

Article

Research on Transformer Temperature Early Warning Method Based on Adaptive Sliding Window and Stacking

Pan Zhang ¹, Qian Zhang ¹, Huan Hu ¹, Huazhi Hu ¹, Runze Peng ^{2,*}  and Jiaqi Liu ²

¹ State Grid Hubei Electric Power Co., Ltd., Xiaogan Power Supply Company, Xiaogan 432000, China; zhangp8@hb.sgcc.com.cn (P.Z.); zhangq42@hb.sgcc.com.cn (Q.Z.); huh24@hb.sgcc.com.cn (H.H.); huhz21@hb.sgcc.com.cn (H.H.)

² School of Computer Science and Engineering, Central South University, Changsha 410083, China; liujiaqi@csu.edu.cn

* Correspondence: 244703045@csu.edu.cn

Abstract: This paper proposes a transformer temperature early warning method based on an adaptive sliding window and stacking ensemble learning algorithm, aiming to improve the accuracy and robustness of temperature prediction. The transformer temperature early warning system is crucial for ensuring the safe operation of the power system, and temperature prediction, as the foundation of early warning, directly affects the early warning effectiveness. This paper analyzes the characteristics of transformer temperature using support vector regression, random forest, and gradient boosting regression as base learners and ridge regression as the meta-learner to construct a stacking model. At the same time, Bayesian optimization is used to automatically adjust the sliding window size, achieving adaptive sliding window processing. The experimental results indicate that the temperature prediction method based on adaptive sliding window and stacking significantly reduces prediction errors, enhances the model's adaptability and generalization ability, and provides more reliable technical support for transformer fault warning.

Keywords: transformer; adaptive sliding window; stacking; Bayesian optimization



Academic Editors: Nilanjan Ray Chaudhuri and Ali Mehrizi-Sani

Received: 6 December 2024

Revised: 15 January 2025

Accepted: 15 January 2025

Published: 18 January 2025

Citation: Zhang, P.; Zhang, Q.; Hu, H.; Hu, H.; Peng, R.; Liu, J. Research on Transformer Temperature Early Warning Method Based on Adaptive Sliding Window and Stacking. *Electronics* **2025**, *14*, 373. <https://doi.org/10.3390/electronics14020373>

Copyright: © 2025 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

As an important piece of equipment in the power system, the transformer is in a high-current and high-voltage working state for a long time, which makes it easy to generate a lot of heat. Excessive temperature may cause aging of insulation materials, deterioration of oil quality, burning of windings, and even cause serious accidents such as fire. Therefore, timely monitoring of the temperature change of the transformer can detect temperature anomalies and issue early warnings in time, helping operation and maintenance personnel to intervene promptly to avoid equipment damage and power supply interruption. Temperature anomalies are often a precursor to transformer failure. Timely temperature monitoring can effectively prevent transformer failures caused by overheating. Therefore, the transformer temperature early warning system can promptly provide temperature anomaly information to operation and maintenance personnel through real-time monitoring and early warning functions, avoid long-term high-temperature operation, and thus extend the service life of the transformer. In addition, the transformer is an important part of the power system, and its safe operation is directly related to the stability of the power supply. Temperature early warning can reduce the occurrence of sudden failures and outages and improve the reliability of the power system. This paper divides the research on transformer temperature early warning methods into two kinds:

transformer temperature early warning based on real-time monitoring and transformer temperature early warning based on predicted temperature.

Early warning research based on real-time transformer temperature monitoring uses the real-time collection, analysis, and alarm of transformer temperature to remind the operation and maintenance personnel of abnormal temperature conditions in a timely manner. Its main feature is the immediate response to the current temperature condition, ensuring that attention can be drawn before the temperature exceeds the safety threshold, making it suitable for early fault detection and real-time response. Transformer temperature early warning based on real-time monitoring temperature early warning usually consists of a temperature sensor module, a data acquisition and transmission module, a data analysis and processing module, and an early warning module. Among them, the temperature sensor is the core of the system, which is used to monitor the temperature of key parts of the transformer in real time. These sensors are installed in transformer windings, cores, oil tanks, etc., to obtain temperature data from these parts. Commonly used temperature sensors include thermocouples, optical fiber temperature sensors, and infrared temperature sensors. The temperature data acquisition and transmission module transmits the temperature data obtained by the sensor to the monitoring center or management platform in a timely manner. The data transmission methods usually include wired transmission, wireless transmission, and edge computing. Among them, edge computing is used to arrange edge computing nodes at the sensor end or near the transformer to perform preliminary data processing, reduce transmission pressure, and improve the real-time performance of data transmission [1]. In addition, temperature data need to be processed and analyzed to identify abnormal temperature changes. Common data processing techniques include using algorithms such as low-pass filtering [2] and Kalman filtering [3] to remove noise and abnormal points in the data to improve data accuracy, and algorithms based on temperature fluctuation analysis, such as moving average [4] and sliding window [5], to detect whether the temperature change exceeds the normal range. Finally, the early warning module triggers an alarm based on preset thresholds or rules through real-time analysis of temperature data, reminding the operation and maintenance personnel to take intervention measures. The system usually sets multiple early warning levels [6], such as mild warning, moderate warning, severe warning, etc. Each level corresponds to different temperature thresholds, alarm information, and response methods.

With the development of technology, the real-time temperature monitoring and early warning system of transformers is also constantly improving and optimizing. Multiple sensors are used for data fusion [7], and the accuracy and reliability of temperature monitoring are improved through complementary data from different types of sensors. For example, the temperature sensor and current sensor are combined to monitor the impact of excessive load on temperature, or the oil temperature, winding temperature, and ambient temperature data are combined to more accurately predict faults. With the development of big data and cloud computing, the transformer temperature monitoring system is gradually transforming to a cloud platform. By uploading temperature monitoring data of the transformer to the cloud, operation and maintenance personnel can realize remote monitoring, fault analysis, and historical data queries through the cloud platform. Big data analysis technology can help identify potential fault modes, make intelligent predictions, and optimize scheduling based on big data models [8]. There is also a proposal to use electromagnetic energy transmission and wireless sensor equipment to establish a transformer monitoring system [9], which can help staff to accurately evaluate transformer alarm faults in a timely manner. The core of the transformer temperature early warning research, based on predicted temperature, is used to predict the temperature variation in advance by modeling and analyzing the trend of transformer temperature data, issue early warnings,

and take corresponding preventive measures. Predictive temperature warning research uses historical data, load data, environmental factors, and other factors combined with intelligent algorithms and big data analysis to predict future temperatures so that active protection measures can be taken before abnormal temperatures arrive [10]. This warning method is more forward-looking than early warning research based on the real-time transformer temperature monitoring using the real-time collection. It can predict abnormal temperature rises and deal with them in advance to avoid power outages caused by sudden failures. It can also reasonably arrange inspections and maintenance through predictions, reduce unnecessary shutdowns and repairs, improve equipment utilization, and help further ensure the safety and stability of the power system. The essential difference between this method and the early warning method based on real-time transformer temperature monitoring using the real-time collection is that it constructs a temperature prediction model through data modeling, feature extraction, and prediction algorithm selection.

Transformer temperature is affected by many factors, including load, current, voltage, ambient temperature, humidity, etc. Therefore, when modeling data, it is necessary to comprehensively consider the key factors affecting temperature and construct a feature vector that reflects the temperature change law of the transformer. The prediction algorithm is the core of the prediction temperature early warning. Commonly used algorithms include the ARIMA model [11], exponential smoothing [12], and other time series analysis algorithms, which can be used for short-term temperature trend prediction; regression analysis, support vector machine [13], random forest, and other machine learning models, which can achieve higher accuracy temperature prediction through big data training; and long short-term memory network [14] and other deep learning algorithms, which are particularly suitable for capturing the nonlinear temperature change law. In addition, combining multiple models and integrating the advantages of different models can improve the prediction effect. For example, ref. [15] combines integrated empirical mode decomposition and the long short-term memory network to realize a transformer temperature prediction model with high noise resistance. Moreover, ref. [16] extract temperature factors and load factors as feature data to train a fully connected neural network model for transformer temperature prediction.

The limitation of existing transformer temperature warning technology is that it is difficult to obtain transformer temperature data in real power grid scenarios, and many existing research methods are not suitable for small sample data. Therefore, this article proposes a stacking model suitable for small sample data, which combines the advantages of multiple models to improve the robustness of overall model prediction. In addition, this article proposes an adaptive sliding window algorithm to improve the accuracy of model prediction.

2. Stacking Ensemble Learning Model

There are a lot of machine learning and deep learning methods used in transformer temperature prediction, as described in the Section 1. This paper chooses to establish a transformer temperature prediction model based on an ensemble learning algorithm because it integrates multiple models to obtain higher prediction accuracy, stronger robustness, and a more flexible model structure, and performs well in dealing with complex data, diverse features, and noise resistance. Transformer temperature prediction involves multiple complex factors such as load, ambient temperature, historical temperature trends, etc. A single model is often difficult to accurately model, while ensemble learning can combine the advantages of multiple models to make prediction results more comprehensive. In addition, a single model may be prone to overfitting or underfitting when facing complex or noisy transformer temperature data, while ensemble learning is usually more robust.

2.1. Stacking Model Analysis

Commonly, ensemble learning includes bagging, boosting, and stacking. Among these, bagging is usually based on the same type of model, such as random forest, and these models have relatively limited adaptability to nonlinear features. In transformer temperature prediction, the data may have strong nonlinearity or multivariate correlations, the effect of bagging may be limited, and the temporal nature of the data are not considered, so it cannot capture long-term and short-term trends well in the time series data of transformer temperature prediction. The boosting method will continuously strengthen the error samples of each round, so it is sensitive to noise data. In temperature data, if there is a large fluctuation or erroneous data input, the boosting method may produce a high error, and in the case of complex data and high noise, the boosting method is more likely to overfit, resulting in a decrease in generalization ability, and the volatility of transformer temperature data will increase the risk of overfitting. Stacking can combine multiple different types of basic models and further optimize them on the secondary learner. In contrast, bagging and boosting are usually based on the same types of basic models. Stacking allows the combination of multiple algorithms to capture the linear, nonlinear, and temporal characteristics of temperature data using the advantages of various models. In transformer temperature prediction, stacking has higher flexibility and generalization than bagging and boosting. Its multi-model and multi-level fusion method enables it to better adapt to complex and changeable temperature data and improve the accuracy and stability of prediction. Therefore, this paper chooses to establish a transformer temperature prediction model based on stacking.

Stacking is an ensemble learning method that improves the performance of the model by combining the prediction results of multiple base learners. The basic idea is to train multiple base learners to learn different parts of the data, respectively, and then use a meta-learner to integrate the outputs of these base learners to form the final prediction results. The overall stacking model structure proposed in this paper is shown in Figure 1.

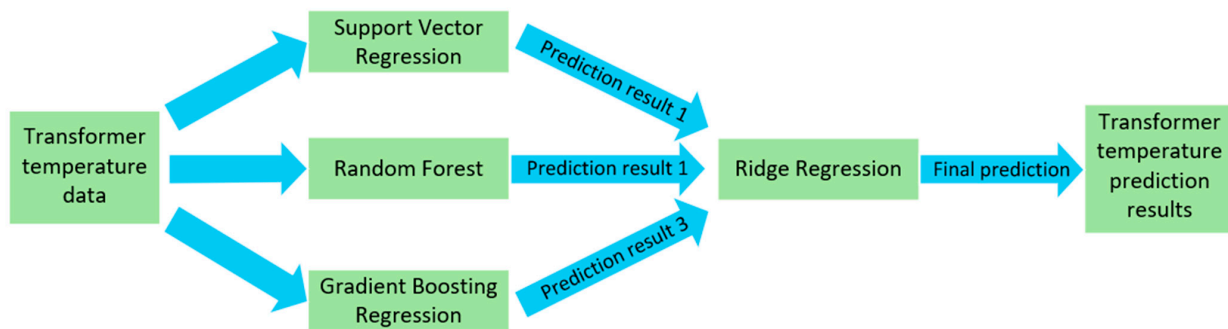


Figure 1. Stacking model structure diagram.

Base learners are the key to the effect of the stacking model, and include support vector regression, random forest, and gradient boosting regression in this paper; the meta-learner is usually a simple model, which, in this paper, is ridge regression. This kind of method can reduce the bias and variance of the prediction and improve the generalization ability of the model by combining multiple base learners. The meta-learner can capture more information through the prediction results of different models. Selecting base learners with greater diversity can improve the effect of stacking. Therefore, this paper selects support vector regression, random forest, and gradient boosting regression with their own characteristics as base learners. Among them, the support vector regression algorithm is suitable for handling nonlinear problems in high-dimensional space and is applicable to small sample data. The random forest algorithm reduces the risk of overfitting through random sampling and feature selection. Gradient boosting regression has stronger flexibility and can provide very high prediction accuracy. This article aims to combine the advantages of three models

and develop a prediction algorithm model that is suitable for small sample data and has high robustness and accuracy.

2.2. Support Vector Regression

Support vector regression is a regression method based on support vector machines that aims to make predictions by minimizing the trade-off between model complexity and training errors. The support vector regression model finds an optimal regression hyperplane such that most training data points lie within a specific tolerance band, using a function to fit the data, as shown in Formula (1):

$$f(x) = \omega^T \Phi(x) + b \quad (1)$$

Among them, ω is the weight vector, x is the input feature, $\Phi(x)$ is the feature mapping, and b is the bias. This model mainly includes selection parameters, namely the width of the margin of tolerance ε and the penalty parameter C . ε defines the tolerance range for prediction errors, and C controls the degree of penalty for training errors. An optimization problem is constructed, which minimizes a loss function to balance the model's complexity and fitting error, as shown in Formula (2):

$$F = \min \left(\frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n (\zeta_i + \zeta_i^*) \right) \quad (2)$$

Among them, ζ_i and ζ_i^* are slack variables used to handle the situation where data points exceed the tolerance band; constraints are added to ensure that the prediction error for each data point does not exceed ε , as shown in Formula (3):

$$\begin{cases} y_i - (\omega^T \Phi(x) + b) \leq \varepsilon + \zeta_i \\ (\omega^T \Phi(x) + b) - y_i \leq \varepsilon + \zeta_i^* \end{cases} \quad (3)$$

Among them, y_i is the true target value of the i -th training sample, and $\omega^T \Phi(x) + b$ represents the model's predicted value; to solve the optimization problem, the Lagrange multiplier method is used to combine the objective function and the constraints, constructing the Lagrange function and solving it to obtain the optimal weight vector ω and bias b ; model prediction is used when predicting new data points. Support vector regression effectively captures the complex relationships within data by finding the optimal regression hyperplane in the high-dimensional feature space, combining the tolerance band and penalty mechanism, and is suitable for a variety of regression tasks.

2.3. Random Forest

Random forest is an ensemble learning method, mainly used for classification and regression tasks, that improves the accuracy and robustness of the model by constructing multiple decision trees and combining their predictions. The steps include the following: randomly drawing samples from the original training dataset with put-back to form multiple different training subsets, each of which is used to train an independent decision tree; selecting a specific number of features randomly to evaluate at each node split of each decision tree instead of using all of them, in which randomization increases the diversity of the model and reduces the risk of overfitting; constructing multiple decision trees based on the pocketed samples and the randomly selected features. Multiple decision trees are constructed, each of which can have a different structure because they are trained on different subsets of data and feature sets; finally, for the regression task, the average of the predictions from all trees is taken. Random forest utilizes the integration of multiple decision trees to improve prediction accuracy and robustness. By introducing randomness,

random forest performs well in dealing with complex data. It is suitable for classification and regression problems and is one of the most widely used algorithms in machine learning.

2.4. Gradient Boosting Regression

Gradient boosting regression minimizes the loss function by gradually building weak learners, thereby improving the predictive ability of the model. Its steps include initialization, calculating the mean $F_0(x)$ of the target value, as shown in Formula (4):

$$F_0(x) = \frac{1}{n} \sum_{i=1}^n y_i \quad (4)$$

where y_i is the target value of the i -th training sample; iterative training, for each round $m = 1, 2, \dots, M$, calculate the residual, as shown in Formula (5):

$$r_i^{(m)} = y_i - F_{m-1}(x_i) \quad (5)$$

The weak learner $h_m(x)$ is then trained, that is, a new weak learner is trained using the residual $r_i^{(m)}$ as the target, and then update the model, as shown in Formula (6):

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (6)$$

Among them, γ_m is the learning rate, which is used to control the influence of the new model on the general model; the final prediction is to predict the new sample. Gradient boosting regression has performed well in many machine learning competitions and practical applications and is a very popular choice in regression tasks. By optimizing the loss function, gradient boosting regression can effectively capture complex patterns in the data.

2.5. Ridge Regression

The meta-learner of the Stacking model in this paper is ridge regression, which constrains the complexity of the model by introducing L2 regularization terms, thereby improving the generalization ability of the model. The process includes selecting regularization parameters and determining regularization strength λ . This parameter is used to control the complexity of the model. The larger the value, the simpler the model is and the more likely it is to be underfitted. The smaller the value, the more complex the model is and the more likely it is to be overfitted. When constructing an optimization problem, the goal of ridge regression is to minimize the loss function, as shown in Formula (7):

$$L(\beta) = \sum_{i=1}^n (y_i - X_i \beta)^2 + \lambda \sum_{j=1}^p \beta_j^2 \quad (7)$$

Here, β is the regression coefficient and X_i is the eigenvector of the i -th sample; solve the optimization problem and obtain the regression coefficient β by minimizing the loss function. The analytical solution of ridge regression is shown in Formula (8):

$$\beta^* = (X^T X + \lambda I)^{-1} X^T y \quad (8)$$

Among them, I is the identity matrix and λ is the regularization parameter. In Stacking, ridge regression as a meta-learner can make full use of its robustness and ability to process high-dimensional data, thereby enhancing the overall performance of the ensemble learning model.

3. Adaptive Sliding Window

Transformer temperature prediction belongs to time series prediction. In time series prediction, the sliding window algorithm is a commonly used data processing method. By

creating subsequences to help the model capture trends and patterns in time series, time series data can be transformed into supervised learning problems, so that regression models and neural networks can better learn the dependencies of sequences. This paper also combines the sliding window algorithm when training the Stacking model, and uses Bayesian optimization to find the optimal sliding window size to implement an adaptive sliding window algorithm.

3.1. Sliding Window Algorithm

The basic idea of the sliding window algorithm is to use a certain period of historical data of the sequence data to predict one or more data points in the future so that non-time series models can also be used for time series prediction. The algorithm needs to determine the window size W first. The size of the sliding window determines the amount of data used in each time the window is slid, that is, the input length of each training sample, and is also a key factor affecting the final time series prediction effect; each time the window is slid, the first $W - 1$ data points in the window are used as input features, and the last data point is used as the target output. The window is continuously slid forward to generate a series of input-output pairs; the generated sample is used to train the model, that is, the support vector regression, random forest, and gradient boosting regression models, and finally stacked into a stacking model.

The sliding window converts the time series into input and output pairs, so that non-sequence models can also be used for time series prediction, capturing local time dependencies, that is, short-term trends within the window, so that the model has better performance in short-term prediction. However, the sliding window model usually generates prediction results step by step, and the error will accumulate as the number of prediction steps increases. Therefore, the selection of the sliding window size is very critical. However, in the transformer temperature prediction, the transformer temperature time series is affected by many factors, and its change trend cycle is different in different seasons and different environments. Using a fixed size sliding window constantly will lead to a decrease in the general prediction accuracy of the prediction model and a loss of its robustness. Therefore, this paper proposes to use the Bayesian optimization algorithm to find the optimal sliding window size under the current data to achieve an adaptive sliding window and improve the accuracy and robustness of the overall transformer temperature prediction model.

3.2. Bayesian Optimization

Bayesian optimization is an efficient global optimization algorithm suitable for optimizing black box functions with high optimization costs. Its main idea is to approximate the objective function by building a proxy model and select the next best sampling point based on the prediction and uncertainty of the proxy model so as to achieve efficient optimization of the objective function. The algorithm needs to define the proxy model first, that is, using the prior distribution to represent the proxy model of the objective function. This model does not contain any data in the initial stage, so it is based on prior assumptions; then, sampling and updating are performed, that is, selecting several initial sampling points, evaluating them through the true objective function, and using these sample points to update the posterior distribution of the proxy model. The proxy model estimates the value and uncertainty of the objective function; then, the next sampling point is selected; that is, the acquisition function is used to select the next point for sampling based on the current proxy model. The acquisition function determines the new sampling position by balancing exploration and development, thereby guiding the algorithm to find the global optimal solution; finally, the proxy model is optimized and the sampling point is updated, that is, the true value of the new sampling point is calculated, and the proxy model is updated again, and the above process is repeated until a certain termination condition is met.

In the Bayesian optimization algorithm used in this paper, the proxy model uses the Gaussian process, which can flexibly express uncertainty and is suitable for representing unexplored areas. The Gaussian process assumes that any function value obeys a multivariate normal distribution, and its mean and covariance are updated through the prior distribution and observed data. The Bayesian optimization algorithm is used to find a sliding window size that minimizes the mean square error of the training model, so the objective function is defined as:

$$f(x) = \min \left(\frac{1}{n} \sum_{i=1}^n \text{MSE} \left(y_{\text{test},i}, y_{\text{pred},i} \right) \right) \quad (9)$$

Among them, x is the corresponding sliding window size, $y_{\text{test},i}$ is the actual temperature of the i -th feature, $y_{\text{pred},i}$ is the predicted temperature of the i -th feature, n is the number of features, and MSE is the mean square error calculation function. For the objective function f , given several observation points $D = \{(x_i, y_i) | y_i = f(x_i) + \varepsilon\}$, its estimation model is $f(x) \sim GP(\mu(x), k(x, x'))$, where $\mu(x)$ is the mean function, $k(x, x')$ is the covariance function, i.e., the kernel function, as shown in Formula (10):

$$k(x, x') = \sigma^2 \exp \left(-\frac{(x - x')^2}{2l^2} \right) \quad (10)$$

where σ is the signal variance and l is the scale parameter. The acquisition function is used to determine the next evaluation point to optimize the objective function. The expected improvement used in this paper selects the sampling point by maximizing the expected improvement of the current optimal value. Given the current optimal value $f(x_{\text{best}})$ and the predicted distribution of the proxy mode, the expected improvement can be expressed as:

$$EI(x) = (f^* - \mu(x)) \Phi \left(\frac{f^* - \mu(x)}{\sigma(x)} \right) + \sigma(x) \phi \left(\frac{f^* - \mu(x)}{\sigma(x)} \right) \quad (11)$$

Among them, f^* is the optimal solution in the current observation, Φ is the cumulative distribution function of the standard normal distribution, and ϕ is the probability density function of the standard normal distribution. Through the Bayesian optimization algorithm, this paper can find close to the optimal parameters with fewer evaluation steps, and, through the probability model and acquisition function, it can strike a balance between exploration and utilization to avoid falling into the local optimal solution.

4. Transformer Early Warning Based on Predicted Temperature

In practice, the ultimate effectiveness of the transformer warning system is based on the warning rule base. The early warning rule base design in the transformer early warning method proposed in this paper is based on the dynamic threshold method of historical data. The dynamic threshold method, based on historical data, is a common method in transformer temperature early warning. It uses the historical temperature data of the equipment to set the early warning threshold so that the threshold can better reflect the actual operating status of the transformer. This method can dynamically adjust the threshold based on factors such as seasonality and load fluctuations to improve the accuracy and flexibility of early warning.

First, to establish an early warning rule base, a large amount of historical temperature data from transformers must be collected. Then, in view of the periodic fluctuation of transformer temperature data during the day and night, this paper divides the historical temperature data of transformers into different sets by hour, and the temperature data in the same set are the temperature at the same time in each day. After obtaining the processed

data, the early warning rule base is divided into three early warning levels. Among them, the first-level early warning is a minor early warning, which requires staff to continuously monitor the temperature change of the transformer online, and no further operation is required; the second-level early warning is a medium early warning, which requires staff to check the operating status of the transformer on site; the third-level early warning is a serious early warning, which requires the transformer to be shut down immediately for inspection, detailed fault analysis and troubleshooting conducted, and electrical connection and insulation status determined. The above three early warning levels correspond to three temperature change thresholds, which are represented by ΔT_1 , ΔT_2 , and ΔT_3 , respectively, and $\Delta T_1 < \Delta T_2 < \Delta T_3$. Since the temperature change in the transformer has a certain time periodicity, this method uses the mean of the processed dataset at that moment as the judgment criterion. If the difference between the predicted temperature and the mean is less than ΔT_1 , no warning operation is performed; if the difference between the predicted temperature and the mean is greater than or equal to ΔT_1 but less than ΔT_2 , the first-level warning operation will be performed; if the difference between the predicted temperature and the mean is greater than or equal to ΔT_2 but less than ΔT_3 , the second -level warning operation will be performed; if the difference between the predicted temperature and the mean is greater than or equal to ΔT_3 , the third-level warning operation will be performed. Through the analysis of historical data, we set dynamic thresholds for transformers in different time periods and load levels to build a warning rule library. In this way, the system can more flexibly adapt to the temperature changes of the equipment in different operating environments, ensuring that the early warning will not be frequently triggered due to small fluctuations, and can also alarm in time when there exists a real abnormality, improving the effectiveness and accuracy of the early warning system.

Finally, the process of the transformer temperature early warning model based on adaptive sliding window and stacking is summarized, as shown in Figure 2, which includes nine steps in total, and we can see the relationship between each algorithm module from them. Different colors represent different modules of the model: the green part represents the data acquisition module, the orange part represents the Bayesian optimization module, the blue part represents the stacking model module, and the cyan part represents the temperature warning module.

Step one collects the historical temperature data of the transformer, that is, it extracts the temperature data in continuous time from the transformer temperature sensor. Step two calculates the sliding window size. The sliding window is one of the important algorithms for time series prediction, and its size is a key parameter affecting the effect of the algorithm. Step three is used to train the support vector regression model based on the transformer temperature data collected in step one and the sliding window size obtained in step two. Step four is to train the random forest model based on the transformer temperature data collected in step one and the sliding window size obtained in step two. Step five is to train the gradient boosting regression model based on the transformer temperature data collected in step one and the sliding window size obtained in step two. It should be noted that steps three, four, and five can be performed simultaneously. Step six is to stack the models obtained in steps three, four, and five to obtain a stacking model and calculate whether the prediction mean square error of the current transformer temperature prediction model based on adaptive sliding window and stacking is the minimum value. If it is the minimum, it goes to step 7, otherwise it returns to step 2. Step 7 is to use a transformer temperature prediction model based on the final optimal performance to predict future temperatures. Step 8 is to build a warning rule base based on historical data and professional knowledge as the judgment standard for warning after transformer temperature prediction, and this step has no dependency relationship with step 7. Step 9 is to judge and warn based on the

transformer predicted temperature obtained in step 7 and the warning rule base obtained in step 8. Among them, the specific details of each step can be seen in Sections 2–4.

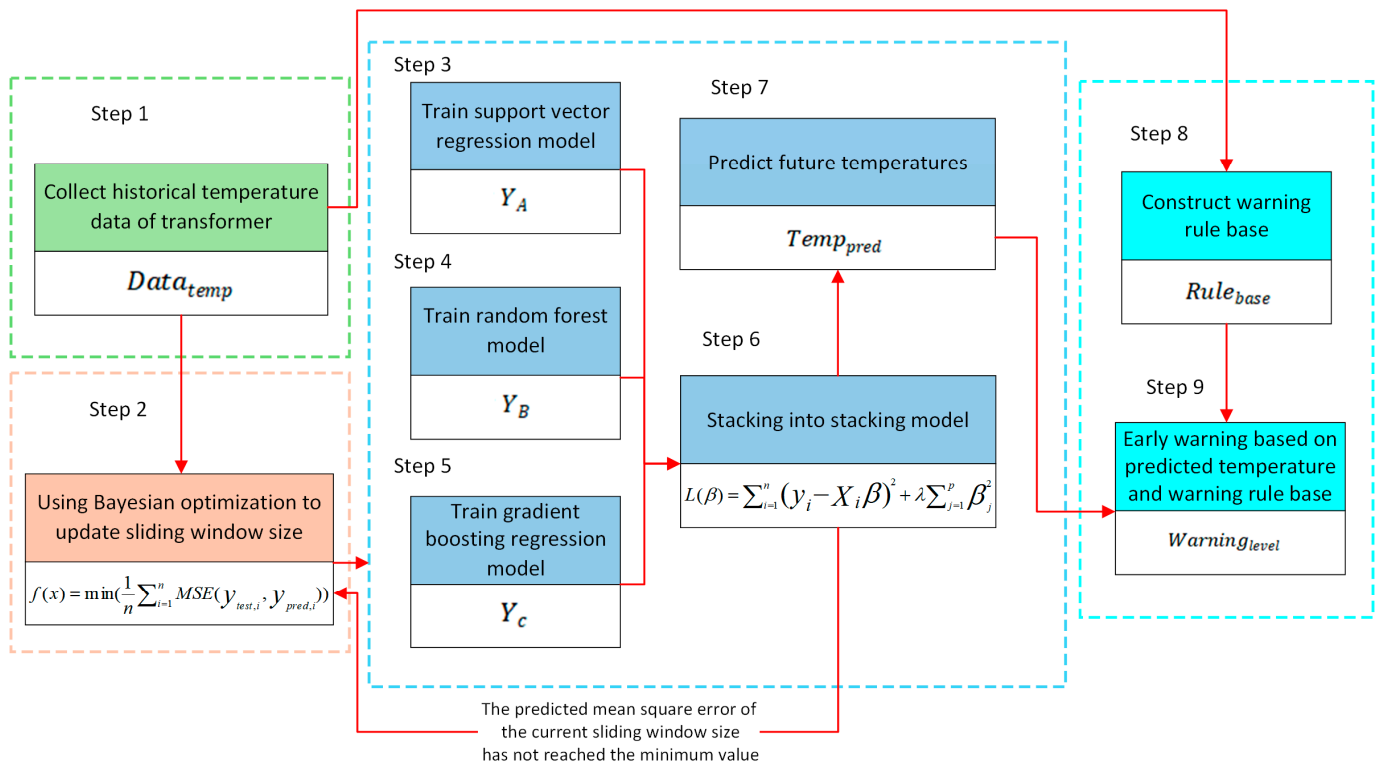


Figure 2. Flowchart of transformer temperature warning method based on adaptive sliding window and stacking.

5. Results

In the transformer temperature early warning system, the most critical factor affecting the early warning function is the temperature prediction of the transformer. If the prediction error is small, the early warning will play a greater role. Therefore, in order to demonstrate the prediction effect of the transformer temperature prediction method proposed in this paper, which is based on adaptive sliding window and stacking, this paper conducted an experiment based on the A, B, and C three-phase temperatures of the high-voltage side optical fiber of a substation main transformer for twenty consecutive days with a data interval of one hour. Specifically, this article divides the transformer temperature data for 20 consecutive days and 480 h into two parts, 360 h and 120 h, which are used for model training and prediction, respectively. A comparative analysis was conducted with the method described in reference [17]. The experimental results prove the advantages of this method and the role it can play in practical applications.

This paper uses the first 15 days of data in the dataset for model training and the last 5 days of data for testing the prediction effect. As shown in Figures 3 and 4, this paper predicts the three-phase temperatures of optical fibers A, B, and C on the high-voltage side of the main transformer. Figure 3 shows the results of the transformer temperature prediction model based on fixed sliding windows and stacking. It can be seen that the transformer temperature prediction model based on fixed sliding windows and stacking can accurately predict future temperatures, and the predicted temperature trend is basically consistent with the actual temperature trend, but there are still individual data points with large errors. Figure 4 shows the results of the transformer temperature prediction model based on adaptive sliding windows and stacking. Figure 5 shows the results of the transformer temperature prediction results of empirical mode decomposition bidirectional

long short-term memory. It can be seen that compared to the empirical mode decomposition bidirectional long short-term memory, the error reduction is more significant. It can also be seen that the error is reduced compared to the fixed sliding window and stacking methods. Therefore, it can be explained that the transformer temperature prediction method based on adaptive sliding windows and stacking improves the accuracy and robustness of transformer temperature prediction.

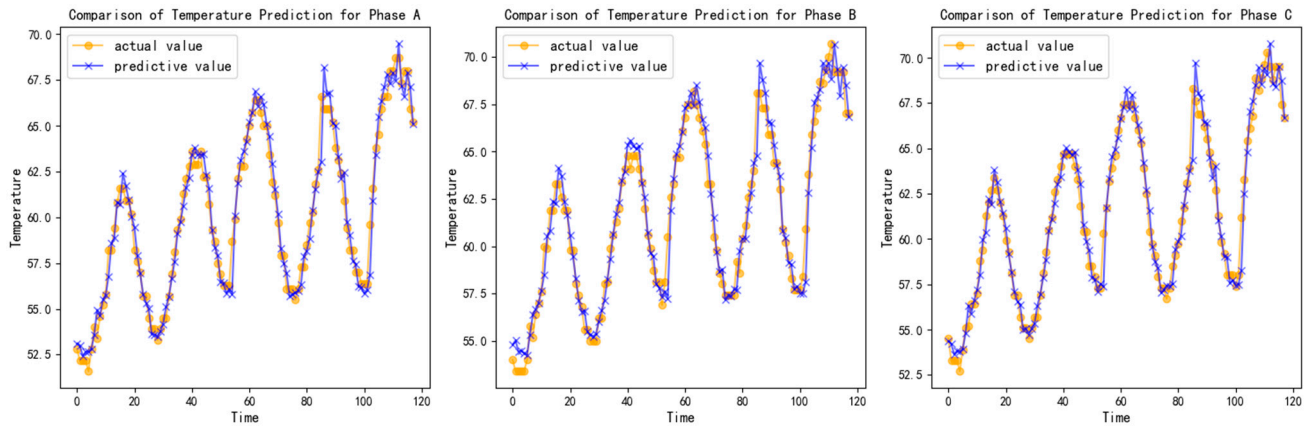


Figure 3. Transformer temperature prediction results of fixed sliding window and stacking.

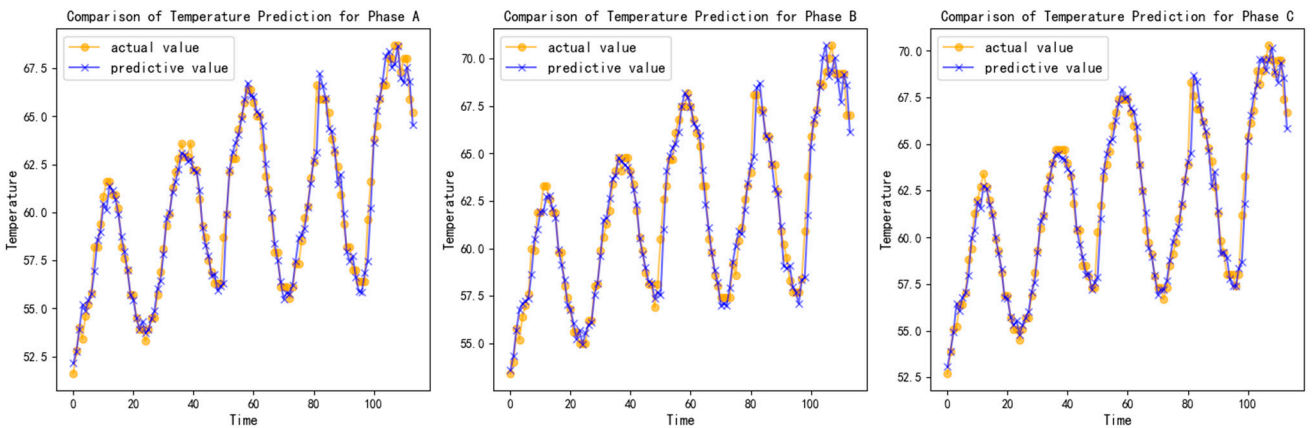


Figure 4. Transformer temperature prediction results of adaptive sliding window and stacking.

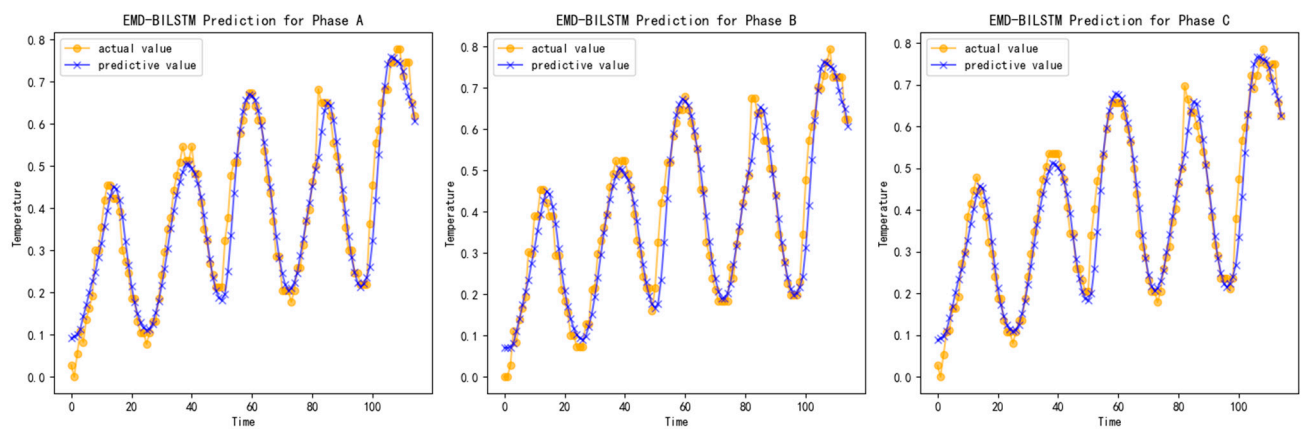


Figure 5. Transformer temperature prediction results of empirical mode decomposition bidirectional long short-term memory.

In addition, we also presented the difference curves between the predicted and actual temperatures of phases A, B, and C using the fixed sliding window stacking model and

the adaptive sliding window stacking model, as shown in Figures 6 and 7. Figure 6 shows the error curve of the fixed sliding window stacking model for predicting the temperatures of phases A and C, corresponding to the error curve in Figure 3. Figure 7 shows the error curve of the adaptive sliding window stacking model for predicting the temperatures of phases A and C, corresponding to the error curve in Figure 4. From the comparison between Figures 6 and 7, it can be seen that Figure 7 has a smaller overall amplitude and is more concentrated around the zero value on the vertical axis. Therefore, it can be seen from Figures 6 and 7 that the adaptive sliding window stacking model has a smaller overall error and more stable prediction performance compared to the fixed sliding window stacking model. However, the difference in single point prediction error between these two algorithm models is relatively small, and only a rough distribution can be seen from the graph, making it impossible to conduct detailed numerical analysis. Therefore, this article will provide a detailed comparison of error values in Table 1, which will be analyzed in detail later.

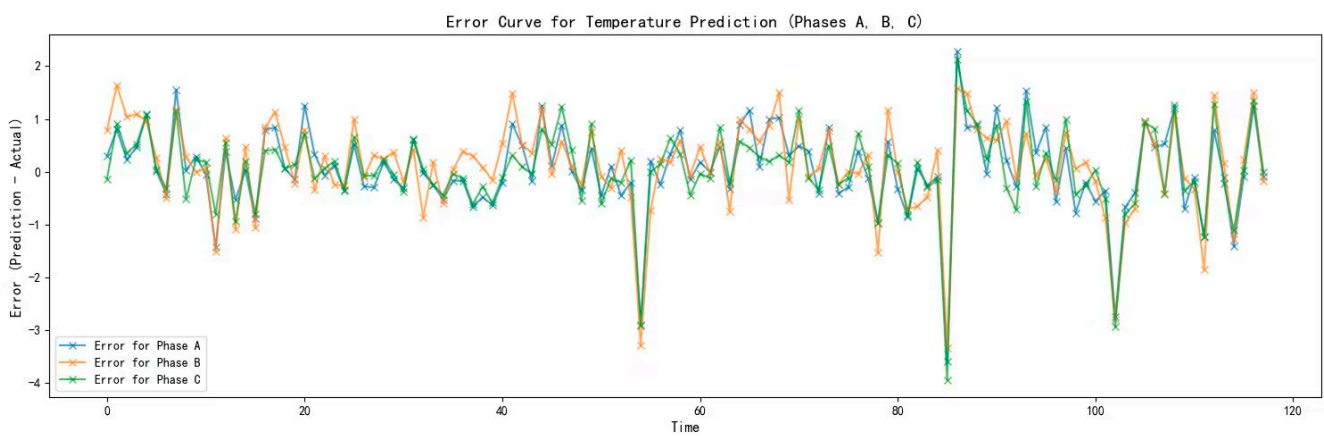


Figure 6. Fixed sliding window stacking model transformer temperature prediction error curve graph.

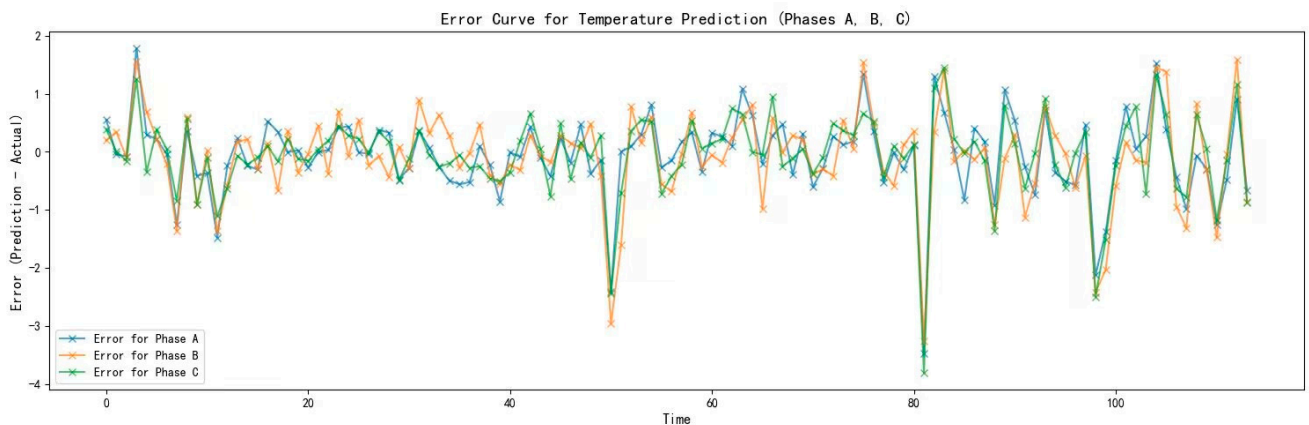


Figure 7. Adaptive sliding window stacking model transformer temperature prediction error curve graph.

Table 1. Prediction results evaluation table.

Evaluation Parameters	Transformer Temperature Prediction with Fixed Sliding Window and Stacking		Transformer Temperature Prediction Method with Adaptive Sliding Window and Stacking		Transformer Temperature Prediction Results of Empirical Mode Decomposition Bidirectional Long Short-Term Memory	
	Mean Square Error	Total Error	Mean Square Error	Total Error	Mean Square Error	Total Error
A-phase Temperature	0.6686	67.9764	0.5060	55.2039	1.0415	84.3839
B-phase Temperature	0.7773	76.4719	0.6489	63.4301	1.1952	87.1573
C-phase Temperature	0.6423	63.8835	0.5266	55.0079	1.0124	80.3528

Since the prediction performance of the above three methods is good, it is difficult to evaluate their advantages and disadvantages by comparing them only using the figure. This paper provides Table 1 to more intuitively reflect the advantages of the transformer temperature prediction method based on the adaptive sliding window and stacking. From Table 1, it can be seen that the mean square error of the prediction results of the above three methods is small. However, compared to the empirical mode decomposition bidirectional long short-term memory, the adaptive sliding window and Stacking transformer temperature prediction method reduces the total error of A-phase temperature prediction by about 34.58%. The total error of B-phase temperature prediction decreased by about 27.22%. The total error of C-phase temperature prediction has been reduced by about 31.54%. In addition, the adaptive sliding window and Stacking transformer temperature prediction methods still reduce the mean square error of the prediction results to a certain extent on the basis of the fixed sliding window and Stacking transformer temperature prediction methods. It can be seen that the total error of A-phase temperature prediction has been reduced by about 18.79%. The total error of B-phase temperature prediction decreased by about 17.05%. The total error of C-phase temperature prediction has been reduced by about 13.89%. Therefore, it can be concluded that the transformer temperature prediction method based on adaptive sliding window and Stacking reduces the prediction error in A, B, and C three-phase temperature prediction compared to existing methods. The transformer temperature warning method and system based on adaptive sliding window and Stacking achieve effective transformer temperature warning by improving the accuracy of transformer temperature prediction.

Finally, to illustrate the importance of the adaptive sliding window algorithm in the transformer temperature warning model proposed in this paper, we took the window range from 5 to 30 as an example and plotted a line graph of the total temperature prediction error as a function of the sliding window size, as shown in Figure 8. It can be seen that different window sizes have different effects on the prediction of different phase temperatures, which makes it difficult to capture patterns. This depends on the various influences of time on the original transformer temperature data. However, the adaptive sliding window algorithm can ultimately calculate the optimal sliding window size, which is the window size that minimizes the sum of the mean square errors of the ABC three-phase temperature prediction. In this example, the optimal sliding window size is 25.

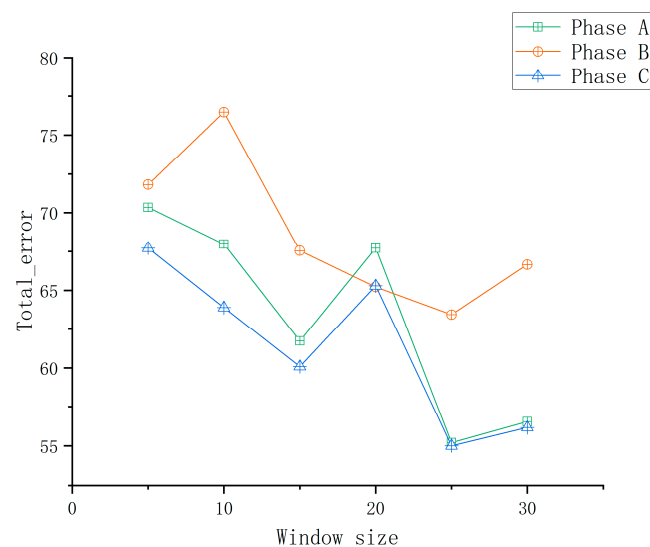


Figure 8. Line graph of total temperature prediction error of phases A, B, and C as a function of sliding window size.

6. Conclusions

This paper proposes a transformer temperature prediction method based on adaptive sliding window and stacking ensemble learning. It combines base learners such as support vector regression, random forest, and gradient boosting regression, and uses ridge regression as a meta-learner to achieve high-precision temperature prediction of multi-model fusion. Under the nonlinear and time-varying characteristics of temperature data, this method dynamically adjusts the size of the sliding window through Bayesian optimization to adaptively capture the temperature change trend, effectively improving the accuracy and robustness of the prediction. Experimental results show that compared with the traditional fixed sliding window method, the adaptive sliding window and stacking methods significantly reduce the temperature prediction error, reducing the total error by 18.79%, 17.05%, and 13.89% on the three-phase temperature of A, B, and C, respectively. This improvement significantly enhances the early warning system's ability to identify transformer temperature anomalies, reduces the possibility of false alarms and missed alarms, and provides more reliable technical support for transformer temperature management and safety assurance.

In practical applications, this method is suitable for long-term transformer temperature monitoring and fault warning, especially for dealing with temperature fluctuations under different ambient temperatures and load conditions. By dynamically adjusting the sliding window size, temperature anomalies can be detected in a high temperature environment or during peak power consumption, helping maintenance personnel take preventive measures, thereby improving the overall stability of the substation and the reliability of the power system.

Future research can be expanded in the following aspects: First, it is possible to consider integrating more deep learning algorithms to better handle the long-term and short-term dependencies in temperature data, thereby further improving the prediction accuracy of the model. Second, by combining multiple sensor data such as current, voltage, and humidity, multi-source data fusion can be achieved to improve the adaptability of the temperature prediction model. In addition, with the development of the Internet of Things and edge computing technologies, the real-time data processing of the temperature early warning system can be transferred to the edge to reduce transmission delays and improve the real-time response capability of the system. Finally, by introducing big data analysis and automated fault diagnosis mechanisms on the cloud platform, intelligent operation and maintenance management of transformers can be realized, which will help promote the digital and intelligent transformation of the entire power system.

Author Contributions: P.Z., Q.Z. and H.H. (Huan Hu) designed the project and drafted the manuscript, as well as collected the data. H.H. (Huazhi Hu), R.P. and J.L. wrote the code and performed the analysis. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Project of State Grid Hubei Electric Power Co., Ltd. Xiaogan Power Supply Company (Research on proactive warning method of abnormal operation of substation equipment based on deep learning). The contract number is SGHBXG00XJS2310925.

Data Availability Statement: The data presented in this study are available in this article.

Conflicts of Interest: Authors Pan Zhang, Qian Zhang, Huan Hu and Huazhi Hu were employed by State Grid Hubei Electric Power Co., Ltd., Xiaogan Power Supply Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Ahmad, I.; Singh, Y.; Ahamad, J. Machine learning based transformer health monitoring using IoT Edge computing. In Proceedings of the IEEE 5th International Conference on Computing, Communication and Security, (ICCCS), Patna, India, 14–16 October 2020.
2. Li, Y. Study on the Precision Optimization Method of Gas Concentration Monitoring System in Transformer Oil by Photoacoustic Spectroscopy. Master's Thesis, Hubei University of Technology, Wuhan, China, 2020.
3. Luo, Y.; Wang, L.; Jiang, J.; Li, D.; Lai, S.; Liu, J.; La, Y.; Sun, W.; Shi, M. Top oil temperature prediction at a multiple time scale for power transformers based on adaptive extended Kalman filter. *Front. Energy Res.* **2024**, *12*, 1440338. [[CrossRef](#)]
4. Abbasi, A.R.; Mahmoudi, M.R.; Arefi, M.M. Transformer winding faults detection based on time series analysis. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–10. [[CrossRef](#)]
5. Li, C.; Chen, J.; Xue, C.; Liu, Z.; Davari, P. Simultaneous Multispot Temperature Prediction of Traction Transformer in Urban Rail Transit Using Long Short-Term Memory Networks. *IEEE Trans. Transp. Electr.* **2023**, *9*, 4552–4561. [[CrossRef](#)]
6. Shi, M.; Xu, H.L.; Li, H.; Luo, Y.T.; Jiang, J.F. Classification Early Warning Model of Oil-immersed Distribution Transformers Based on Multi-modal Data Fusion. *Guangdong Power Grid* **2022**, *35*, 110–117.
7. Huy, P.C.; Minh, N.Q.; Tien, N.D.; Anh, T.T.Q. Short-term electricity load forecasting based on temporal fusion transformer model. *IEEE Access* **2022**, *10*, 106296–106304. [[CrossRef](#)]
8. Li, K.; Liu, D.; Li, Y. Research on power transformer test plan based on big data processing algorithm. *J. Phys. Conf. Ser.* **2020**, *1648*, 022130. [[CrossRef](#)]
9. Xu, S.; Shang, Y.; Li, Z.; Lu, Y.; Liu, M.; Liu, W.; Wang, Z.; Tang, W. Transformer Monitoring with Electromagnetic Energy Transmission and Wireless Sensing. *Sensors* **2024**, *24*, 1606. [[CrossRef](#)] [[PubMed](#)]
10. Lei, L.; Liao, F.; Liu, B.; Tang, P.; Zheng, Z.; Qi, Z.; Zhang, Z. Research on Transformer Temperature Rise Prediction and Fault Warning Based on Attention-GRU. In Proceedings of the IEEE 5th International Conference on Smart Power & Internet Energy Systems, (SPIES), Shenyang, China, 1–4 December 2023.
11. Saleh, N.; Azis, N.; Jasni, J.; Kadir, M.Z.A.; Talib, M.A. Prediction of a transformer's loading and ambient temperature based on SARIMA approach for hot-spot temperature and loss-of-life analyses. In Proceedings of the IEEE International Conference on the Properties and Applications of Dielectric Materials, (ICPADM), Johor Bahru, Malaysia, 12–14 July 2021.
12. Taksana, R.; Janjamraj, N.; Romphochai, S.; Bhummkittipich, K.; Mithulananthan, N. Design of Power Transformer Fault Detection of SCADA Alarm Using Fault Tree Analysis, Smooth Holtz–Winters, and L-BFGS for Smart Utility Control Centers. *IEEE Access* **2024**, *12*, 116302–116324. [[CrossRef](#)]
13. Tang, H.; Guo, T.; Li, S.; Huang, F.; Lang, X.; Li, D. Research on Transformer Temperature Warning Algorithm Based on Support Vector Machine. In Proceedings of the IEEE 2023 International Conference on Internet of Things, Robotics and Distributed Computing, (ICIRDC), Rio De Janeiro, Brazil, 29–31 December 2023.
14. Liu, H.; Wang, Y.; Sun, J.; Zhang, Y.; Du, Y.; Zhang, B. Research on Temperature Prediction and Fault Early Warning of Transformer Bushings. In Proceedings of the IEEE 7th International Electrical and Energy Conference, (CIEEC), Harbin, China, 10–12 May 2024.
15. Yang, Z.J.; SiMa, W.X.; Yang, M.; Li, W.H.; Yuan, T.; Sun, P.T. A fast prediction method of hot spot temperature of oil-immersed transformer based on flow-thermal field coupling simulation and EEMD-LSTM network. *High Volt. Technol.* **2024**, 1–14. [[CrossRef](#)]
16. Feng, J.; Feng, Z.; Jiang, G.; Zhang, G.; Jin, W.; Zhu, H. A Prediction Method for the Average Winding Temperature of a Transformer Based on the Fully Connected Neural Network. *Appl. Sci.* **2024**, *14*, 6841. [[CrossRef](#)]
17. Chen, H.; Meng, L.; Xi, Y.; Xu, C.; Wang, Y.; Xin, M.; Chen, G.; Chen, Y. Power transformer oil temperature prediction based on empirical mode decomposition-bidirectional long short-term memory. *J. Algorithms Comput. Technol.* **2023**, *17*, 17483026231176196. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.