnormal condition. For normal conditions, the temperature signal is output directly. For abnormal situations, the three-way data are calculated by the fault location algorithm and the abnormal situations are classified according to seven abnormal scenarios. Then, the LSTM predictor is started, bringing the predicted value P into this abnormal situation, and finally determining the faulty sensor and diagnosing the fault type. If it does not exceed the threshold, it means that the data are still normal data at this time, and the temperature signal is output directly. If the threshold is exceeded, one or more sensors are faulty. This part of the data is isolated and the temperature signal predicted by the predictor is output.

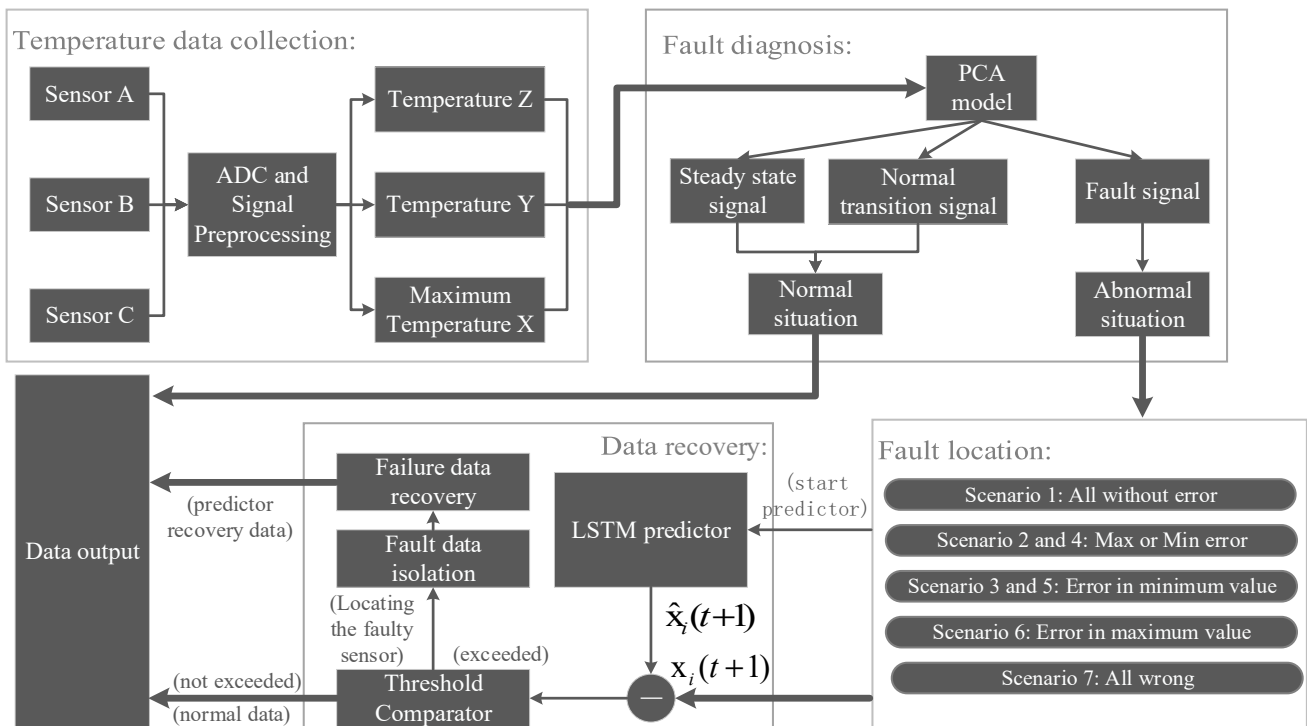


Figure 4. The sensor fault diagnosis and restoration algorithms.



### 3. Results and Discussion

#### 3.1. Sensor Failure Simulation Experiment

Since the internal temperature of dry-type transformers is mainly composed of the coil temperature rise, it is also influenced by the local air temperature [24]. The air temperature is characterized by seasonal changes, where the coil temperature will rise and fall with the increase or decrease in load and show daily periodicity. To better verify the algorithm's performance of each function, the verification set should contain the above two features. Therefore, the experiment selected the measured temperature of the whole-day on 10 April in spring and 10 December in winter as examples for verification. Since the sampling frequency of the monitor was once a second, there were 86,400 monitoring data of three-way sensors per day. The sensor fault diagnosis function of the algorithm needed to be verified with the fault data. The fault temperature data could be obtained by the method of laboratory fault simulation because the measured data came directly from the normal temperature monitored by the sensor. By superimposing the fault states based on normal data, it simulated the shock faults, drift faults, power failure faults, constant output, and deviation faults [25,26]. The specific method was as follows:

Impulse faults were caused by random disturbances, surges, and spark discharges in the power and ground wires. This caused a sharp jump in the data measured by the sensor in a short period. A linear function could be superimposed on this to make the temperature rapidly increase and fall back in a short time to simulate a shock failure. Drift faults were caused by the gradual shift in the sensor measurement values. Generally speaking, drift faults are linear drift faults. These can be fitted by a superimposed linear relationship, and a drift step size set to 0.01 °C. That is, if the drift is 0.01 °C per second, the drift is 0.6 °C per minute. The power failures were caused by broken signal lines, disconnected chip pins, or poor circuit contact. This kind of fault would cause the measured output value to be 0 °C. Therefore, the temperature value of this period can be directly set to 0 °C for the simulation. The constant value output belongs to the hard fault in the sensor type fault. When the fault occurs, the sensor loses its measurement ability and keeps a constant value. It is possible to represent a constant temperature value by setting a constant. The bias faults were caused by the bias current or bias voltage. This would make the data measured by the sensor deviate from the normal measurement value, which could be simulated by superimposing the deviation value.

The above five fault simulation methods were added to the time series monitored by the three-way sensors A, B, and C. The whole-day temperature measured on 10 December 10 in winter was used as the test set. Shock faults were added to sensors A and B between 1000 and 2000 s. Open circuit faults were added to sensor B between 8000 and 13,000 s. Then, we made sensor C power down between 20,000 and 26,000. Next, sensor C had a drift fault in the interval of 31,000 to 35,000 s, drifting at 0.01 °C per second. For the deviation fault, three faults were designed: high, low, and multi-sensor deviation. From 40,000 to 44,000 s, sensor A was 10 °C higher, and from 50,000 to 54,000 s, sensor B was 10 °C lower, and between 64,000 and 74,000 s, sensor B and C had low faults, where B was 10 °C lower and C was 20 °C lower. The all-day fault temperature simulated by the above method is shown in Figure 5.

#### 3.2. Accuracy Simulation Experiment of Fault Diagnosis

The fault data obtained by the simulation was used as the test set. The simulated sensor faults were tested by the PCA fault diagnosis model. The test results are shown in Figure 6 below.

The simulation results showed that when the temperature of the transformer was in a stable state, its SPE statistic (Q statistic) also remained in a stable state, and the projection on the residual space was smaller than the SPE threshold. This showed that the normal steady-state signal could be accurately identified by the PCA model. When the dry-type transformer temperature jumped normally, the sensor output value of the monitor also changed, and the projection in the residual space also jumped. However, the SPE value

of the temperature after the jump was still smaller than the set SPE threshold, so the PCA model could identify this situation as a normal mutation signal rather than a fault signal. When a fault occurred, the SPE value of the fault was higher than the set threshold. Moreover, the jump amplitude of multi-sensor faults in the residual space was larger than that of the single-sensor faults. According to the number of transitions higher than the SPE threshold, the number of faults could be counted.

The results of the simulation experiments showed that the PCA model could accurately diagnose different types of sensor faults as well as multiple sensor faults.

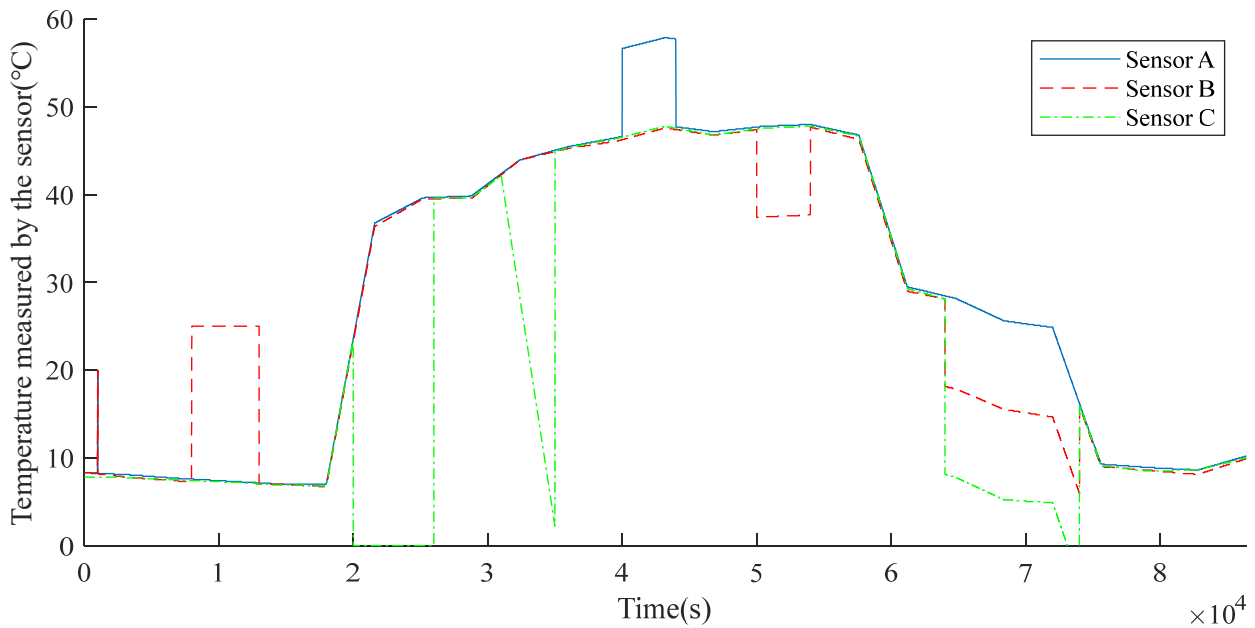


Figure 5. Simulation of the all-day fault temperature.

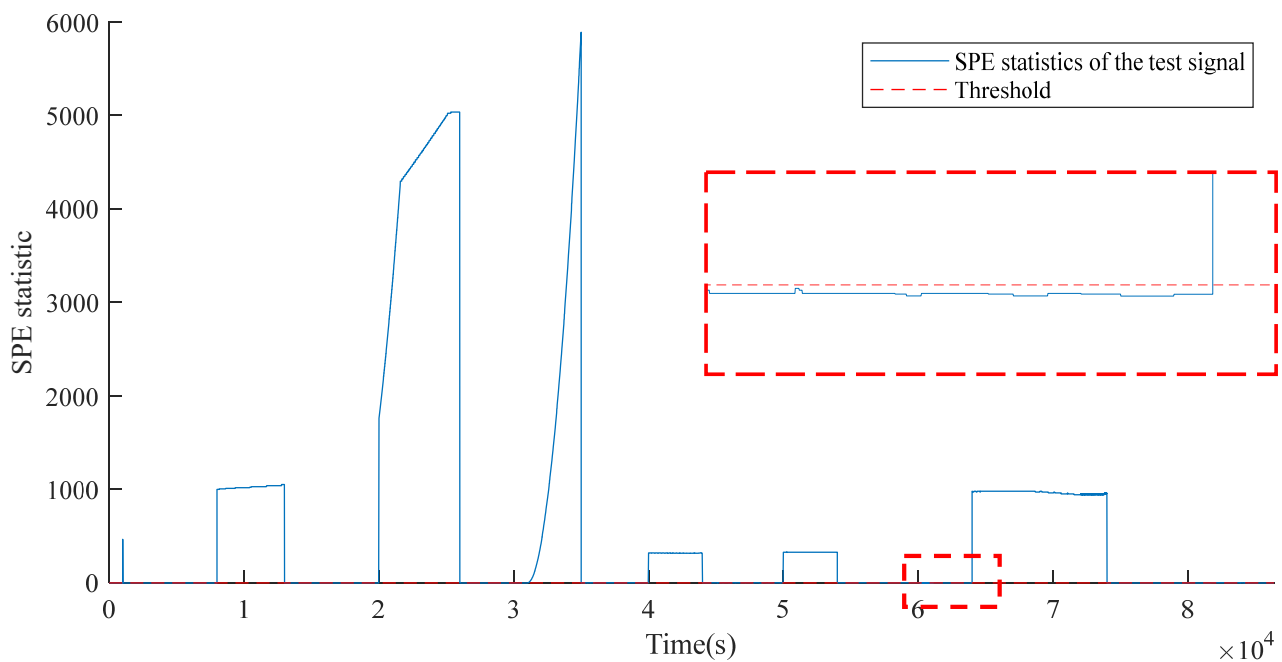


Figure 6. The SPE statistic troubleshooting results.

To verify the diagnostic accuracy of the PCA model through the above fault simulation method, five typical faults were added to the whole-day data at random times. Each fault was tested 50 times, and the experimental results are shown in Table 1.

**Table 1.** The diagnostic accuracy of the different faults.

	Impact Fault	Drift Fault	Power Failure	Constant Output	Deviation Fault
Diagnosis rate/%	100	100	100	98	96

### 3.3. Data Recovery Model Training and Optimization

The hyperparameters of the LSTM network were optimized by using the K-fold cross-validation and grid search optimization to make the trained model meet the requirements. The LSTM network training used 90% of the data as the training set and 10% as the validation set, then normalized the data. The hyperparameters of the LSTM network were optimized by a grid search combined with 5-fold cross-validation. The residual sum of squares (RSS) was used as the loss. Finally, the optimizer type was Adam, the activation function was sigmoid, the number of hidden neurons was 12, and the maximum number of cycles of the neural network was 1000. For the dataset in this experiment, a time step of 24 was found to be advantageous over longer time steps during training. This meant that the data of the past 24 moments could be used to determine the output of the current moment. The data were divided into a training set and validation set according to 4:1. The initial learning rate of the Adam optimizer was set to 0.01, and the loss on the validation set was monitored to use the learning rate decayed callback function. If the loss did not decrease after five consecutive iterations, the learning rate was reduced to be half and the lower limit of the learning rate reduction was set to 0.001. When the loss of the validation set did not decrease after five consecutive iterations, the callback function of the early stop was triggered to avoid overfitting.

### 3.4. Accuracy and Generalization Performance Analysis of Data Recovery Model

Two sets of experimental data were selected for the experiment: 8:00 on 10 April in spring and 18:00 on 10 December in winter. Then, we analyzed the forecast accuracy for the future period at two-time points. Since these two sets of test sets did not appear in the process of model training and tuning, they could reflect the generalization performance of the model and could characterize the periodic change process of temperature. At the same time, the BP and SVM predictors were compared with the LSTM predictors to analyze the accuracy of the LSTM predictors.

The experimental results are shown in Figures 7–10. The experimental time in Figure 7 was 8:00 in the morning. After this time, the load gradually increased, and the temperature of the transformer also increased in increments. The LSTM model always followed the temperature of the transformer. From Figure 8, it could be found that the relative error of the LSTM model was the smallest compared to the BP and SVM models and met the accuracy requirements of an error of less than 0.5 °C. The experimental time in Figure 9 was 18:00. After this time, the load gradually decreased, and the temperature of the transformer decreased accordingly. The LSTM could still follow the temperature change well. It can be seen in Figure 10 that the output value of BP had a certain deviation from the real temperature, and the error of the SVM model was too large, which seriously lost the recovery ability. LSTM met the accuracy requirement with an error of less than 0.5 °C.

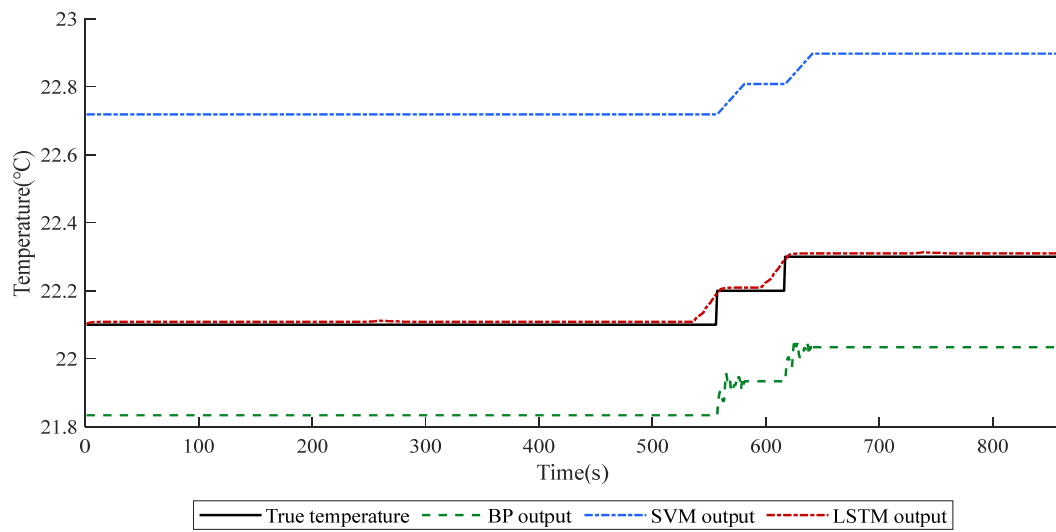


Figure 7. A comparison of the forecast values for the spring data.

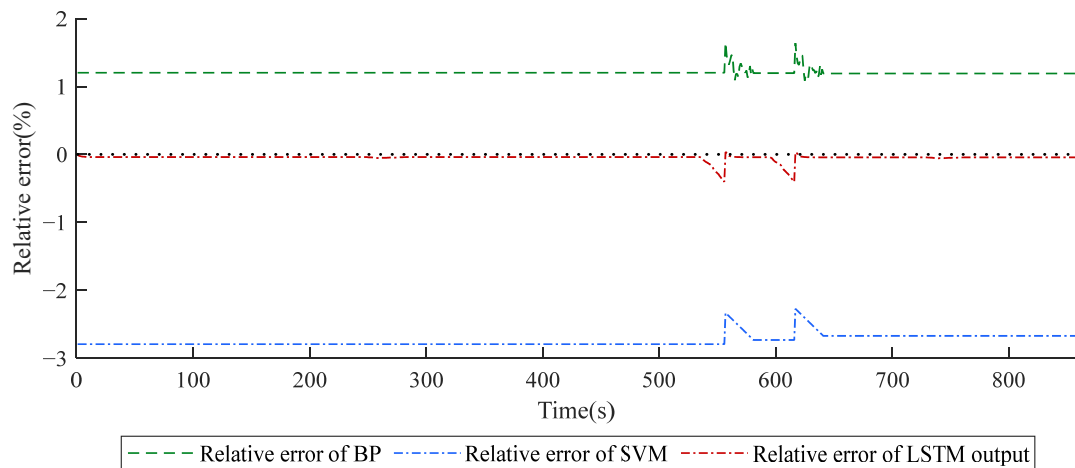


Figure 8. A comparison of the relative errors of the spring data forecast values.

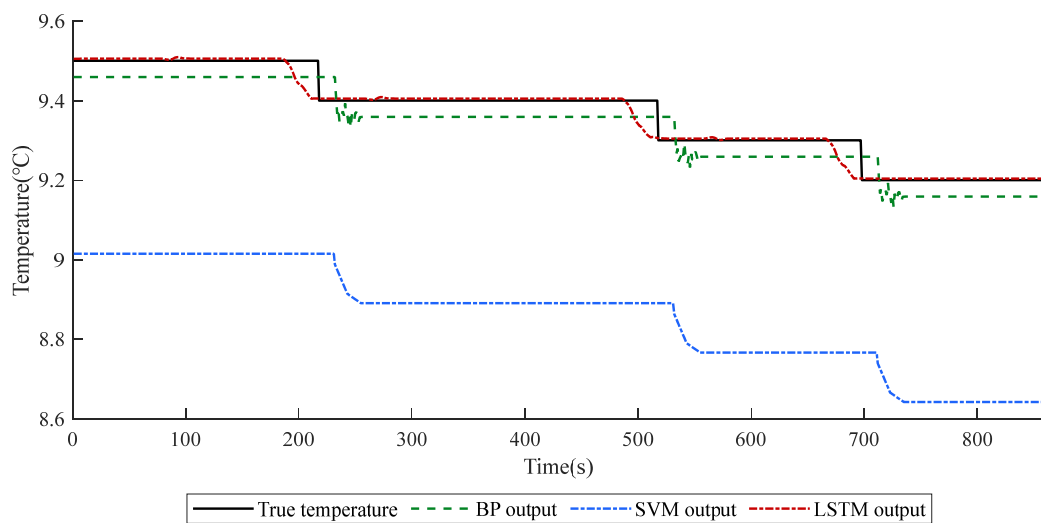
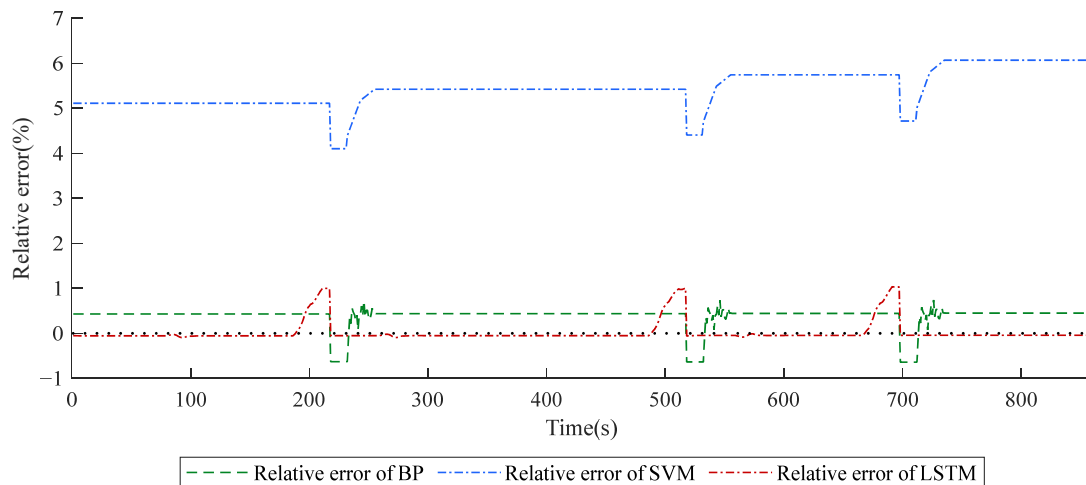


Figure 9. A comparison of the forecast values for the winter data.



**Figure 10.** A comparison of the relative errors of the winter data forecast values.

Table 2 shows the mean absolute error (MAE), root mean square error (RMSE), and maximum relative error (MRE) of the prediction results of each model. It can be seen that the prediction effect of the LSTM model was significantly better than that of the other models.

**Table 2.** The error comparison of the different prediction algorithms.

Error	Algorithm	10 April 8:00	10 December 18:00
RMSE/°C	BP	0.2671	0.0421
	SVM	0.6090	0.5100
	LSTM	0.0146	0.0221
MAE/°C	BP	0.2670	0.0418
	SVM	0.6087	0.5088
	LSTM	0.0109	0.0103
MRE/%	BP	±1.6486	±0.7402
	SVM	±2.7975	±6.0617
	LSTM	±0.4053	±1.0337

### 3.5. Field Test Results

The algorithm was transplanted into the designed temperature monitor, then the fault diagnosis and recovery ability of the algorithm were tested through field experiments. The experimental time was 86,400 s throughout the day. The host computer sent a simulated sensor failure at a random time point. In this way, the algorithm's ability to diagnose and recover data from single-sensor failure and multi-sensor failure were tested. At the same time, it verified that the thermostat would not respond incorrectly due to sensor failure, and recorded the real-time data.

The field experiment results are shown in Figures 11–13. The algorithm's fault diagnosis function could accurately diagnose different types of single sensor faults or multi-sensor faults. The diagnosis time was less than 1 ms. The fault location function could accurately locate the faulty sensor, isolated the faulty sensor data immediately, and started the predictor to restore the data of the faulty sensor. The data recovery function tracked the temperature changes well, and replaced the faulty sensor data with the output value of the predictor to achieve data recovery. The difference between the predicted value and the actual value was less than or equal to 0.1 °C, which met the accuracy requirements of the predicted value. The experiments showed that the algorithm was accurate in diagnosis, high in recovery accuracy, and accurate in fault location.

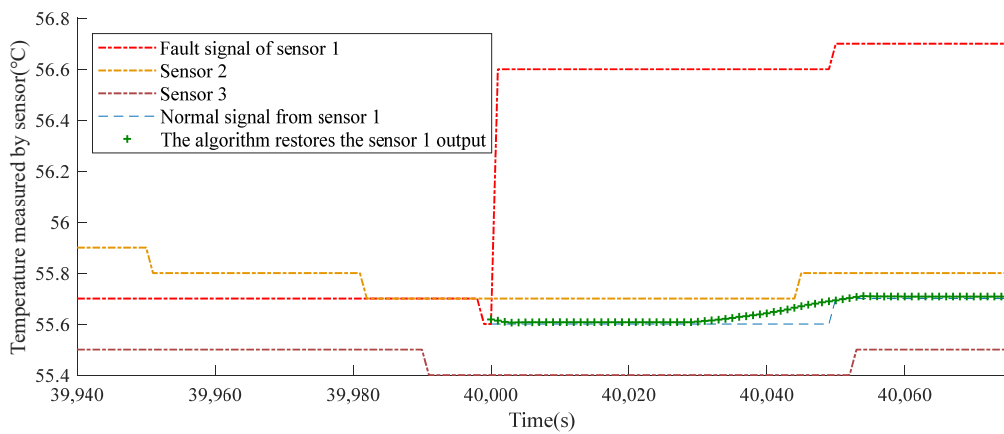


Figure 11. The data recovery for the bias failure of a single sensor.

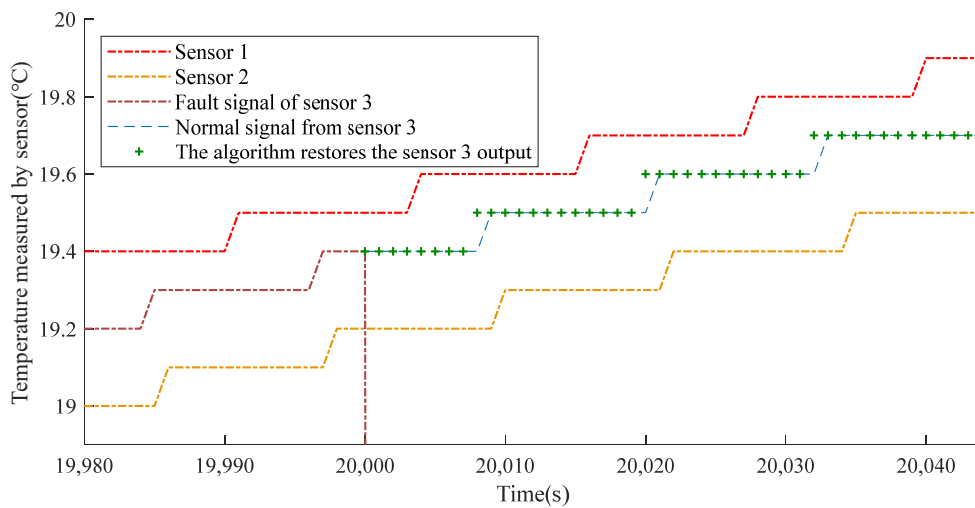


Figure 12. The data recovery for the power failure of a single sensor.

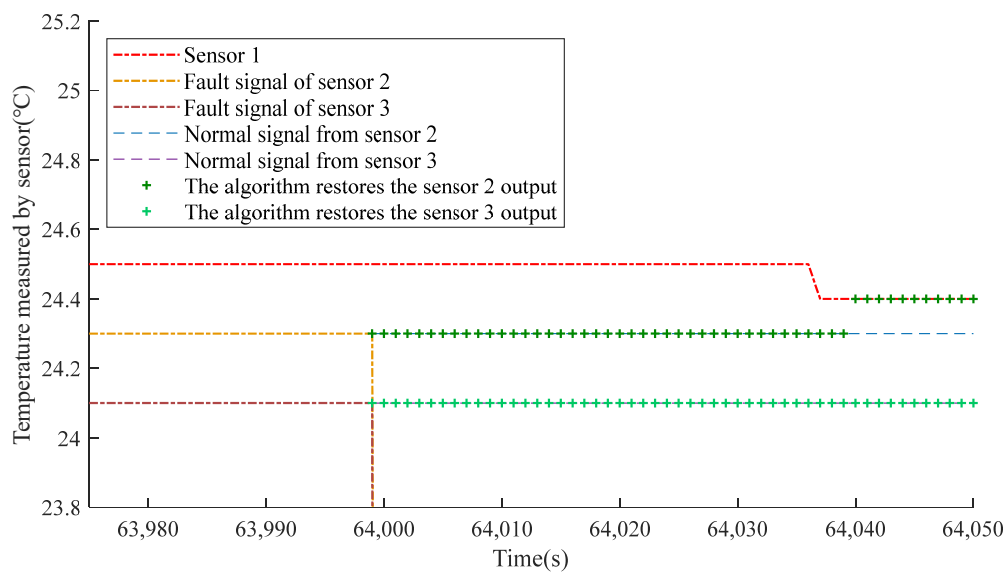


Figure 13. The data recovery for the bias failure of multiple sensors.

During the whole experiment, no false response of the monitor due to sensor failure occurred. This algorithm provided a reliable method for ensuring the normal operation of the dry-type transformer temperature monitor.

#### 4. Conclusions

For dry-type transformer temperature monitors causing erroneous responses due to sensor failure, in this study, a set of temperature monitors was designed. Then, a fault diagnosis and recovery algorithm based on the principal component analysis (PCA), long short-term memory neural network (LSTM), and decision tree was proposed. Finally, the feasibility and scientific experiments of the method were verified by the simulation experiments and field experiments. The research indicated the following:

1. The fault diagnosis function based on PCA could accurately diagnose the impact fault, open circuit fault, power failure fault, drift fault, and deviation fault of single or multiple sensors. The diagnosis rate was above 96%, and the diagnosis time was less than 1 ms.
2. Fault localization could diagnose the faulty sensor through a decision tree and isolated the fault data.
3. The LSTM-based data recovery function could accurately track the temperature changes under dynamic processes. The error of the predicted value was less than or equal to 0.1 °C and the generalization performance was good. Compared with the BP and SVM, it has obvious advantages.
4. The field experiments verified that the algorithm could significantly improve the stability of the monitor. Even if the sensor fails, the dry-type transformer was guaranteed to work within the normal temperature range.

**Author Contributions:** Conceptualization, K.Y. and Y.L.; Formal analysis, M.Z. and Y.L.; Software, M.Z.; Validation, M.Z.; Writing—original draft, M.Z.; Writing—review & editing, M.Z., K.Y., C.S. and Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (NSFC): Yi Luo 41761084, and the Ten Thousand Talent Plans for Young Top-notch Talents of Yunnan Province: Yi Luo YNWR-QNBJ-2019-200.

**Data Availability Statement:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Acknowledgments:** The authors would like to thank all of the authors for their hard work in this research.

**Conflicts of Interest:** The founding sponsors had no role in the design of the study, in the collection, analysis, or interpretation of data, in the writing of the manuscript, or in the decision to publish the results.

#### References

1. Yang, D.; Qin, J.; Pang, Y.; Huang, T. A novel double-stacked autoencoder for power transformers DGA signals with an imbalanced data structure. *IEEE Trans. Ind. Electron.* **2021**, *69*, 1977–1987. [[CrossRef](#)]
2. Sun, Y.; Xu, G.; Li, N.; Li, K.; Liang, Y.; Zhong, H.; Zhang, L.; Liu, P. Hotspot Temperature Prediction of Dry-Type Transformers Based on Particle Filter Optimization with Support Vector Regression. *Symmetry* **2021**, *13*, 1320. [[CrossRef](#)]
3. Liu, Y.; Gao, H.; Gao, W.; Peng, F. Development of a substation-area backup protective relay for smart substation. *IEEE Trans. Smart Grid* **2016**, *8*, 2544–2553. [[CrossRef](#)]
4. Yin, Z.; Hou, J. Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes. *Neurocomputing* **2016**, *174*, 643–650. [[CrossRef](#)]
5. Simbeye, D.S.; Yang, S.F. Water quality monitoring and control for aquaculture based on wireless sensor networks. *J. Netw.* **2014**, *9*, 840. [[CrossRef](#)]
6. Bayar, N.; Darmoul, S.; Hajri-Gabouj, S.; Pierreval, H. Fault detection, diagnosis and recovery using Artificial Immune Systems: A review. *Eng. Appl. Artif. Intell.* **2015**, *46*, 43–57. [[CrossRef](#)]
7. Willsky, A.; Jones, H. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Trans. Autom. Control.* **1976**, *21*, 108–112. [[CrossRef](#)]

8. Yuqing, L.; Tianshe, Y.; Jian, L.; Na, F.; Guan, W. A fault diagnosis method by multi sensor fusion for spacecraft control system sensors. In Proceedings of the 2016 IEEE International Conference on Mechatronics and Automation, Harbin, China, 7–10 August 2016; pp. 748–753.
9. Chen, H.; Jiang, B.; Lu, N.; Mao, Z. Deep PCA based real-time incipient fault detection and diagnosis methodology for electrical drive in high-speed trains. *IEEE Trans. Veh. Technol.* **2018**, *67*, 4819–4830. [[CrossRef](#)]
10. Li, W.; Peng, M.; Liu, Y.; Jiang, N.; Wang, H.; Duan, Z. Fault detection, identification and reconstruction of sensors in nuclear power plant with optimized PCA method. *Ann. Nucl. Energy* **2018**, *113*, 105–117. [[CrossRef](#)]
11. Berriri, H.; Slama-Belkhdja, I. Enhanced parity equations method for sensor fault detection in electrical drives. In Proceedings of the 2010 Conference on Control and Fault-Tolerant Systems (SysTol), Nice, France, 6–10 October 2010; pp. 831–836.
12. Tolouei, H.; Shoorehdeli, M.A. Nonlinear parity approach to sensor fault detection in pH neutralization system. In Proceedings of the 2017 Iranian Conference on Electrical Engineering (ICEE), Tehran, Iran, 2–4 May 2017; pp. 889–894.
13. Wang, B.; Diao, M.; Zhang, H. Fault diagnosis and data recovery of sensor based on relevance vector machine. In Proceedings of the 2014 IEEE International Conference on Mechatronics and Automation, Tianjin, China, 3–6 August 2014; pp. 1822–1826.
14. Zhu, T.B.; Lu, F. A Data-Driven Method of Engine Sensor on Line Fault Diagnosis and Recovery. *Trans. Tech Publ* **2014**, *490*, 1657–1660. [[CrossRef](#)]
15. Oh, B.K.; Glisic, B.; Kim, Y.; Park, H.S. Convolutional neural network-based data recovery method for structural health monitoring. *Struct. Health Monit.* **2020**, *19*, 1821–1838. [[CrossRef](#)]
16. Orosz, T.; Rassólkin, A.; Kallaste, A.; Arsénio, P.; Pánek, D.; Kaska, J.; Karban, P. Robust Design Optimization and Emerging Technologies for Electrical Machines: Challenges and Open Problems. *Appl. Sci.* **2020**, *10*, 6653. [[CrossRef](#)]
17. Zhang, S.; Wang, Y.; Liu, M.; Bao, Z. Data-based line trip fault prediction in power systems using LSTM networks and SVM. *IEEE Access* **2017**, *6*, 7675–7686. [[CrossRef](#)]
18. Electronic Thermostats for Transformers 2005, JB/T 7631-2005. Available online: <https://kns.cnki.net/kcms/detail/detail.aspx?FileName=SCSD000001038584&DbName=SCSD> (accessed on 10 May 2022).
19. Power Transformers-Part 11: Dry-Type Transformers. China Electrical Equipment Industry Association, 2007. GB/T 1094.11-2007. Available online: <https://kns.cnki.net/kcms/detail/detail.aspx?FileName=SCSD000005138709&DbName=SCSD> (accessed on 10 May 2022).
20. Rogers, A.P.; Guo, F.; Rasmussen, B.P. A review of fault detection and diagnosis methods for residential air conditioning systems. *Build. Environ.* **2019**, *161*, 106236. [[CrossRef](#)]
21. Ni, K.; Ramanathan, N.; Chehade, M.N.H.; Balzano, L.; Nair, S.; Zahedi, S.; Kohler, E.; Pottie, G.; Hansen, M.; Srivastava, M. Sensor network data fault types. *ACM Trans. Sens. Netw. (TOSN)* **2009**, *5*, 1–29. [[CrossRef](#)]
22. Dragos, K.; Smarsly, K. Distributed adaptive diagnosis of sensor faults using structural response data. *Smart Mater. Struct.* **2016**, *25*, 105019. [[CrossRef](#)]
23. Borra, S.; Di Ciaccio, A. Measuring the prediction error. A comparison of cross-validation, bootstrap and covariance penalty methods. *Comput. Stat. Data Anal.* **2010**, *54*, 2976–2989. [[CrossRef](#)]
24. Dai, X.; Qin, F.; Gao, Z.; Pan, K.; Busawon, K. Model-based on-line sensor fault detection in Wireless Sensor Actuator Networks. In Proceedings of the 2015 IEEE 13th International Conference on Industrial Informatics (INDIN), Cambridge, UK, 22–24 July 2015; pp. 556–561.
25. Balzano, L.; Nowak, R. Blind calibration of sensor networks. In Proceedings of the 6th International Conference on Information Processing in Sensor Networks, Cambridge, MA, USA, 25–27 April 2007; pp. 79–88.
26. He, X.; Wang, Z.; Liu, Y.; Qin, L.; Zhou, D. Fault-tolerant control for an Internet-based three-tank system: Accommodation to sensor bias faults. *IEEE Trans. Ind. Electron.* **2016**, *64*, 2266–2275. [[CrossRef](#)]