Substation equipment temperature prediction based on multivariate information fusion and deep learning network

Lijie Sun[1], Chunxue Liu[2], Ying Wang[3] and Zhaohong Bing[4]

[1] School of Electronics and Information Engineering, Taizhou University, Taizhou, Zhejiang, China
[2] School of Information, Liaoning University, Shenyang, Liaoning, China
[3] Economic and Technological Research Institute of State Grid Heilongjiang Electric Power Co., Ltd., Haerbin, Heilongjiang, China
[4] Computing Technology Institute of East China, Shanghai, China

Corresponding author
Lijie Sun, lijiesun@tzc.edu.cn

ABSTRACT

Background. Substation equipment temperature is difficult to achieve accurate prediction because of its typical seasonality, periodicity and instability, complex working environment and less available characteristic information.

Methods. To overcome these difficulties, a substation equipment temperature prediction method is proposed based on multivariate information fusion, convolutional neural network (CNN) and gated recurrent unite (GRU) in this article. Firstly, according to the correlation analysis including linear correlation mapping, autocorrelation function and partial autocorrelation function for substation equipment temperature data, the feature vectors from ambient, time and space are determined, that is the multivariate information fusion feature vector (denoted as MIFFV); secondly, the dimension of MIFFV is reduced by principal component analysis (PCA), extract some of the most important features and form the reduced feature vector (denoted as RFV); then, CNN is used for deep learning to extract the relationship between RFV and the high-dimensional space feature, and construct the high-dimensional feature vector of multivariate time series (denoted as HDFV); finally, the high-dimensional feature vector is used to train GRU deep learning network and predict the equipment temperature.

Results. A substation equipment in Taizhou City, Zhejiang Province is conducted by the method proposed in this article. Through the comparative experiment from the two aspects of features and methods, under the two prediction performance evaluation indexes of mean absolute percentage error (MAPE) and root mean square error (RSME), two main conclusions are drawn: (1) MIFFV from three aspects of ambient features, time features and space features have better prediction performance than the single feature vector and the combined feature vector of two aspects; (2) compared with other four related models under the same conditions, RFV is regarded as the input of the models, the proposed model has better prediction performance.

INTRODUCTION

The safe operation of power equipment is the focus and key to ensure the stable operation of substation, in which substation primary equipment is the top priority; therefore, we should attach great importance to the primary equipment of the substation, strengthen management and control, and do a good job in the daily condition monitoring and maintenance of the primary equipment of the substation (*Wang, 2016*). Equipment temperature is an important index to measure the health of equipment, however, online monitoring is mainly for primary equipment (*Sun, 2019*), and many factors will cause the equipment temperature to rise, such as too much voltage load, an insufficiently tightened joint connection, loose bolts at key points, oxidized and corrode conductor surface, too much contact resistance of the contact surface, and so on. If the temperature rises slightly, the relevant electrical equipment will be damaged and burned, which will lead to the operation failure of the substation; more importantly, it will lead to fire and safety accidents, resulting in huge economic losses and social impact of the substation. Therefore, it is very important to know the temperature of each equipment in real time.

In the past, the substation was inspected and measured regularly by manual means which is prone to casualties, and in recent years, the state grid has adopted the intelligent inspection means for the management and monitoring of substation equipment, and installed infrared cameras in the substation, but due to the limited storage space of the equipment, it is generally set for one day or one hour, so sometimes the fault can not be found in time. Through substation equipment temperature prediction, the future temperature information is obtained in advance, and the purpose of equipment fault early warning can be realized.

When the data source and data set have been identified, the completion of equipment temperature prediction task mainly needs to go through two processes: feature engineering and modeling. This article focuses on these two links to solve the problem of accurate prediction of substation temperature.

Feature engineering mainly carries out feature selection and feature extraction. For substation equipment temperature prediction, in addition to the complex working environment of substation equipment, the biggest difficulty is that the information source used for prediction is limited. The research results in this field at home and abroad show that there are more domestic research results and less foreign research results. The research results are mainly concentrated in domestic Huazhong University of Science and Technology, Harbin University of technology, Zhejiang University, North China Electric Power University and some power companies (*Hao et al., 2021*; *Guo et al., 2020*; *Kong, 2015*). The research objects of substation equipment at home and abroad mainly include high-voltage or low-voltage switchgear (*Velásquez, Lara & Melgar, 2019*; *Zeng et al., 2018*; *Bussière et al., 2017*), intelligent electronic equipment (*Sun et al., 2022*), disconnector (*Huang et al., 2022a*), bushing contact (*Huang et al., 2022b*), *etc.* At present, most studies used historical time series as feature extraction source for rolling prediction of equipment temperature, which typically include auto-regressive and moving average model (ARMA) series models (AR, ARMA, ARIMA) (*Baptista et al., 2018*); however, the

simple temperature trend can not accurately predict the future equipment temperature value, resulting in the failure to accurately identify the health status of the equipment and take precautions in advance. Some scholars are also constantly trying to find more feature sources. Through the seasonal analysis of substation equipment temperature data, it was found that there exists typical positive correlation between ambient temperature and equipment temperature. Therefore, the daily maximum temperature and daily minimum temperature are taken as ambient characteristics and equipment temperature at historical time to form a feature vector for equipment temperature prediction (*Yu et al., 2022*); in addition, by analyzing the influencing factors of temperature rise of high-voltage switchgear, *Xu, Xu & He (2016)* established a temperature prediction fusion model based on load current and ambient temperature of high-voltage switchgear by using information fusion technology and back propagation neural networBPNNk, and achieved good prediction performance. As is known, for primary equipment of substation main transformer, load current and equipment monitoring belong to different departments, so it is difficult to obtain load current information, and the daily maximum temperature and daily minimum temperature of the ambient can not clearly reflect the real-time correlation between the weather temperature and the equipment temperature, which will affect the prediction performance. Temperature is a parameter with heat transfer characteristics, and the temperature of adjacent positions in space has the effect of interaction. Based on current research, it can be seen that the traditional substation equipment temperature prediction method ignores the spatial relationship information of equipment in the historical time, resulting in poor prediction accuracy. Thus, it is particularly important to select what characteristics to characterize the temperature for prediction. So, when solving the problem of substation equipment temperature prediction, inspired by considering environmental perspective factor in the research results of the literature (*Hou et al., 2021a*), Feature extraction information comes from three viewpoints of ambient, time and space, and develops ambient feature vector, time feature vector and space feature vector as multivariate information fusion feature vector in this article. Considering that Zhejiang Province is a typical subtropical seasonal climate, the real-time weather temperature and humidity are selected as the ambient characteristics to form the ambient feature vector; the historical temperature time series of the monitoring points of the prediction target is selected as the time feature vector and the temperature of all monitoring points with space correlation for the predicted target monitoring point temperature is composed of space feature vector. Principal component analysis (PCA) (*Zhang et al., 2022*) is a common data analysis method and a linear dimensionality reduction method, whose principle is to map high-dimensional data to low-dimensional space through a certain linear projection, and expect the maximum amount of information (the largest variance) of the data on the projected dimension, so as to use fewer data dimensions and retain the characteristics of more original data points, which can be used to extract the main feature components of data. PCA has the functions of simplifying operation, removing data noise and discovering hidden related variables (*Dai, 2021*; *Song & Yang, 2022*), and it is adopted to reduce the feature vector of multivariate information fusion to form the reduced feature vector, so as to realize the feature extraction process for substation equipment temperature prediction.

The quality of the prediction model is also the main factor affecting the prediction performance. In the last five years, neural networks have been widely used in substation equipment temperature prediction, such as back propagation neural network (*Liu, 2012*), radial basis function neural network (*Wang et al., 2015*), generalized regression neural network (*Kong & Zhang, 2016*), adaptive neural network (*Wang, 2015*), neural network optimized by swarm intelligence algorithm (*Xu, Hao & Zheng, 2020*), support vector machine (SVM) and a series of other machine learning methods (*Zhang et al., 2020*). In the past three years, deep learning networks have made breakthrough, such as pedestrian trajectoryprediction (*Esfahani, Song & Christensen, 2020*), PM2.5 prediction (*Mohammadshirazi et al., 2022*), traffic speed prediction (*Zheng, Chai & Katos, 2022*), estimation of residual capacity for lithium-ion battery (*Hou et al., 2022*) and so on (*Xu, Lin & Zhu, 2020*). In 2021, *Hou et al. (2021b)* solved the problem of temperature prediction of switchgear equipment in substation by using long short-term memory (LSTM) network, and achieved good results, which opens the prelude of solving the problem of substation equipment temperature prediction with deep learning network. The gated recurrent unit (GRU) was proposed by *Gharehbaghi et al. (2022)* and is an effective variant of LSTM (*Cao, Jiang & Gao, 2021*; *Yuan et al., 2022*). In many cases, GRU and LSTM have the same excellent results, but GRU has fewer parameters, so it is relatively easy to train and the over fitting problem is lighter (*Cao, Jiang & Gao, 2021*; *Yuan et al., 2022*). Therefore, GRU network is adopt to predict substation equipment temperature in this article. Before the prediction, taking advantage of CNN's feature extraction (*Khalifani et al., 2022*), CNN network is used for deep learning to extract the relationship between the reduced feature vector and the equipment temperature in the high-dimensional space, and construct the high-dimensional feature vector of multivariate time series, then the high-dimensional feature vector is used to train GRU network and predict the equipment temperature.

RELATED WORK

Correlation analysis

Two functions of autocorrelation function and partial autocorrelation function are adopted to analyze correlation. The autocorrelation functon and partial autocorrelation function are described as follows. (1) As is known, autocorrelation belongs to sequence correlation, which expresses the cross-correlation between the sequence and itself at different moments (*Chachlakis et al., 2021*). The autocorrelation coefficient of the time series is denoted as ACF, that is autocorrelation function. This article quantitatively describes the lag autocorrelation of substation equipment temperature time series by calculating ACF value. ACF is expressed as $\hat{\rho}_k$ in formula Eq. (1):

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}\left(Z_t - \bar{Z}\right)\left(Z_{t+k} - \bar{Z}\right)}{\sum_{t=1}^{n}\left(Z_t - \bar{Z}\right)^2} \tag{1}$$

where, Z_t is the equipment temperature at time t, Z_{t+k} is the equipment temperature at time $t+k$, $\bar{Z}$ is the average value of equipment temperature. (2) Partial autocorrelation is the relationship summary between the time series observation after eliminating interference and the previous time step observation (*Mestre et al., 2021*). That is, consider the correlation after removing the influence of intervention variables $Z_{t+1}, Z_{t+2}, Z_{t+3}, \ldots$ with common linear dependence from Z_t and Z_{t+k}, namely, under the condition of observation Z_{t+1}, the autocorrelation state of Z_t and Z_{t+k} so on. Partial autocorrelation function (PACF) is expressed as P_k in formula Eq. (2):

$$P_k = \frac{Cov\left[\left(Z_t - \hat{Z}_t\right), \left(Z_{t+k} - \hat{Z}_{t+k}\right)\right]}{\sqrt{Var\left(Z_t - \hat{Z}_t\right)}\sqrt{Var\left(Z_{t+k} - \hat{Z}_{t+k}\right)}} \tag{2}$$

where, Cov refers to the covariance at moment t, Var refers to sample variance, $\hat{Z}_t$ is sample estimation at moment t, and $\hat{Z}_{t+k}$ is sample estimation at moment $t+k$.

PCA

Principal component analysis(PCA) is a data dimension reduction method that is widely applied in various fields (*Cao, Sun & Zhao, 2022*), which has the functions of simplifying operation, removing data noise and discovering hidden related variables. Therefore, PCA is selected to screen the input features. By calculating cumulative contribution rate of the input features, the first few important features are selected from multiple features as the principal components to reduce the input dimension and improve the convergence speed.

The main idea of PCA is to relinearly combine p-dimensional linearly related features and map them into k-dimensional linearly independent features ($k < p$). The reacquired k-dimensional features are principal components, which can represent the information of the original features to the greatest extent.

It is assumed that it has p features, and each feature has n observation values, then the initial data matrix C can be obtained.

$$C = \begin{bmatrix} c_{11} c_{12} \cdots c_{1p} \\ c_{21} c_{22} \cdots c_{2p} \\ \vdots \vdots \vdots \vdots \vdots \\ c_{n1} c_{n2} \cdots c_{np}. \end{bmatrix} \tag{3}$$

The implementation process of PAC method is realized by the following six steps:

(1) The original p characteristics are standardized to obtain the standardized feature variables.

$$y_j = \frac{s_j - \mu_j}{s_j}, j = 1, 2, \ldots, p \tag{4}$$

where, $\mu_j = \frac{1}{n}\sum_{i=1}^{n} c_{ij}, s_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(c_{ij} - \mu_j\right)^2}$.

(2) Standardize each feature element to obtain the corresponding data matrix W.

$$W = \begin{bmatrix} w_{11} w_{12} \cdots w_{1p} \\ w_{21} w_{22} \cdots w_{2p} \\ \vdots \vdots \vdots \\ w_{n1} w_{n2} \cdots w_{np} \end{bmatrix} \quad (5)$$

where, $w_{ij} = \frac{c_{ij} - \mu_{ij}}{s_j}, i = 1, 2, \ldots, n; j = 1, 2, \ldots, p.$

(3) According to the matrix W, the correlation coefficient matrix $R = (r_{ij})_{p \times p}$ of W is calculated. Where, $r_{ij} = \frac{\sum_{t=1}^{n} w_{t1} w_{tj}}{n-1}, i, j = 1, 2, \ldots, p.$

(4) Calculate the eigenvalues of matrix R andsort them in descending order $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$, and the standard orthogonalization eigenvector corresponding to each eigenvalue is calculated $u_1, u_2, \ldots, u_p$, where, $\mu_j = [\mu_{1j}, \mu_{2j}, \ldots, \mu_{pj}]^T, j = 1, 2, \ldots, p.$

(5) p new feature vectors are computed with the original p standard orthogonal feature elements, that is,

$$\begin{cases} N_1 = u_{11} y_1 + u_{21} y_2 + \cdots + u_{p1} y_p \\ N_2 = u_{12} y_1 + u_{22} y_2 + \cdots + u_{p2} y_p \\ \vdots \\ N_p = u_{1p} y_1 + u_{2p} y_2 + \cdots + u_{pp} y_p \end{cases} \quad (6)$$

where, N_1 refers to the first principal component; N_2 is the second principal component; N_p is the $p - th$ principal component.

(6) The contribution rate and cumulative contribution rate of each principal component are calculated, and the calculation formula is shown in formula Eq. (7) and formula Eq. (8) respectively.

$$N_j = \frac{\lambda_j}{\sum_{t=1}^{p} \lambda_t}, j = 1, 2, \ldots, p \quad (7)$$

$$\eta_i = \frac{\sum_{t=1}^{i} \lambda_t}{\sum_{t=1}^{p} \lambda_t} \times 100\%, i = 1, 2, \ldots, p. \quad (8)$$

Among them, N_j is the contribution rate of the $j - th$ principal component; η_i is the cumulative contribution rate of the first i principal components.

CNN

CNN is the abbreviation of convolutional neural network, which is a variant of multilayer perceptron (MLP), and it was developed by biologists Huber and Wiesel in their early research on cat visual cortex (*Aslan et al., 2022*). Figure 1 shows the structure of CNN networks. The structure of CNN is described in order, including input layer, convolution layer, activation layer, pool layer, full connection layer and output layer. The convolution layer is the core structure of CNN model, which is usually 1×1 matrix, 3×3 matrix and 5×5 matrix. The weights of neurons on the same feature mapping plane in CNN can
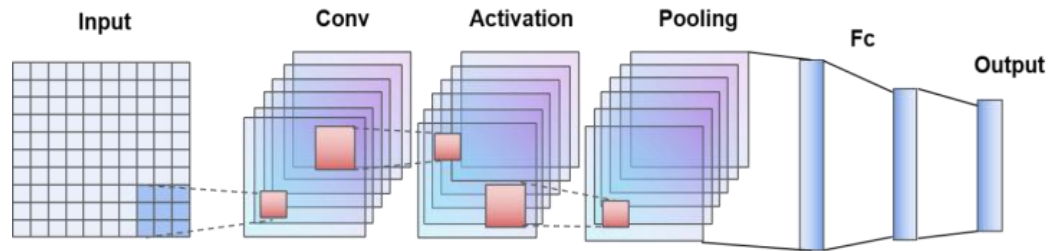
Figure 1 CNN model structure diagram.

be shared locally. Therefore, CNN network supports parallel learning, which can greatly improve the calculation speed and model prediction efficiency. The unique structure of CNN has great advantages in the fields of machine learning, deep learning and prediction field, which is the most widely used depth feature extraction method.

GRU network

Gate recurrent unit (GRU) is a special network structure in neural network (*Ansari, Bartoš & Lee, 2022*), which has only two gate structures of reset gate and update gate, is simpler than the three gate structure of LSTM network and has good prediction effect. These two gating vectors can determine which data can be used as the final output. The basic structure of GRU is shown in Fig. 2.

In Fig. 2, x_t refers to the input data, that is, the high-dimensional feature vector, h_{t-1} refers to the output data of the previous layer, and h_t refers to the output data of the current layer. r_t and z_t are the outputs of reset gate and update gate, and k_t is the candidate set. σ and tanh are *sigmoid* activation function and tanh activation function. The mathematical description of GRU is shown in formula Eq. (9).

$$\begin{cases} z_t = \sigma\left(W_z \cdot \left[h_{t-1}, x_t\right] + b_z\right) \\ r_t = \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right] + b_r\right) \\ k_t = \tanh\left(W_k \cdot \left[r_t \cdot h_{t-1}, x_t\right] + b_k\right) \\ h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot k_t \end{cases} \tag{9}$$

THE PROPOSED METHOD

The substation equipment temperature prediction method proposed in this article is mainly realized by the following five steps:

(1) Correlation analysis. Linear graph correlation, autocorrelation and partial autocorrelation analysis are carried out for the temperature data of substation equipment;

(2) Determine the feature vector of multivariate information fusion. In this article, it includes the features from three aspects of ambient, time and space, which is denoted as *MIFFV*.
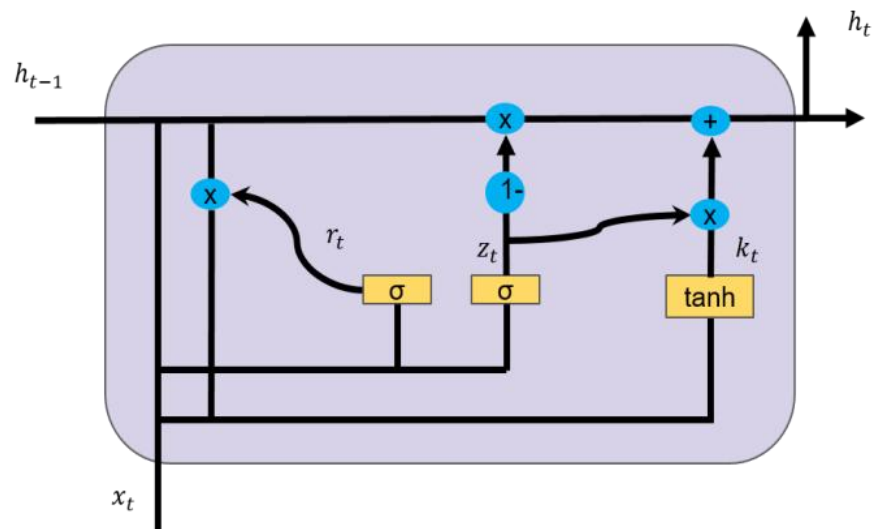
Figure 2   GRU structure diagram.

(3) Obtain the reduced feature vector. PCA is applied to reduce the dimension of multivariate information fusion feature vector to obtain the reduced feature vector, which denoted as *RFV*;

(4) CNN is used to extract the relationship between the reduced feature vector and the equipment temperature in the high-dimensional space, and construct the high-dimensional feature vector of multivariate time series, which is denoted as *HDFV*;

(5) *HDFV* is used to train GRU deep learning network and predict the equipment temperature.

Flow chart of proposed method is shown in Fig. 3.

EXPERIMENTS AND RESULT ANALYSIS

Temperature data acquisition of substation equipment

The research object of this article is primary equipment of main transformer in a substation from Taizhou City, Zhejiang Province. The substation adopts the intelligent inspection system. The temperature of each monitoring point for the equipment is measured by the infrared camera and stored in the form of multi-dimensional intelligent inspection history curve analysis report, including the substation equipment temperature monitoring serial number, organization, measurement position, inspection time, measured value and description (describe the equipment status, whether it is normal or not).

Substation equipment monitoring points are distributed at 110 kV side and 220 kV side, which includes four parts of bushing, conservator, heat sink and panorama. In this article, the data at 220 kV side are selected for the experiment. Monitoring point information is shown in Table 1.

The infrared camera of the equipment is set to monitor once every hour, and the data acquisition time is 15 months from December 11, 2020 to March 10, 2022. However,
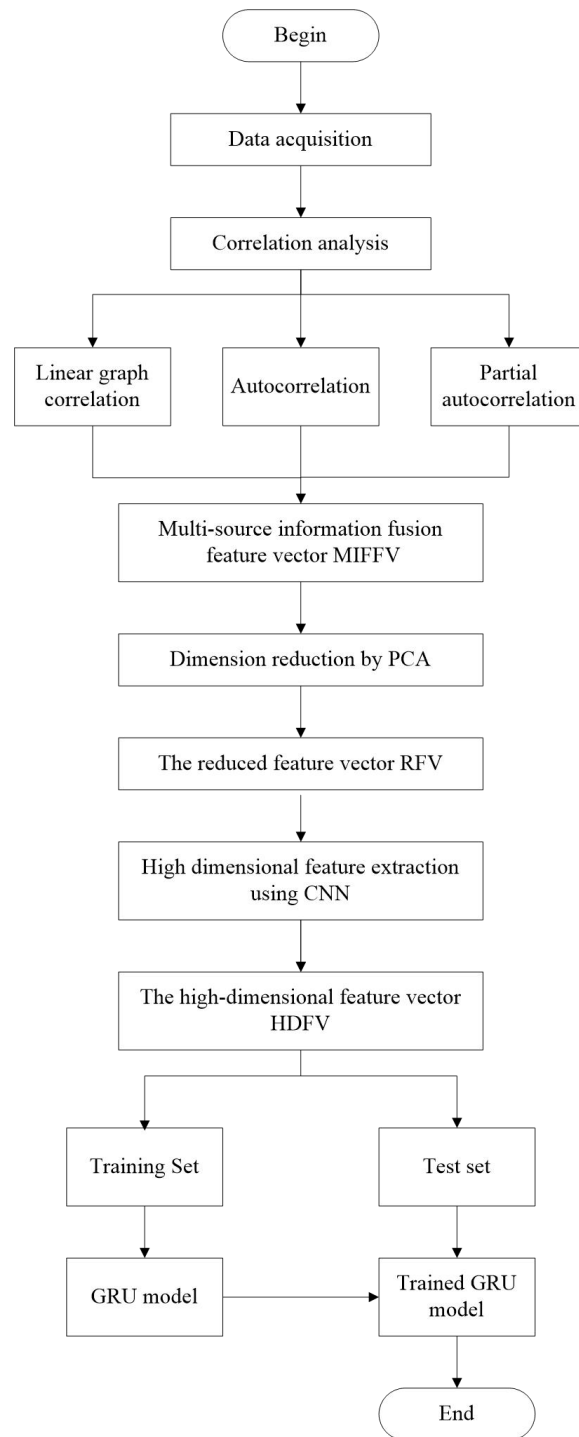
Figure 3 Flow chart of proposed method.

Table 1 Monitoring point information table.	
Serial number	**Monitoring point name**
1	220 kV bushing phase A contact
2	220 kV bushing phase B contact
3	220 kV bushing phase C contact
4	Conservator on 220 side
5	No. 1 heat sink on 220 side
6	No. 2 heat sink on 220 side
7	220 side equipment panorama

there are power outage maintenance and bad points in the monitoring process. Therefore, this article adopts the method of direct elimination, and finally obtains 3,906 effective experimental data.

The selection of data will directly affect the effectiveness of the prediction model. According to the typical seasonal characteristics of temperature, this article selects the data of the first 12 months in the experimental data for training, that is, 3,086 data from December 11, 2020 to December 10, 2021, and 820 data from December 11, 2021 to March 10, 2022.

For the primary equipment of main transformer in substation, the temperature of bushing has the greatest impact on the equipment, so the temperature of A contact from bushing phase is selected as the prediction target for the experiment. Figure 4 shows the thermal imaging diagram of phase A contact at 220kV side bushing on October 1, 2020.

Feature engineering
A. Data Analysis and Feature Selection

Figure 5 shows the temperature data of all monitoring points on 220 kV side from the primary equipment of No. 2 main transformer, and two conclusions can be drawn: (1) With seasonal changes, the equipment temperature also changes significantly. The corresponding performance of the same monitoring point in different seasons is different. The average temperature in winter is about 20 °C and the average temperature in summer is about 50 °C. It can be seen that there is obvious correlation between equipment temperature and environmental factors. Therefore, when predicting the equipment temperature, it is necessary to consider the ambient temperature factor. (2) The temperature trend of different monitoring points for the same equipment shows obvious consistency, which means that there is typical linear correlation between the temperature of equipment space correlation monitoring points.points for the same equipment shows obvious consistency, which means that there is typical linear correlation between the temperature of equipment space correlation monitoring points.

In addition, Fig. 6 shows the correlation analysis results of historical temperature time series from phase A contact of bushing. According to the analysis results of autocorrelation and partial autocorrelation, it can be determined that the temperature time series of phase A contact for bushing is an unstable series. From ACF and PACF between the temperature time series and its first-order difference series, it can be seen that they are trailing, indicating

Figure 4 **Thermal imaging of phase a contact at 220 kV side bushing on October 1, 2020.**

that the historical temperature of substation equipment has strong correlation, and the influence of past time decreases gradually with the passage of time.

In summary, the substation equipment temperature has typical seasonality, periodicity and instability. Therefore, when predicting the equipment temperature, this article determines that the feature vector of multivariate information fusion is composed of the characteristics of ambient, time and space, which is recorded as $MIFFV = [A, T, S]$, where, A refers to the ambient feature, T refers to the time feature and S refers to the space feature. The specific description is as follows:

(1) Ambient feature. In Part A, it is found that the substation equipment temperature is greatly affected by the ambient temperature. Therefore, the weather conditions are taken as the ambient feature in this article, which are recorded as $A = [A_1, A_2, A_3, \ldots\ldots, A_{d1}]$, $d1$ is the dimension of ambient feature. In addition, considering that Zhejiang Province belongs to a typical subtropical monsoon climate, with low temperature and little rain in winter, prevailing northwest wind, high temperature and rain in summer, prevailing southeast wind and muggy, this article determines to take real-time weather temperature and humidity as ambient characteristics to form the ambient feature vector($A = [A_1, A_2]$, that is, set $d1 = 2$). Because the temperature of substation
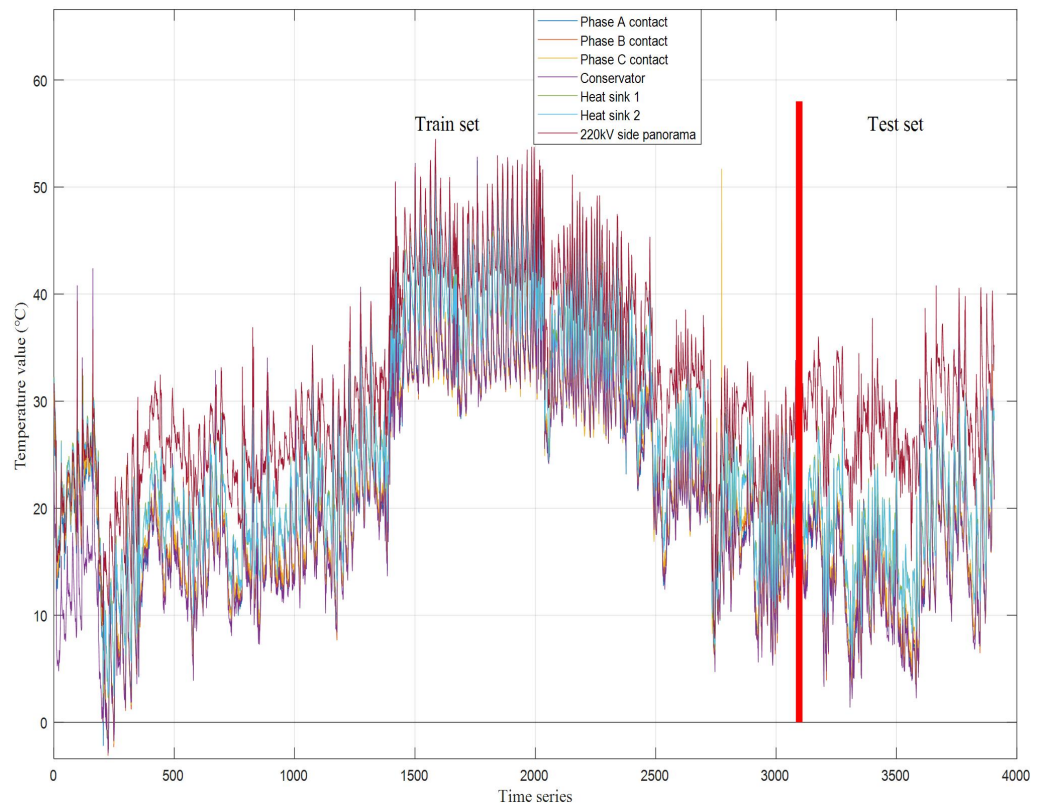
Figure 5 **Temperature trend diagram of all monitoring points on 220 kV side.**

equipment is set to be collected every hour, in order to obtain ambient characteristics, Java programming is used to collect weather conditions every hour through the weather interface of Juhe API (website: http://www.juhe.cn), and two columns of weather temperature and humidity are selected as ambient characteristics.

(2) Time feature. According to the working experience of substation operation and maintenance personnel and the autocorrelation and partial autocorrelation analysis results, the time series of substation equipment temperature has strong lag correlation. Therefore, the historical temperature time series of substation equipment is selected as the time feature vector, which is recorded as $T = [T_1, T_2, \ldots\ldots, T_{d2}]$. Although the lag correlation is relatively large, considering that this article adopts the feature vector of multi information fusion, in order to avoid the inclination of the feature vector in the time feature due to too many time features, the equipment temperature values of the past three times are selected as the time feature, that is $d2 = 3$.

(3) Space feature. The primary equipment of No. 2 main transformer is taken as the research object. For such substation equipment, including 110 kV and 220 kV sides, and both sides are relatively independent, the article selects the temperature of phase A contact for bushing on 220 kV side as the prediction target for the experiment. Therefore the temperature of all monitoring points with space correlation with phase A contact for
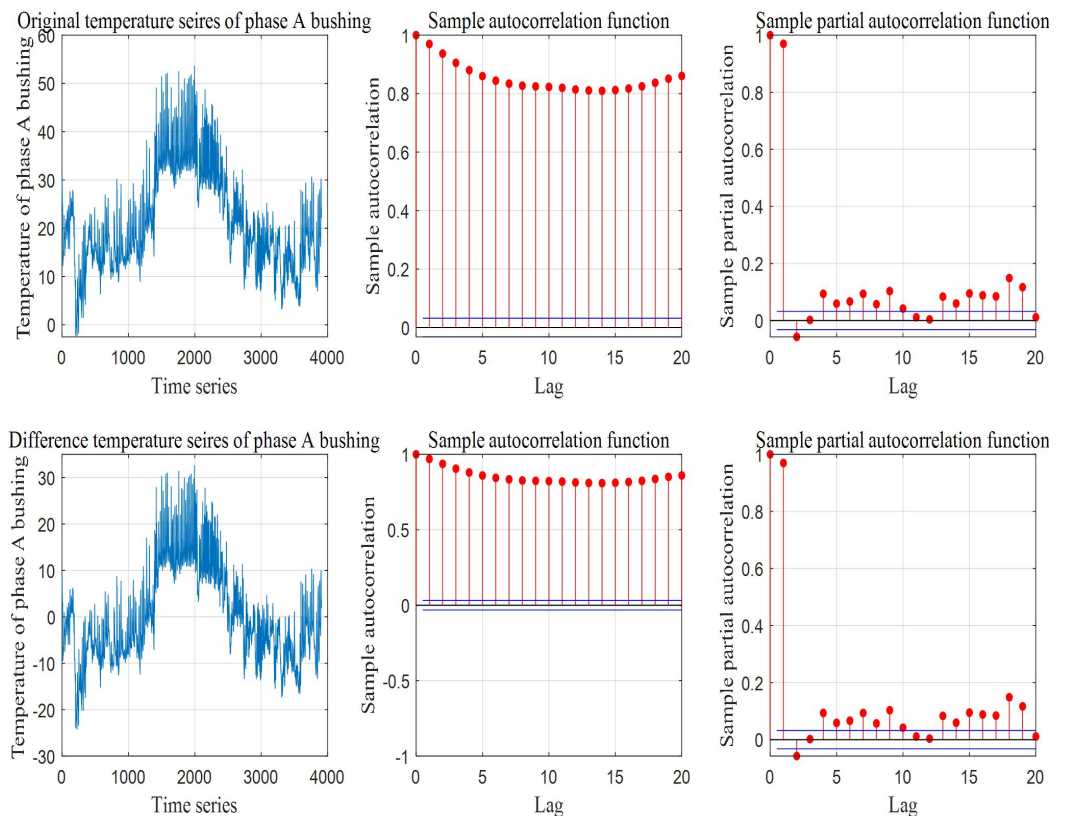
Figure 6 ACF and PACF analysis of equipment temperature time series.

bushing is composed of space feature vector, which is recorded as $S = [S_1, S_2, \ldots, S_{d3}]$. The names of all monitoring points are recorded in Table 1. There are seven infrared temperature monitoring points on the 220 kV side, that is, in addition to the bushing phase A contact, there are six spatial correlation monitoring points, namely bushing B-phase contact, bushing c-phase contact, conservator, No. 1 heat sink, No. 2 heat sink temperature and 220 kV side panoramic temperature. Therefore, set $d3 = 6$.

B. Feature extraction—reduced feature vector based on PCA
There are 11 characteristics in *MIFFV* of multivariate information fusion composed of three aspects of ambient, time and space, which can comprehensively characterize the temperature. While, too much input data can not improve prediction accuracy, but it is easier to produce information redundancy. Therefore, PCA is adopted to reduce the dimension. In general, the eigenvector composed of eigenvalues with cumulative contribution rate of 85%–95% is used as the principal component. Through many experiments, it is verified that the effect of the eigenvalue prediction is the best when the cumulative contribution rate reaches 98%, therefore, the principal components are taken as the reduced feature vector (denoted as *RFV*) under 98% cumulative contribution rate in this article. In the experimental process, PCA dimensionality reduction mapping matrix is shown in Fig. 7, and the feature contribution rate pie chart is shown in Fig. 8.
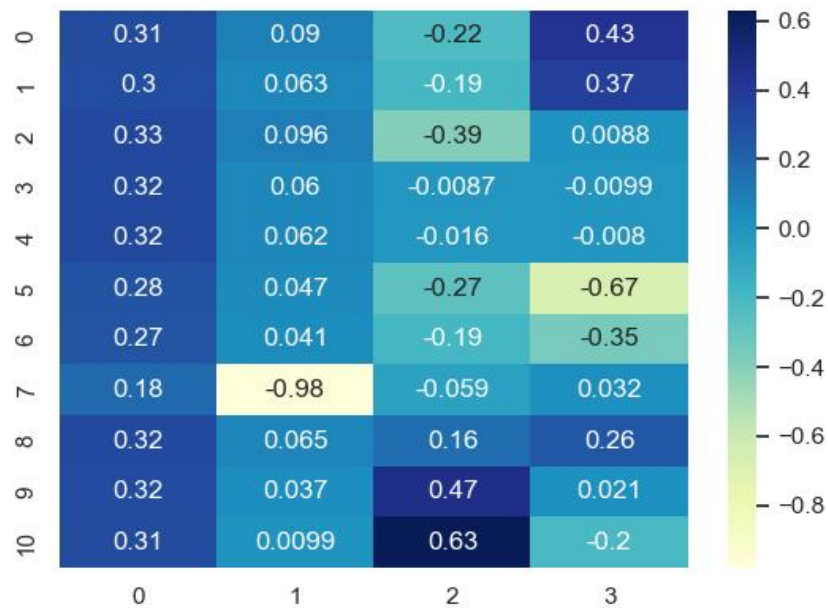
Figure 7 PCA dimensionality reduction mapping matrix.
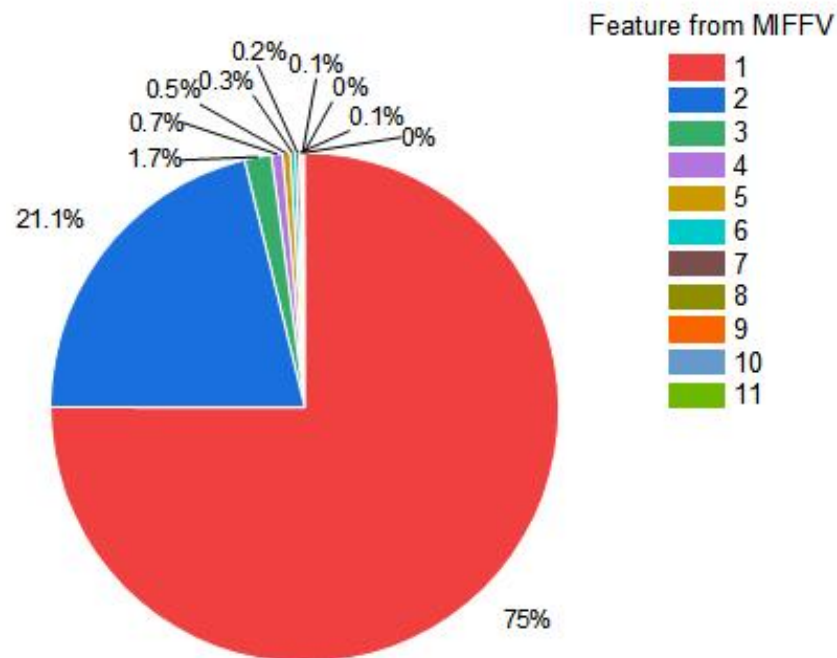
Figure 8 Feature contribution rate pie chart.

C. Feature extraction—depth feature mining based on CNN

Before establishing the prediction model, take advantage of CNN feature extraction, apply it to deep learning, and extract the relationship between reduced feature vector and
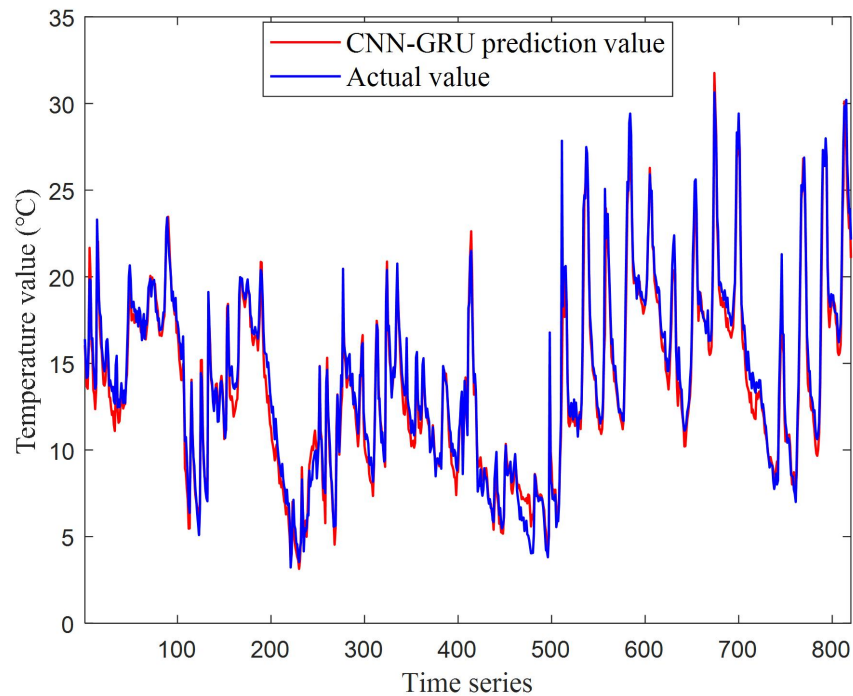
Figure 9 The prediction results for test set based on CNN-GRU.

equipment temperature in high-dimensional space; that is, the reduced feature vector *RFV* obtained by PCA is taken as input data of CNN model, *RFV* of low dimension is mapped to high dimension space, and the high-dimensional feature vector of multivariate time series is constructed, which is *HDFV*, and it is the output of the CNN model.

Temperature prediction of substation equipment

Deep learning network based on CNN and GRU (CNN-GRU) is adopted to predict the phase A contact of bushing, where, CNN filter size is 10; the training cycle is 24 times per round, 60 rounds in total, and the total number of iterations is 1,440; the learning rate is 0.005 and the error threshold is 0.001.

The prediction results for test set based on CNN-GRU are shown in Fig. 9, and the testing relative error is shown in Fig. 9. From the above results, it can be summed up that the temperature prediction effect of bushing phase A contact based on CNN-GRU network is good, the relative error remains between ±0.2, and there is a relatively large error between the sample 450 and 500 in the test set from Fig. 10. The results show that because too many missing points and bad points are eliminated during this period, resulting in the model not obtaining a perfect model for a period of time. In the future, when dealing with missing points, it can be considered using fuzzy c-means clustering and other methods to complete the data to improve the prediction performance of the model.
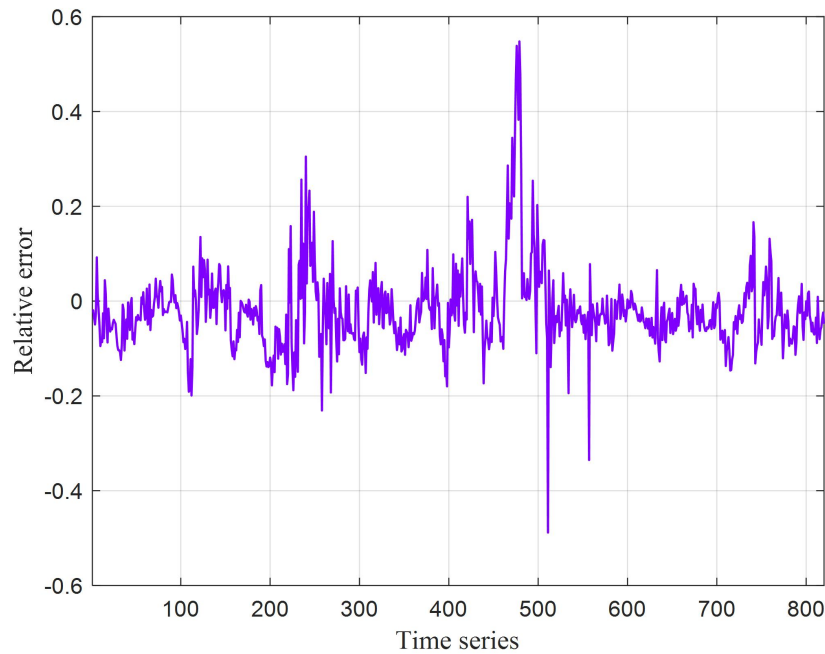
Figure 10 **The testing relative error based on CNN-GRU.**

PREDICTION PERFORMANCE TEST AND RESULT ANALYSIS

Evaluation index of predictive performance

(1) MAPE

MAPE refers to mean absolute percentage error, which is expressed by the formula Eq. (10):

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| \tag{10}$$

where, y_i is true value of equipment temperature, and $\hat{y}_i$ is the predicted value of equipment temperature. The range of MAPE belongs to $(0, +\infty)$, MAPE value of 0% means perfect model, and MAPE value greater than 100% indicates relatively poor model. (2) RMSE

RMSE refers to root mean square error, which is expressed by the formula Eq. (11):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\hat{y}_i - y_i \right)^2} \tag{11}$$

where, y_i and $\hat{y}_i$ means the same with the formula (10); the range of RMSE is $(0, +\infty)$, and the error is positively correlated with RMSE value. When the predicted value is exactly the same as the actual value, it is equal to 0.

Table 2 Prediction performance comparison of CNN-GRU network under different feature vectors.

Feature vector	MAPE	RMSE
A	19.65	244.84
T	12.58	212.14
S	18.52	245.05
MIFFV	6.98	122.12
RFV (MIFFV+PCA)	5.48	95.54

Table 3 Prediction performance comparison results of different models.

Model	MAPE	RMSE
BPNN	5.98	101.47
WaveNN (morlet)	7.93	120.93
LSTM	6.50	131.55
CNN-LSTM	6.26	100.17
CNN-GRU	5.48	95.54

Comparative experiments

Aiming at verifying the effectiveness of this method, comparative experiments from two aspects are carried out in this article:

(1) The prediction performance under different characteristics is compared. Comparative features include only time feature T, only ambient feature A, only spatial feature S, multivariate information fusion feature vector $MIFFV$ and the reduced feature vector RFV, and CNN-GRU network is adopt as the prediction model to predict the temperature of phase A contact. The comparison results are listed in Table 2.

(2) The prediction performance of different models is compared under the same conditions. The reduced feature vector RFV proposed in this article is taken as the input data, and CNN-GRU network is compared with four other network models of BPNN (back propagation neural network), WaveNN (wavelet neural network, in which the Morlet wavelet is adopt), LSTM (long short term networks) and CNN-LSTM. During the comparative experiments, the parameters such as iteration times, learning rate and error threshold are the same in all prediction models. Prediction results of different models are compared in Table 3.

Analysis of prediction results

According to the above comparative experiments, this article analyzes the prediction results from multiple angles and draws the following conclusions from the statistical results from Tables 2 and 3:

(1) CNN-GRU was applied to the prediction performance comparison experiment under different feature conditions, and the results showed that the multi-source information fusion feature vector constructed from the three aspects of ambient, time and space is better than the single feature prediction effect, in which MAPE and RMSE were reduced by one order of magnitude; that is, $MIFFV$ includes rich information than the A, T and S feature;

(2) The reduced feature vector *RFV* composed of principal components extracted after PCA dimensionality reduction had better prediction performance than *MIFFV* (MAPE is decreased from 6.98 to 5.48, and RMSE is decreased from 122.12 to 95.54), which shows that feature extraction plays a significant role in the prediction process, and the feature engineering scheme proposed in this article has the best effect on the temperature prediction of substation equipment.

(3) Compared with CNN-LSTM, CNN-GRU had better performance, which shows that although GRU with two gating structures are simpler than LSTM three gating structures, GRU has better effect in temperature prediction of substation equipment;

(4) CNN-LSTM had better effect than LSTM, which shows that CNN can mine the characteristics of equipment temperature depth when it is used for high-dimensional feature extraction, and provides a guarantee for the prediction model to achieve better prediction effect;

(5) The depth network models of LSTM, CNN-LSTM and CNN-GRU had better prediction effect than the shallow networks of BPNN and WaveNN shallow networks, which shows that the deep learning network has obvious advantages in the field of prediction compared with the shallow networks in traditional machine learning.

CONCLUSIONS

In the process of substation equipment temperature prediction, the prediction effect is not ideal due to less information sources; the problem is solved from the two links of feature engineering and prediction modeling. In the aspect of feature engineering, linear graph correlation, autocorrelation and partial autocorrelation function analysis are applied to establish the feature vector of multi-source information fusion from the three aspects of environment, time and space. After PAC dimension reduction, the principal component is obtained as the reduced feature vector. Finally, the equipment temperature is predicted through CNN-GRU double-layer depth network model, in which CNN realizes depth feature extraction. The effectiveness of this method is fully proved by comparative experiments from two aspects of different feature vectors and different prediction models. However, in practice, it is usually necessary to obtain the equipment temperature at more times in advance, so the next goal is to realize the multi-step accurate prediction of substation equipment temperature.

Funding
The authors received no funding for this work.

Competing Interests
Ying Wang is employed by Economic and Technological Research Institute of State Grid Heilongjiang Electric Power Co. Ltd. Zhaohong Bing is employed by Computing Technology Institute of East China. The authors declare there are no competing interests.

Author Contributions

- Lijie Sun conceived and designed the experiments, prepared figures and/or tables, data acquisition, supervision, and approved the final draft.
- Chunxue Liu performed the experiments, prepared figures and/or tables, and approved the final draft.
- Ying Wang analyzed the data, authored or reviewed drafts of the article, and approved the final draft.
- Zhaohong Bing performed the computation work, authored or reviewed drafts of the article, and approved the final draft.

Data Availability

The raw data and code are available in the Supplemental Files.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.1172#supplemental-information.

REFERENCES

Ansari MS, Bartoš V, Lee B. 2022. GRU-based deep learning approach for network intrusion alert prediction. *Future Generation Computer Systems* **128**:235–247 DOI 10.1016/j.future.2021.09.040.

Aslan N, Koca GO, Kobat MA, Dogan S. 2022. Multi-classification deep CNN model for diagnosing COVID-19 using iterative neighborhood component analysis and iterative ReliefF feature selection techniques with X-ray images. *Chemometrics and Intelligent Laboratory Systems* **224**:104539 DOI 10.1016/j.chemolab.2022.104539.

Baptista M, Sankararaman S, deMedeiros IP, Nascimento Jr C, Prendinger H, Henriques EMP. 2018. Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. *Computers & Industrial Engineering* **115**:41–53 DOI 10.1016/j.cie.2017.10.033.

Bussière W, Rochette D, Clain S, Andréa P, Renard JB. 2017. Pressure drop measurements for woven metal mesh screens used in electrical safety switchgears. *International Journal of Heat and Fluid Flow* **65**:60–72 DOI 10.1016/j.ijheatfluidflow.2017.02.008.

Cao H, Sun P, Zhao L. 2022. PCA-SVM method with sliding window for online fault diagnosis of a small pressurized water reactor. *Annals of Nuclear Energy* **171**:109036 DOI 10.1016/j.anucene.2022.109036.

Cao K, Jiang M, Gao S. 2021. Spectrum availability prediction based on RCS-GRU model. *Physical Communication* **49**:101479 DOI 10.1016/j.phycom.2021.101479.

Chachlakis DG, Zhou T, Ahmad F, Markopoulos PP. 2021. Minimum Mean-Squared-Error autocorrelation processing in coprime arrays. *Digital Signal Processing* **114**:103034 DOI 10.1016/j.dsp.2021.103034.

Dai S. 2021. Quantum cryptanalysis on a multivariate cryptosystem based on clipped hopfield neural network. *IEEE Transactions on Neural Networks and Learning Systems* **33(9)**:5080–5084 DOI 10.1109/TNNLS.2021.3059434.

Esfahani HN, Song Z, Christensen K. 2020. A deep neural network approach for pedestrian trajectory prediction considering heterogeneity. In: *99th annual meeting of the Transportation Research Board*. Washington, D.C., USA.

Gharehbaghi A, Ghasemlounia R, Ahmadi F, Albaji M. 2022. Groundwater level prediction with meteorologically sensitive Gated Recurrent Unit (GRU) neural networks. *Journal of Hydrology* **612**:128262.

Guo WQ, Dong Y, Li QH, Zhang MM, Wang LX. 2020. Application of PSO-BP neural network in temperature prediction of switchgear equipment. *Journal of Shanxi University of Science and Technology* **38(01)**:149–153.

Hao XH, Yang ZJ, Hao Y, Han ZX, Ma H, Zhang DX. 2021. Medium and long term prediction method of wind power based on improved gm-arma combination model. *Electrotechnical* **7**:11–13.

Hou J, Chen FW, Li PH, Zzhu HQ. 2021a. Gray-box parsimonious subspace identification of Hammerstein-type systems. *IEEE Transactions on Industrial Electronics* **68(10)**:9941–9951 DOI 10.1109/TIE.2020.3026286.

Hou JW, Wu WCH, Li LF, Tong X, Hu RJ, Wu WB, Cai WZH, Wang HL. 2022. Estimation of remaining capacity of lithium-ion batteries based on X-ray computed tomography. *Journal of Energy Storage* **55**:105369 DOI 10.1016/j.est.2022.105369.

Hou YY, Zheng ER, Guo WQ, Li JW, Dong Y. 2021b. Prediction of switchgear equipment based on long-term and short-term memory cyclic neural network. *Journal of Shaanxi University of Science and Technology* **39(4)**:148–155.

Huang KQ, Zheng RF, Qu FF, Liu YF, Lu XB, Huang X. 2022a. Phase failure analysis of 110 kV GIS disconnector. *Sichuan Electric Power Technology* **45(01)**:87–90.

Huang MH, Jiang T, Dong JJ, Wang K, Zhao HSH. 2022b. High voltage bushing temperature prediction of box transformer based on LSTM. *Electrical Measurement and Instrument* **04(06)**:1–7.

Khalifani S, Darvishzadeh R, Azad N, Rahmani RS. 2022. Prediction of sunflower grain yield under normal and salinity stress by RBF, MLP and, CNN models. *Industrial Crops and Products* **189**:115762 DOI 10.1016/j.indcrop.2022.115762.

Kong XH. 2015. Design of substation equipment temperature early warning system. Master's thesis, Control Engineering, Jinan University, Guangzhou, China.

Kong XH, Zhang HF. 2016. Substation equipment temperature prediction based on Optimized Generalized Regression Neural Network. *China Power* **49(09)**:54–59.

Liu ZL. 2012. Research on temperature monitoring and early warning management system of electromechanical equipment based on BP neural network. Master's thesis, Management Science and Engineering, Taiyuan University of Technology, Taiyuan, Shanxi, China.

Mestre G, Portela J, Rice G, San Roque AM, Alonso E. 2021. Functional time series model identification and diagnosis by means of auto- and partial auto-correlation analysis. *Computational Statistics & Data Analysis* **155**:107108 DOI 10.1016/j.csda.2020.107108.

Mohammadshirazi A, Kalkhorani VA, Humes J, Speno B, Rike J, Ramnath R, Clark JD. 2022. Predicting airborne pollutant concentrations and events in a commercial building using low-cost pollutant sensors and machine learning: a case study. *Building and Environment* **213**:108833 DOI 10.1016/j.buildenv.

Song HF, Yang WW. 2022. GSCCTL: a general semi-supervised scene classification method for remote sensing images based on clustering and transfer learning. *International Journal of Remote Sensing* **43(15–16)**:5976–6000 DOI 10.1080/01431161.2021.2019851.

Sun HH. 2019. Research on emergency capacity evaluation of large-area power outage based on scenario construction. Master's thesis, Safety Engineering, Capital University of Economics and Business, Beijing, China.

Sun LJ, Chen S, Zhu JF, Li JH. 2022. Substation equipment temperature prediction method considering local spatio-temporal relationship. *Scientific Programming* **2022**:4414093 DOI 10.1155/1970/4414093.

Velásquez RMA, Lara JVM, Melgar A. 2019. Reliability model for switchgear failure analysis applied to ageing. *Engineering Failure Analysis* **101**:36–60 DOI 10.1016/j.engfailanal.2019.03.004.

Wang CQ, Yang CH, Sun YT, Zhang TQ, Xu YP. 2015. Infrared temperature prediction method of substation equipment based on improved RBFNN. *Electrical Application* **34(23)**:6669–6674.

Wang T. 2015. Research on substation equipment temperature prediction system based on adaptive neural network algorithm. *Information and Computer* **24**:51–54.

Wang T. 2016. Research on temperature prediction and fault early warning of substation electrical equipment. M.S. thesis, Department of Industrial and Manufacturing System Engineering Huazhong Science and Technology University, Wuhan, China.

Xu GJ, Xu C, He J. 2016. Research on temperature early warning of high voltage switchgear based on information fusion technology. *Electrotechnical* **10**:18–20.

Xu XF, Hao J, Zheng Y. 2020. Multi-objective artificial bee colony algorithm for multi-stage resource leveling problem in sharing logistics network. *Computers & Industrial Engineering* **142(4)**:106338 DOI 10.1016/j.cie.2020.106338.

Xu XF, Lin ZR, Zhu J. 2020. DVRPLS with variable neighborhood region in refined oil distribution. *Annals of Operations Research* **309**:663–687 DOI 10.1007/s10479-020-03780-9.

Yu T, Gan Q, Feng G, Han G. 2022. A new fuzzy cognitive maps classifier based on capsule network. *Knowledge-Based Systems* **250**:108950 DOI 10.1016/j.knosys.2022.108950.

Yuan Q, Ma CH, Liu JL, Gui HQ, Li MY, Swang HL. 2022. Correlation analysis-based thermal error control with ITSA-GRU-A model and cloud-edge-physical collaboration framework. *Advanced Engineering Informatics* **54**:101759 DOI 10.1016/j.aei.2022.101759.

Zeng JY, Chen GY, Hou H, Bo W, Zeng QY. 2018. Risk assessment of switchgears in distribution system considering environmental dependency. *Procedia Computer Science* **130**:1140–1145 DOI 10.1016/j.procs.2018.04.157.

Zhang H, Srinivasan R, Yang X, Ahrentzen S, Coker ES, Alwisy A. 2022. Factors influencing indoor air pollution in buildings using PCA-LMBP neural network: a case study of a university campus. *Building and Environment* **225**:109643 DOI 10.1016/j.buildenv.2022.109643.

Zhang HY, Zhang DL, Zhao Y, Wang Z. 2020. Temperature prediction of switchgear based on least squares support vector machine optimized by PCA and GA algorithm. *Electrical Applications* **39(02)**:59–63.

Zheng G, Chai WK, Katos V. 2022. A dynamic spatial–temporal deep learning framework for traffic speed prediction on large-scale road networks. *Expert Systems with Applications* **195**:116585 DOI 10.1016/j.eswa.2022.116585.