processes

MDPI

Article

A Transformer Oil Temperature Prediction Method Based on Data-Driven and Multi-Model Fusion

Lin Yang [1], Liang Chen [2], Fan Zhang [3,*], Shen Ma [2], Yang Zhang [2] and Sixu Yang [2]

[1] Sichuan Electric Power Research Institute, Chengdu 610041, China; yangcome@163.com
[2] State Grid Meishan Power Supply Company, Meishan 620010, China; liangchen777@hotmail.com (L.C.); shen_ma1248@163.com (S.M.); yang_zhang7197@163.com (Y.Z.); sixu_yang5334@163.com (S.Y.)
[3] School of Electronic Information, Sichuan University, Chengdu 610065, China
* Correspondence: fan.zhang@scu.edu.cn

Abstract: A power transformer is an important part of the power system, and the oil temperature of the transformer is an important state parameter that reflects the operation state of the transformer. The accurate prediction of the oil temperature of the transformer can ensure the safe and stable operation of the transformer. Given the lack of a practical and effective data processing process and the problem that most of the current research is conducted on small-scale ideal datasets, this paper proposes a transformer oil temperature prediction method based on data-driven and multi-model fusion. The method first analyses and processes the actual transformer inspection data; it then uses the multi-model fusion method to model and predict the transformer oil temperature. The base model was trained using the machine learning method, and the secondary learning model was trained using the improved TSSA-BP neural network. The improved sparrow search algorithm (TSSA) was used to optimise the parameters of the BP neural network to improve the convergence accuracy and prediction performance of the model. The transformer data are classified according to cooling mode, operating voltage, and other indicators, and then eight groups of experimental datasets under different actual conditions are constructed for modelling and prediction. The experimental results show that the maximum root mean square error and the mean absolute percentage error of this method on different datasets are 1.0877 and 1.58%, and compared with other prediction methods, the prediction accuracy of this method is better than other methods, which verifies the practicability and feasibility of modelling and predicting for the actual transformer inspection data.

Keywords: transformer; oil temperature; temperature prediction; data processing; improved SSA

check for updates

Academic Editor: Jean-Louis Dirion

Received: 21 November 2024
Revised: 1 January 2025
Accepted: 20 January 2025
Published: 22 January 2025

Citation: Yang, L.; Chen, L.; Zhang, F.; Ma, S.; Zhang, Y.; Yang, S. A Transformer Oil Temperature Prediction Method Based on Data-Driven and Multi-Model Fusion. *Processes* **2025**, *13*, 302. https://doi.org/10.3390/pr13020302

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

The transformer is one of the important equipment of the power system, and its operation status is related to the safety and stability of the whole power system. Transformer failure will cause great harm to the power system. The internal failure and life reduction in the transformer is usually caused by the degradation of insulation capability, and the internal overheating of the transformer is the main reason for the degradation of its insulation capability, so it is important to obtain an accurate internal temperature of the transformer [1]. Hot spot temperature and oil temperature are important parameters of the internal temperature of the oil-immersed transformer; hot spot temperature is the temperature of the hottest spot inside the transformer, and oil temperature is the temperature of the transformer-insulating oil, both of which are important indexes for

assessing the operational status of the transformer. Oil temperature is usually measured by temperature sensors. There are two ways to obtain the hot spot temperature: (1) directly measured using fibre optic sensors or (2) calculated according to the hot spot temperature analysis calculation method; however, fibre optic sensors have a high cost of temperature measurement and are difficult to maintain. According to the load guidelines and numerical calculation methods, to calculate the hot spot temperature, the environment, time, and other factors cannot be taken into account and therefore cannot accurately calculate the hot spot temperature [2].

Elevated oil temperature is usually caused by transformer overload operation, cooling system abnormalities, equipment ageing, and other reasons. Transformer equipment can usually be overloaded in a short period of time, and this behaviour will lead to oil temperature rise. However, the prolonged overload operation of the transformer may shorten the service life of the transformer or lead to transformer damage. An abnormal transformer cooling system will lead to the abnormal heat dissipation of the transformer, which will cause the oil temperature to rise. Ageing of the equipment will increase the magnetic flux loss of the transformer and reduce the insulation capacity, resulting in an increase in oil temperature. Rising oil temperature will have an effect on the insulation capacity of the transformer. Rising oil temperature accelerates the ageing of insulating materials, resulting in a decrease in their insulation capacity. High temperature will accelerate the oxidation process of the insulating oil, and the oxidation products will reduce the insulating ability of the transformer. As the temperature continues to rise, the dielectric loss of the insulating oil increases, and even the breakdown of the insulating oil may occur, seriously affecting the stable operation of the transformer.

The current mainstream research is to establish a prediction model for the hot spot temperature of the transformer, but considering that it is difficult to obtain accurate hot spot temperature data on the transformer and the oil temperature data have the same reference value, this paper models and predicts the oil temperature data using artificial intelligence algorithms based on historical data to extract the internal potential characteristics and establish a prediction model, which provides a predictive model for the dynamic capacity increase in the transformer in the context of big data mining, analysis, and so on. Dynamic capacity increases work to provide predictive and efficient load reduction suggestions and countermeasures [3]. Through the accurate prediction of transformer oil temperature, we can effectively monitor and adjust the transformer's operating status, as well as reduce the waste of energy due to overheating or overcooling, thus improving the efficiency of energy utilisation [4].

In previous studies on transformer temperature prediction, deep learning methods were first applied to transformer hot spot temperature prediction by Daponte P et al. However, the inputs of this network model only considered the environmental factors and load currents; after that, Pradhan MK et al. proposed an improved artificial neural network model by taking the oil temperature parameter as the input to the model based on Daponte P [5]; Li Mengli et al. [6] increased the input feature dimension to five dimensions, which made the prediction results more accurate; Li Shuqing et al. [7] applied the grey neural network model to transformer temperature prediction, which enabled them to improve the prediction accuracy under the limited amount of data; Jiang Bing et al. [8] used the BP neural network for the prediction of hot spot temperatures of transformer windings, but local convergence problems may occur, so other researchers followed up to establish various optimisation algorithms for the BP neural network to predict the hot spot temperatures of transformer windings. Researchers also established various optimisation algorithms to optimise the hyperparameters of the BP neural network to avoid this problem;

for example, the literature [9] optimised the BP neural network to achieve the prediction of transformer oil temperature by improving the PSO algorithm [9].

The support vector machine (SVM) method was initially used to solve classification problems, but considering the advantages of SVMs for dealing with nonlinear, high-dimensional data, researchers have applied it to transformer temperature prediction. However, support vector machines are sensitive to parameters, and their sensitivity and the improper selection of parameters can significantly affect the results, so the use of this method requires the optimal solution of its parameters [10]. Chen Weigen et al. [11] used a genetic algorithm to optimise the parameters of the support vector machine, and the prediction accuracy was better than the results using the default parameters. Yu Xi et al. [12] used the particle swarm algorithm to optimise the parameters of SVM and compared the results with the genetic algorithm to conclude that particle swarm algorithm optimisation is better. Liu Gang et al. [13] overcame the shortcomings of traditional PSO, whereby it is easy to fall into local minima, and improved the convergence speed of the algorithm by introducing a contraction factor to the particle update speed based on PSO of LSSVM. Liang Feng et al. [14] used the ant colony optimisation (ACO) algorithm to optimise the SVM parameters and compare it with the PSO algorithm; the transformer hot spot temperature prediction results are better than the PSO-SVM model, but the selection rules for input features are vague. Jingke Liu et al. [15] established the PCA-IHHO-LSSVM model to predict the transformer's temperature. First, the model input features are determined using PCA, and then the parameters of LSSVM are optimised using the Harris Hawk algorithm, which improves the normality of feature selection, and the model has a better prediction accuracy. However, the above researchers using the SVM are working on small-scale or ideal experimental datasets and are unable to determine its prediction effect on real large-scale datasets.

The experimental dataset of the above method is mostly a small-scale dataset with a data size of about several hundred groups, while the dataset sizes in actual engineering are relatively large, and it is not possible to determine the prediction effect of the above method in actual large-scale datasets. In addition, most of the previous studies used ideal datasets and ignored the process of data processing, and the actual transformer data are complex and cannot be directly applied to the model; they first need to be analysed and processed. Therefore, this paper proposes a transformer oil temperature prediction method based on data-driven and multi-model fusion. The method firstly analyses and processes the real transformer operation and inspection data, eliminates duplicate data, fills in the missing data, and screens the abnormal data, and then the abnormal data detection process is completed through the theoretical detection method of outliers combined with the actual operation and inspection experience. The model construction is divided into two parts; the first part is the base model, which is constructed using five different machine learning methods to enable feature extraction for different feature dimensions of the dataset. The second part is the improved SSA-BP neural network model, where the oil temperature prediction results derived from the base model are used as inputs to the neural network model; the hyperparameters of the BP neural network are optimised using the improved sparrow search algorithm (TSSA), and the BP neural network performs a quadratic prediction of the initial prediction values to produce the final prediction results. The purpose of establishing the oil temperature prediction model is to establish a high-precision oil temperature prediction model through the normal operation data of the transformer, after which the real-time monitoring data are input into the model, and the model prediction value is compared with the actual value to determine whether the current oil temperature data are abnormal or not. In this paper, the method is used to predict transformer oil temperature for several different datasets, and the results show good prediction results for all of them.

The Section 1 of this paper writes about data analysis and data processing, where the original dataset is subjected to data transformation and classification, model input feature selection, and data cleaning. The Section 2 of this paper carries out the algorithmic introduction of the model, which introduces the principle of the base model algorithm and the application of the TSSA-BP method, respectively. The Section 3 is the experimental design and conclusion analysis, which describe the presentation and analysis of the experimental results. The Section 4 is the concluding remarks, which summarise and look forward to the research methodology of this paper. Considering the large number of acronyms in this paper, a list of acronyms is provided in Table 1.

Table 1. Abbreviation list.

Abbreviation	Full Form
BP	Back Propagation
CART	Classification and Regression Tree
KNN	K-Nearest Neighbours
MAPE	Mean Absolute Percentage Error
ONAN	Oil Natural Air Natural
ONAF	Oil Natural Air Forced
ODAF	Forced Directed Oil Circulation Air Cooling
RMSE	Root Mean Square Error
SSA	Sparrow Search Algorithm
SVM	Support Vector Machine

2. Data Analysis and Data Processing

The data analysis and data processing parts are to perform data analysis and data processing for the original dataset. Data analysis mainly completes the analysis and presentation of data size and data types, and data processing mainly completes data classification, feature selection, and data cleaning.

2.1. Data Analysis

In this paper, the real-time monitoring data of 34 substations in a region for ten months from July 2023 to May 2024 are obtained. The original format of the data is all the transformer monitoring data in one SQL file for each month, and the size of the SQL file for each month is about 4 GB; the data size is about 8 million items, and the overall data size is about 46 GB, which needs to be converted into CSV files. The following shows the raw data size of the selected 10 months, respectively, and the data size is shown in Table 2.

From Table 1, it can be seen that the amount of data per month is around 4 GB, and the data size is around 8 million groups, and the different data sizes in different months are usually due to substation maintenance or the loss of the data transmission process.

The data categories in the obtained dataset are shown in Table 3.

As can be seen from Tables 1 and 2, the acquired data are characterised by large data volume, multiple data types, and data covering the actual operation of multiple transformers at different times of the day and in different seasons, and because it is impossible to adequately capture and explain the complex relationships in the data at the traditional theory-driven level, this paper establishes a data-driven transformer oil temperature prediction model based on the data. The data-driven prediction model can

capture the load conditions of transformer temperature rise changes and improve the prediction accuracy and generalisation ability of the model.

Table 2. Raw data size.

Dataset	SQL File Size	CSV File Size
23.07	4.26 GB	7,896,147 (sets)
23.08	4.21 GB	7,867,448 (sets)
23.09	4.43 GB	8,278,574 (sets)
23.10	4.29 GB	8,016,948 (sets)
23.11	3.88 GB	7,250,760 (sets)
23.12	4.32 GB	8,073,011 (sets)
24.01	4.27 GB	7,979,573 (sets)
24.02	3.80 GB	7,101,259 (sets)
24.03	4.38 GB	8,185,135 (sets)
24.04	4.41 GB	8,241,197 (sets)
24.05	4.29 GB	8,020,014 (sets)

Table 3. Types of data.

Serial Number	Data Category	Unit
1	oil temperature	°C
2	load current	A
3	Phase A,B,C current	A
4	active power	W
5	reactive power	Var
6	load factor	
7	Maximum load of the day	A
8	Average load ratio	
9	environmental temperature	°C

A part of the real data is shown in Table 4.

Table 4. Part of the real data.

Device ID	Load Current (A)	Phase A Current (A)	Phase B Current (A)	Environmental Temperature (°C)	Active Power (W)	Reactive Power (Var)	Load Factor	Oil Temperature (°C)
36	92.1856	229.613	231.370	34.1404	91.6285	10.1126	61.4567	50.0961
36	77.5621	191.842	191.348	29.3889	77.5628	2.9871	0	50.6245
36	92.8863	2320.492	232.776	29.6454	92.5663	7.7026	61.9241	52.1176
36	89.5527	221.532	223.816	34.8114	89.5522	0	59.7014	51.0629
36	998.712	245.601	248.060	32.9141	98.5945	4.8225	65.8082	52.3812
36	98.2766	243.4926	246.6548	31.8886	98.1257	5.4253	62.5170	52.9086
36	97.9914	242.789	245.952	30.5331	97.9917	4.3563	65.3278	54.2269

The table shows a part of the real data situation, and there are more feature categories in the real data; from the table, we found that there are anomalous data and missing values. For example, in the current load of the 5th data for 998.712, the A phase current of the 3rd data for 2320.492 is anomalous data, and the reactive power and the load factor appeared as 0 in the data; these anomalous data points are taken directly as the input of the model, which will affect the model's robustness and prediction ability, so the real data need to be processed.

2.2. Data Processing

Because of the large size of the data, the complexity of the data situation, and the possibility of anomalies in the dataset, data processing of the original dataset is required. The data processing process is shown in Figure 1.

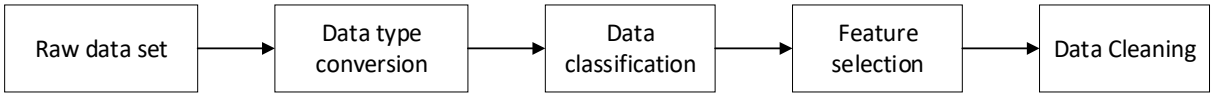


Figure 1. Data processing procedure.

After the dataset type conversion, the scale of the data in a single file reaches millions of articles, which cannot be directly cleaned; taking into account the actual operation and inspection data of different categories of transformers of the change rule and potential characteristics of the relationship between the existence of a large number of differences, the category as a model input feature for prediction is not good, so the transformer data are categorised. Different models of transformers, cooling methods, rated capacity, operating voltage, and other categories of information can be used as a basis for classification, but this study found that there is a large difference in the temperature rise in transformers with different cooling methods and operating voltages, so the transformers are classified according to the cooling method and operating voltage [16]. The data are obtained for two types of operating voltages, 110 kV and 220 kV, and three types of cooling methods, oil-immersed self-cooled (ONAN), oil-immersed air-cooled (ONAF), and forced-oil circulating air-cooled (ODAF), and the model is trained using the historical data of a specific category when it is necessary to predict the oil temperature of a specific category of transformer.

Feature selection is carried out after data classification because there are many features in the raw transformer data, but not all of them are the main correlation features of oil temperature, so grey correlation analysis is used to select the main features. The grey correlation analysis method is used to measure the degree of association between factors in a system; the grey correlation coefficient is calculated to determine the similarity between the reference and comparison series, and then the grey correlation coefficient is averaged to obtain the degree of association, through which the degree of influence between different factors is measured [17]. After performing grey correlation analysis, it is determined that the model input features are ambient temperature, load factor, load current, and active power, and their correlation values are calculated and shown in Table 5.

Table 5. Input feature correlation calculation.

	Environmental Temperature	**Load Factor**	**Load Current**	**Active Power**
Correlation	0.6595	0.7534	0.6473	0.6547

Data cleaning is performed after transformer classification is completed, and data cleaning is mainly completed with duplicate value detection, missing value filling, and

outlier screening [18]. Duplicate value detection is a key step in data cleaning, detecting duplicate values in the dataset and deleting them to ensure the specification of the dataset. Missing value replenishment first requires missing value detection, and the detected missing values are filled by mean values. Outliers are detected through a combination of theoretical methods and practical O&M rules. The theoretical approach uses the 3σ principle, which is used to identify outliers and variability in data based on the properties of the normal distribution, which is a continuous probability distribution where the shape and location of the distribution are determined via the mean and standard deviation [19]. In a dataset that obeys a normal distribution, the standard data will lie within three standard deviations of the mean, and data points that lie outside $\pm 3\sigma$ of the mean are identified as outliers. The core of this lies in the construction of the probability density function, which is

$$f(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

where μ is the mean, and σ is the standard deviation. Outliers are detected in the dataset using the 3σ principle, and the detected outliers are not actual operational outliers but usually represent theoretical outliers, and its removal facilitates the feature extraction of the model.

The actual operation and inspection rules are determined according to the operation and maintenance experience and the operation law of the transformer; for example, under the normal operation state of the transformer, the data when the oil temperature exceeds the threshold value of 90 degrees are considered as an abnormal value; the data when the continuous jump of the oil temperature data exceeds 1 degree are considered as an abnormal value; the data when the continuous jump of the load ratio exceeds 100% are considered as an abnormal data, and so on. The theoretical method is based on the mathematical analysis method to detect the theoretical abnormal values, while the actual operation and inspection rules are used to screen the real abnormal values, and the combination of the two can lead to a more accurate detection of transformer data abnormal values, thereby improving the normality and accuracy of the dataset.

The experimental data for cleaning are the August dataset of two transformers with an operating voltage of 110 kV cooling mode ONAN and ONAF, respectively, and the December dataset of three transformers with an operating voltage of 220 kV cooling mode ONAN, ONAF, and ODAF, respectively, and the cleaning results are shown in Table 6.

Table 6. Data cleaning results.

Transformer Serial Number	Raw Data	Abnormal Data	Cleaning Data
1	78,456	2104	76,352
2	76,561	2147	74,414
3	78,456	1288	77,168
4	74,421	560	73,861
5	78,456	484	77,972

The histogram comparing the data before and after cleaning is shown in Figure 2.

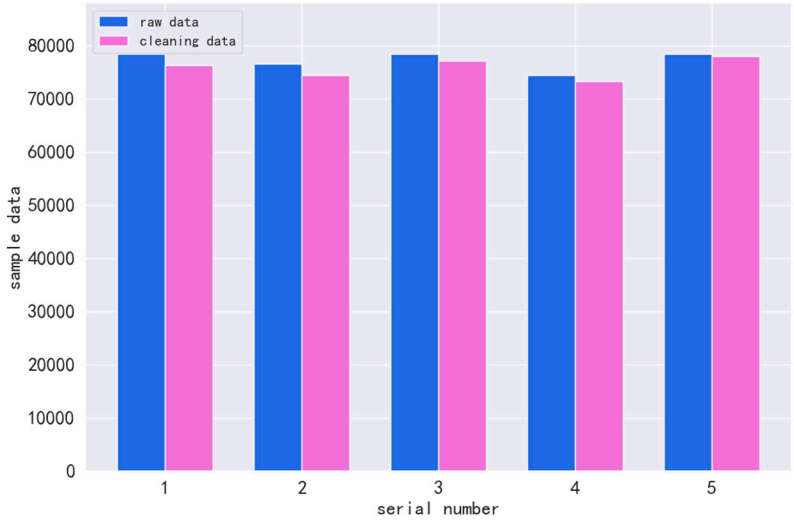


Figure 2. Histogram of data cleaning results.

3. Introduction of Model Algorithm

The data-driven multi-model fusion prediction method achieves an accurate prediction of oil temperature by integrating the base model and the secondary prediction model, and the overall process of model construction is shown in Figure 3.

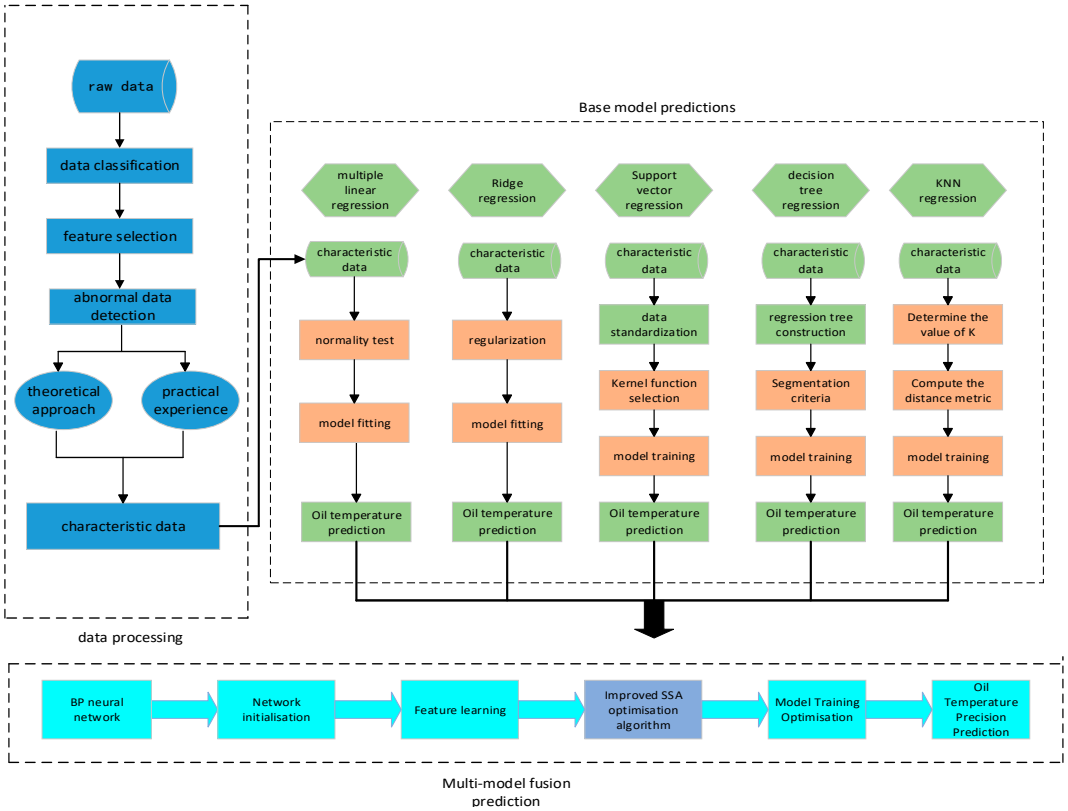


Figure 3. The overall process of model building.

The process is as follows: the feature dataset is obtained through data processing; the feature data are input into the five base models for the initial prediction of oil temperature; the resulting predicted value of oil temperature is used as the input to the BP neural network; the improved SSA optimisation algorithm is used for parameter searching of the BP neural

network, and the BP neural network is trained according to the optimal combination of the parameters, and the final prediction of the oil temperature is finally realised.

3.1. Introduction to Base Modelling Algorithms

The base model prediction method is used to make the initial prediction of transformer oil temperature using five machine learning methods; the machine learning methods use multiple linear regression, ridge regression, support vector regression, decision tree regression, and KNN regression methods, and the algorithm principles are described below.

(1) Multiple linear regression: Multiple linear regression is a statistical method used to analyse the linear relationship between multiple independent variables and a single dependent variable [20]. In multiple linear regression, the prediction accuracy is made best by finding a set of best-fit coefficients, which usually performs better when there is a good linear relationship between the inputs and outputs, which is given by the formula:

$$f(x) = l_1x_1 + l_2x_2 + l_3x_3 + \cdots + l_nx_n + b \tag{2}$$

where $f(x)$ represents the dependent variable; l represents the weight coefficients trained by the model; x represents the independent variable, and b is the bias. The fitting process for multiple linear regression is usually performed by determining the coefficients using the least squares method;

(2) Ridge regression: Ridge regression is a linear regression method for dealing with covariate data. In multivariate linear regression, ordinary least squares estimation produces unstable results when there are multiple covariates in the predictor variables, and Ridge regression introduces a regularisation term to solve this problem [21]. The objective function of linear regression is

$$J(\beta) = \sum(y - X\beta)^2 \tag{3}$$

To ensure that the regression coefficients β are summable, Ridge regression adds the following penalty term of L2 paradigm to the objective function:

$$J(\beta) = \sum(y - X\beta)^2 + \sum\lambda\beta^2 \tag{4}$$

where λ is a non-negative number, and the larger λ is, the smaller the regression coefficient β is in order to minimise $J(\beta)$. The inclusion of the L2 paradigm penalty term makes the matrix full rank and ensures invertibility. However, the inclusion of the penalty term makes the estimation of the regression coefficient β no longer unbiased. Regularisation improves model stability and reduces the risk of overfitting;

(3) Support vector regression: The core of the support vector machine method is to map the data from a low-dimensional space to a high-dimensional space through the kernel function in order to find a hyperplane in the high-dimensional space so that the error between the training samples and the hyperplane is minimised, and finally, it completes the regression prediction through the kernel function to weight the support vectors by averaging them [22]. Considering that the radial basis kernel function has a better generalisation performance, this paper chooses the radial basis kernel function, which is mathematically expressed as

$$K(\mu, v) = \exp\left(\frac{-||\mu - v||^2}{2\sigma^2}\right) \tag{5}$$

where μ and v are sample data, and σ is the kernel function.

SVM optimises the model by maximising the interval and minimising the total loss. When using support vector machines for nonlinear regression prediction, it is necessary

to assign initial values to the penalty factor C and the kernel function σ. However, the determination of the initial values is related to the prediction effect of the model, so it is necessary to use some bionic algorithms to optimise its hyperparameters, and in this paper, we use the genetic algorithm to optimise its parameters;

(4) Decision tree regression: The decision tree algorithm is a basic method in machine learning; regression decision tree mainly refers to the CART algorithm, and its so-called regression is based on the feature vector to determine the corresponding output value; the regression tree is the feature space, which is divided into several units, and each division module has a specific output. The specific process of regression tree construction is in the input space where the training data are located, which is determined by recursively dividing each region into two sub-regions and deciding the output value on each sub-region to construct a decision binary tree [23]. The decision tree generating function is

$$f(x) = \sum_{m=1}^{M} c_m I(x \in R_m) \quad (6)$$

where M is the number of partitioned regions; c_m is the partitioned region output value, and I is the indicator function.

Decision tree construction typically starts at the root node and recursively creates child nodes based on the best segmentation point for the feature. When selecting segmentation criteria, the decision tree algorithm needs to determine how to evaluate the feature segmentation effect. Segmentation criteria include the selection of features and thresholds that minimise the mean square error, with the ultimate goal of determining the optimal features for segmenting the data;

(5) KNN regression: KNN regression, also known as K-nearest neighbour regression, is based on the core idea that for n-dimensional input variables, each corresponding to a point in the feature space, the output is the category label or predicted value corresponding to that feature vector [24]. Regression prediction using the KNN algorithm is performed by finding the k-nearest neighbours of a new prediction instance, and then taking the mean of the target of these k samples as the sample prediction value, which is expressed with the following formula:

$$y = \frac{1}{K} \sum_{i=1}^{K} y_i \quad (7)$$

Prediction using KNN regression requires the selection of a distance metric to calculate the similarity between the data, and commonly used data metrics are Euclidean distance and Manhattan distance [25]. In this paper, Euclidean distance is chosen to calculate data correlation.

3.2. Introduction to Multi-Model Fusion Methods

A multi-model fusion method is used to combine multiple models to improve the model's generalisation ability and its prediction accuracy, and its basic idea is to combine the prediction results of multiple models to obtain better prediction results [26]. The general multi-model fusion method is used to obtain the final prediction result by voting or the weighted average of the prediction results of the base learners, while in this paper, we improve the prediction accuracy of the model by establishing an improved SSA-BP neural network for potential feature extraction and training using the initial prediction results.

3.2.1. Principles of Improved Sparrow Search Algorithm

The sparrow search algorithm (SSA) is a population optimisation method inspired by the foraging behaviour of sparrows. In nature, sparrows automatically fall into two types of roles when searching for food: discoverers and followers. The role of discoverers is to scout for food sources, and they indicate to the group the location of the food as well as

the search path. Followers rely on the guidance of the discoverers to obtain food. During foraging activities, sparrows observe the movements of their peers, and when high-intake peers are detected, attackers compete with them to increase their feeding opportunities. At the same time, when sparrow populations are threatened, they adopt strategies to escape predators [27].

However, the sparrow search algorithm still has the disadvantages of easily falling into local optimum and high randomness [28]. The improved sparrow search algorithm, also known as the adaptive t-distributed sparrow search algorithm (TSSA), reduces the probability of SSA falling into local optimum solutions and improves the optimisation ability of the algorithm. Compared with SSA, TSSA samples the t-distribution of each sparrow's position when updating its position and redefines the sparrow position updating method. TSSA dynamically selects the probability P to regulate the use of the adaptive t-distribution variation operator. TSSA achieves the adaptive adjustment of sparrow searching to avoid falling into the local optimal solution and improves the algorithm convergence speed, its ability to search for the global optimal solution through the introduction of the adaptive mechanism, and its ability to search for globally optimal solutions [29].

In the original SSA framework, the finder position update at each iteration is described as

$$X_{i,j}^{t+1} = \begin{cases} X_{i,j}^{t} \exp\left(-\frac{i}{\alpha iter_{max}}\right), R_2 < ST \\ X_{i.j}^{t} + QL, R_2 \geq ST \end{cases} \tag{8}$$

In the equation, t represents the current iteration number; $iter_{max}$ is the maximum iteration number; $X_{i,j}$ represents the position of the ith sparrow in the jth dimension; R_2 represents the warning value, and the range is [0, 1]; ST is the safety value, and the range is [0.5, 1]; Q is a random number obeying a normal distribution; L is a 1 × d matrix, and all the elements in the matrix are 1; when $R_2 < ST$, it indicates that there is no predator in the surroundings of the environment; when $R_2 \geq ST$, it indicates that all the sparrows need to move to other safe areas for foraging. When there is no predator, an extensive search can be performed; when $R_2 \geq ST$, it means that a sparrow in the population has found a predator, and it sent a message to the other sparrows; at this time, all the sparrows need to move to other safe areas to forage. The follower's location update is described as

$$X_{i,j}^{t+1} = \begin{cases} Qexp\left(\frac{X_w^t - X_{i,j}^t}{i^2}\right), i > \frac{n}{2} \\ X_P^{t+1} + \left|X_{i,j}^t - X_P^{t+1}\right| \cdot A^+ \cdot L, i \leq \frac{n}{2} \end{cases} \tag{9}$$

where X_P represents the global optimal position; X_W represents the global worst position; A^+ is a 1 × d matrix with all matrix elements 1 or −1 and satisfies $A^+ = A^T(AA^T)^{-1}$. When $i > n/2$, it means that the ith follower did not forage successfully, which leads to its low survival rate, and it needs to go to other areas to forage, and other position information is determined based on the position update formula.

When the population is threatened by a natural enemy, the scout sparrow sends out an early warning signal, and the population engages in anti-predatory behaviour; the position update is mathematically described as

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^{t} + \beta \cdot \left|X_{i,j}^t - X_{best}^t\right|, f_i > f_{best} \\ X_{i,j}^{t} + K \cdot \left(\frac{\left|X_{i,j}^t - X_W^t\right|}{(f_i - f_W + \varepsilon)}\right), f_i = f_{best} \end{cases} \tag{10}$$

In the above expression, X_{best} is the global optimal position; β is the step-size adjustment parameter, which is a normally distributed random number; K is the step-size

adjustment parameter, which indicates the movement position of the sparrow; f_i is the fitness value of the ith sparrow; f_{best} is the current global optimal fitness value; f_w is the worst fitness value; when $f_i > f_{best}$, it means that the sparrow is foraging in the edge area, where it is easily detected by predators. When $f_i = f_{best}$, it means that the sparrow is located in the centre of the population and detects the presence of predators around it, and it needs to move to other areas to avoid predators.

TSSA introduces an adaptive t-distribution at the time of position update, and the position update is described as

$$X_i^{t+1} = X_i^t + X_i^t \cdot t(iter) \tag{11}$$

where $X_i{}^{t+1}$ is the sparrow position after perturbation; $X_i{}^t$ is the current position of the sparrow, and $t(iter)$ is a distribution variation algorithm with the number of iterations as the parameter of degrees of freedom. Dynamic selection probability p is used to regulate the change in the adaptive t-distribution variation operator, which is mathematically expressed as

$$p = w_1 - w_2 \times \left(\frac{iter_{\max} - iter}{iter_{\max}}\right) \tag{12}$$

where w_1 is the upper limit of dynamic probability selection; w_2 is the magnitude of change in dynamic selection probability; $iter_{max}$ is the maximum number of iterations, and $iter$ is the current number of iterations.

3.2.2. Introduction to BP Neural Networks

Because of the nonlinear modelling ability and data-driven characteristics of multilayer neural networks, they have good characteristics in solving regression-type problems. The BP neural network is a multilayer feed-forward neural network, which adjusts the weights and biases in the network through the characteristics of signal forward propagation and error backpropagation [30]. The structure of the BP neural network is divided into an input layer, a hidden layer, and an output layer, and the number of neurons in the input layer is determined by the input features of the model; the number of neurons in the output layer is determined by the number of outputs of the model, and the number of neurons in the hidden layer is derived through empirical formulas or model tuning. The network structure is shown in Figure 4.

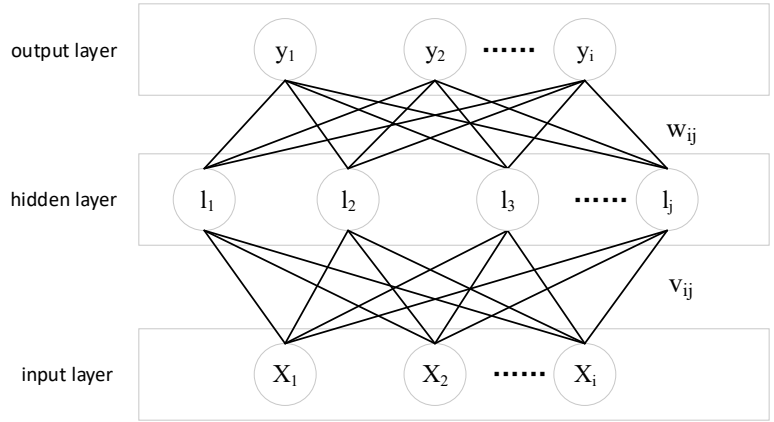


Figure 4. BP neural network structure diagram.

3.2.3. Improving the SSA-BP Model

To avoid the BP neural network from falling into local minima as well as to improve its prediction accuracy, an improved SSA-BP neural network model is established. When

the BP neural network performs a local search, it can find out the updating position of weights and thresholds faster, provide better effect hyperparameters for the training of the BP neural network, and optimise the prediction accuracy and convergence speed of the BP neural network [31]. The flowchart of the improved SSA to optimise the BP neural network is shown in Figure 5.

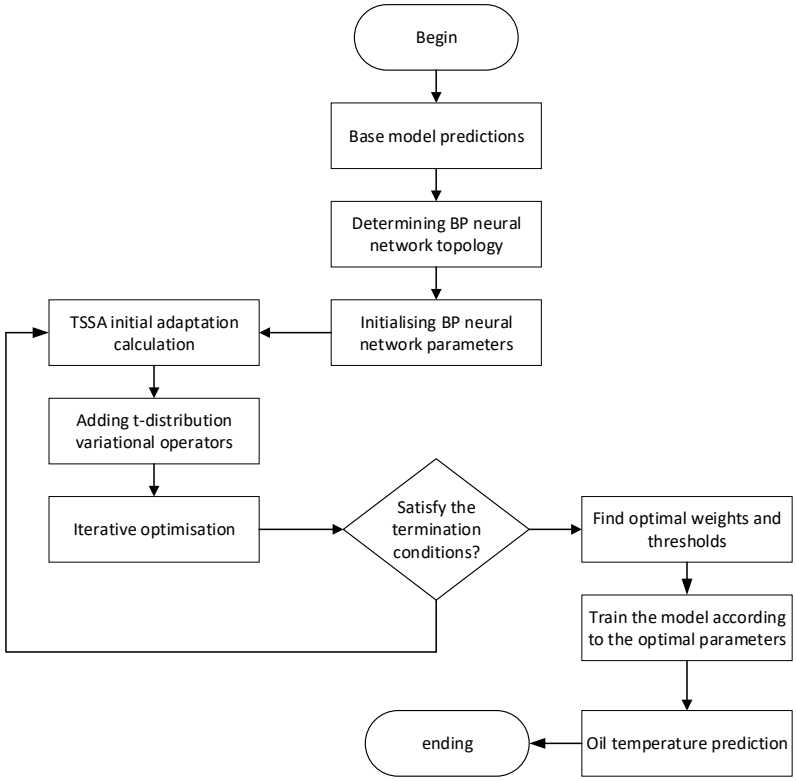


Figure 5. TSSA optimises the BP neural network process.

The process first takes the base model prediction results as input to the BP neural network and initialises the parameters of the BP neural network. After that, the weights and bias of the BP neural network are optimised using the TSSA optimisation algorithm, and the process starts with initialising the sparrow population position, after which the t-distribution variance operator is added to the position updating process, with the discoverer updating the position to simulate the global search and the follower updating the position to simulate the local search, and the optimal individual of the population position is finally determined. After judging that the termination conditions are satisfied, the optimal weights and thresholds are output, and the model is trained according to the optimal parameters, and then the final prediction of transformer oil temperature is completed.

In the multi-model fusion prediction method, the base model includes multiple linear regression, Ridge regression, support vector regression, KNN nearest neighbour regression, and decision tree regression. Among them, the computational complexity of multiple linear regression, support vector regression, Ridge regression, and decision tree regression is relatively low and can be ignored in the overall complexity calculation. But brute-forced KNN nearest neighbour regression needs to be considered for its computational complexity, and its prediction time complexity is $O(n\ f\ +\ k\ f)$,where k is the number of nearest neighbours. In BP neural networks, the training complexity depends on the number of layers of the network and the number of neurons in each layer, which is usually denoted as $O(p{\cdot}q{\cdot}r + q{\cdot}r{\cdot}s)$, where p is the number of neurons in the input layer; q is the number of neurons in the hidden layer; r is the number of hidden layer layers, and s is the number of neurons in the output layer. Model simplification and approximation methods

can be performed to reduce the computational complexity when dealing with large-scale datasets. Model simplification can usually be used to reduce the computational complexity by reducing the number of layers and neurons of the neural network. Approximation methods for KNN nearest neighbour regression use approximate search methods to reduce the computation.

4. Experiment Design and Result Analysis

The experimental design and result analysis section is divided into three subsections, which include dataset division, base model prediction, multi-model fusion experiment, and result analysis. Dataset division is used to divide the processed data into 8 experimental datasets according to different practical situations. The base model prediction is to show the prediction effect of different base models on the divided datasets. The multi-model fusion experiment and result analysis show the results of verification and comparison experiments and analyse them.

4.1. Dataset Partitioning

To verify the feasibility of the multi-model fusion method in the real complex data situation while considering the effects of the season and transformer load state on the model prediction results, 8 sets of experimental datasets are constructed, which are the small-scale and larger-scale datasets with high load in summer, low load in summer, high load in winter, and low load in winter, respectively, and the small-scale dataset consist of 200 sets of experimental data, of which 180 sets are used for model training, and 20 sets are used to validate the effect; the larger-scale dataset consist of 2000 sets of experimental data, of which 1800 sets are used for model training, and 200 sets are used as a test set to validate the model effect.

To verify the effectiveness of the experiments driven by larger-scale real data, this paper first conducted experiments using the base model.

4.2. Base Model Predictions

The multi-model fusion approach first requires the use of a base model for prediction, which provides an initial prediction of the transformer oil temperature through five machine learning methods. Based on the grey correlation calculation, it is determined that the inputs to the model consist of load current, ambient temperature, load factor, and active power, and the output of the model is oil temperature. The small-scale dataset and larger-scale dataset of summer high load are selected for the presentation of the base model's prediction results; a total of 180 sets of data from the summer high load small-scale dataset are used for training, and 20 sets of data are used for testing. The summer high load larger-scale dataset used 1800 sets for training and 200 sets for testing.

The base model approach, in terms of parameter settings, sets alpha for Ridge regression to 0.5, and the optimiser selects auto mode; the penalty factor C for support vector regression is set to 1000, and the kernel function selects the radial basis kernel function; epsilon is set to 0.1, and gamma is set to 0.05. The split quality function for decision tree regression selects the mean squared error; the splitting strategy is defaulted, and the maximum depth of the tree is set to five, and the maximum number of leaf nodes is set to ten. The number of neighbours for KNN regression is set to 30; distance is used for regression weights; the algorithm is selected as a brute, and the distance metric is set as Euclidean. The experiments are performed on the Python software (version 3.13), and the prediction results are shown in Figure 6.

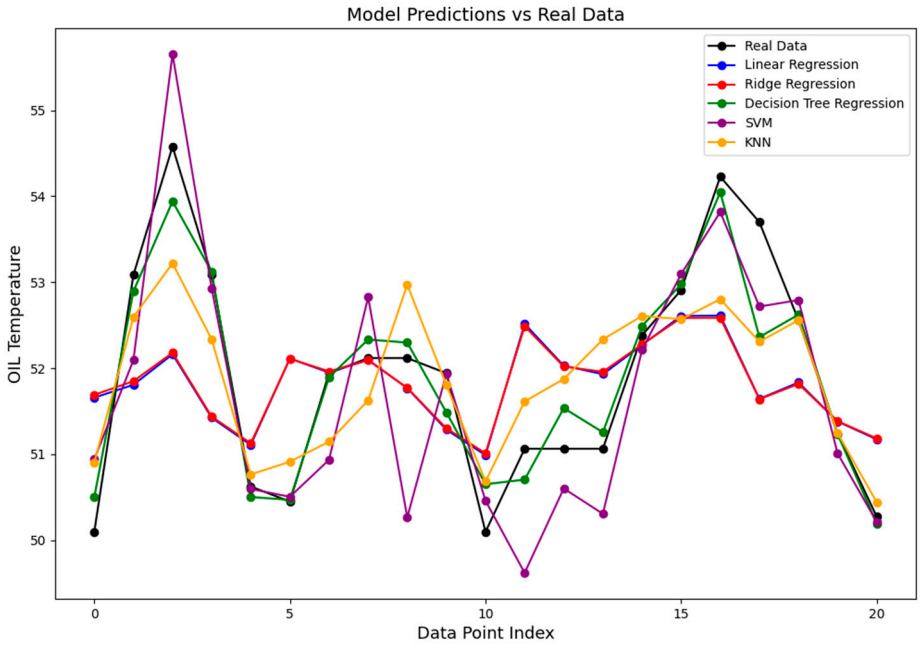


Figure 6. Base model small-scale dataset prediction results.

By analysing the base model's prediction results, it can be obtained that the decision tree method has the best prediction effect and is consistent with the trend of the real data, and the error of multiple linear regression is large, but the prediction trend is consistent with the real data.

The test set for the summer high-load, larger-scale dataset has 200 sets of data that cannot be visualised on a line graph, so the base model prediction results are presented in tabular form.

From Table 7, it can be seen that when predicting larger-scale datasets, due to factors such as increased data volume and complex data situation, the prediction accuracy of the above base model decreases in comparison to small-scale datasets, so a multi-model fusion method is established to improve the prediction ability on larger-scale datasets.

Table 7. Base model larger-scale dataset prediction results.

Base Model Approach	RMSE
multiple linear regression	2.2966
Ridge regression	1.8747
support vector regression	1.7468
decision tree regression	1.4261
KNN regression	1.6359

4.3. Multi-Model Fusion Experiment and Result Analysis

4.3.1. Experimental Design

According to the prediction results of the base model, it can be seen that the prediction accuracy of the base model in larger datasets is lower, so we verify the prediction effect of the multi-model fusion method through further experiments; for this purpose, this paper designs the following experiments for verification, as shown in Table 8.

Table 8. Experimental design.

	Experiment No.	Dataset	Experimental purpose
Verification experiment	Experiment 1	High-load, small-scale datasets in the summer	Validation of model predictions under high load in summer
	Experiment 2	High-load larger datasets in the summer	
	Experiment 3	Low-load, small-scale datasets in the summer	Validation of model predictions under low load in summer
	Experiment 4	Low-load larger datasets in the summer	
	Experiment 5	High-load, small-scale datasets in the winter	Validation of model predictions under high load in winter
	Experiment 6	High-load larger datasets in the winter	
	Experiment 7	Low-load, small-scale datasets in the winter	Validation of model predictions under low load in winter
	Experiment 8	Low-load larger datasets in the winter	
Comparison experiment	Dataset	Prediction model	Experimental purpose
	High-load, small-scale larger datasets in summer	SSA-BP prediction model	Comparing the prediction accuracy of multi-model fusion prediction methods with other prediction methods
		GA-BP prediction model	
		Multi-model fusion prediction	

The validation experiment verifies the prediction effect of the multi-model fusion method in different practical situations by targeting eight different datasets. Comparison experiments use different prediction methods for the same dataset to compare and analyse the prediction effect of different prediction methods.

4.3.2. Verification Experiment

The multi-model fusion approach retrains the base model prediction results through the TSSA-BP model, which is mathematically expressed as

$$f(x) = TSSA - BP(x_1, x_2, x_3, x_4, x_5) \tag{13}$$

where $f(x)$ is the prediction result of the integrated learning model; TSSA-BP stands for the improved sparrow algorithm-optimised BP neural network model, and x_1 to x_5 in parentheses represent the prediction results of the five base models.

In conducting the validation experiments, the experimental validation is first carried out for the small-scale dataset; for example, Experiment 1, Experiment 3, Experiment 5, and Experiment 7 are carried out, and their results are analysed.

In terms of the parameter settings of the multi-model fusion method, the parameter settings for the small-scale and larger-scale datasets are not exactly the same. The parameter settings for the small-scale dataset are as follows: the number of sparrow populations is set to 20; the number of sparrow search iterations is set to 30; the optimisation parameter dimensionality is set to two dimensions; the warning value ST is set to 0.8, and the percentage of discoverers is set to 20%. The number of nodes in the input layer of the BP neural network is set to five. The number of nodes in the hidden layer is set to ten, and the number of nodes in the output layer is set to one. The activation function is selected as ReLU; the learning rate is set to 0.01; the number of iterations is set to 500, and the loss function is selected as MSE.

The evaluation metrics are selected as root mean square error and mean absolute percentage error. Root mean square error (RMSE) is an evaluation metric used to measure the difference between the predicted value of the model and the true value; the closer the value of *RMSE* is to zero, the more accurate the prediction of the model is. This is because it squares the error, thereby giving more weight to larger prediction errors. Mean absolute percentage error (MAPE) is a measure of prediction accuracy, which evaluates the model's performance by calculating the absolute value of the difference between the predicted value and the true value as a percentage of the true value, and the smaller the value of *MAPE*, the better the prediction performance of the model. Both indicators are used to measure the difference between the predicted value and the actual value, and the root mean square error is calculated as

$$RMSE(X,h) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(h(x_i) - y_i)^2} \tag{14}$$

The mean absolute percentage error is calculated as

$$MAPE(X,h) = \frac{1}{m}\sum_{i=1}^{m}\left|\frac{h(x_i) - y_i}{y_i}\right| \tag{15}$$

In the formula, m denotes the sample; $h(x_i)$ denotes the predicted value of the ith sample, and y_i denotes the actual value of the ith sample.

The small-scale datasets with high summer load, low summer load, high winter load, and low winter load are modelled and predicted using multi-model fusion prediction methods to analyse the prediction performance of multi-model fusion prediction methods on different situations and data sizes, respectively. The prediction results of Experiment 1, Experiment 3, Experiment 5, and Experiment 7 are shown in Figures 7–10.

Figure 7 shows the prediction results of Experiment 1, with an RMSE of 0.5581 and an MAPE of 0.81%, whose prediction trends are consistent with the actual values, but the prediction is generally effective at some peak points.

Figure 8 shows the prediction results of Experiment 3, with an RMSE of 0.6525 and an MAPE of 0.96%, from which it can be seen that the trend of the predicted and actual values of the oil temperature is the same, and the prediction error is larger at the peak point compared to the other sample points.

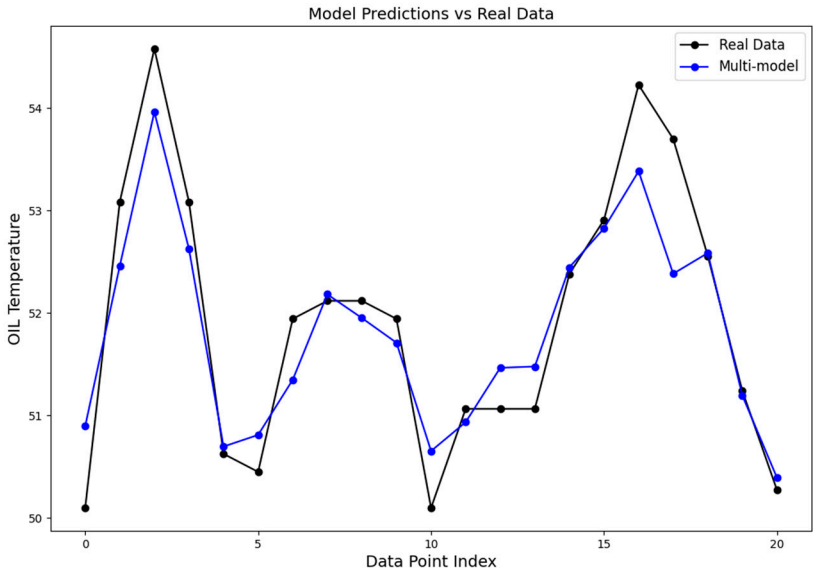


Figure 7. Results of high-load, small-scale data prediction in summer.

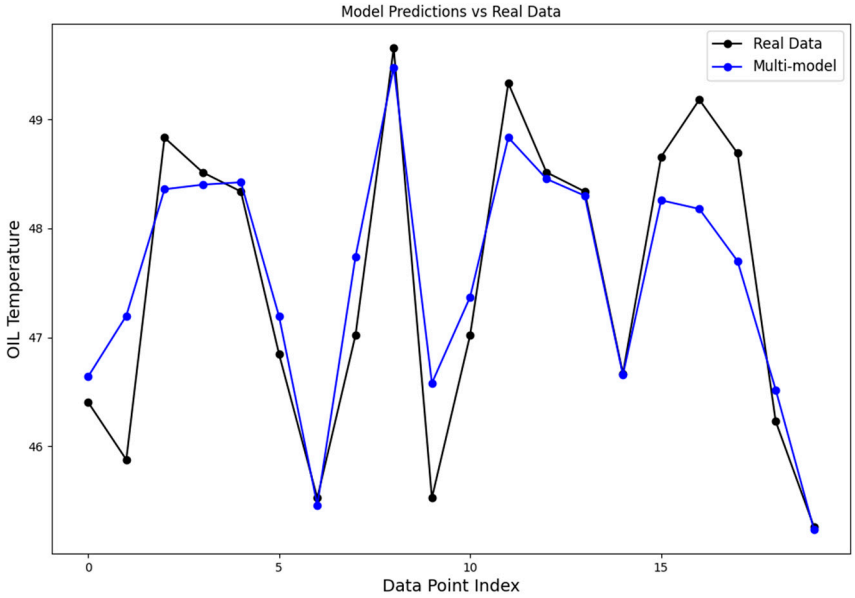


Figure 8. Results of low load small-scale data prediction in summer.

Figure 9 shows the prediction results of Experiment 5, with an RMSE of 0.1567 and an MAPE of 0.38%, from which it can be seen that the prediction errors for all test sample points are small, indicating that the multi-model fusion method has a better prediction effect on this dataset.

Figure 10 shows the prediction results of Experiment 7; the RMSE is 0.3692, and the MAPE is 0.67%. From the figure, it can be seen that the trend of the predicted value is basically consistent with the real value, and the model prediction error is small.

As can be seen from Figures 7–10, the predicted values of the data-driven multi-model fusion-based approach on small-scale datasets are consistent with the trend of the actual values. It has good prediction accuracy, both in the cases of summer low-load and high-load and winter low-load and high-load small-scale datasets.

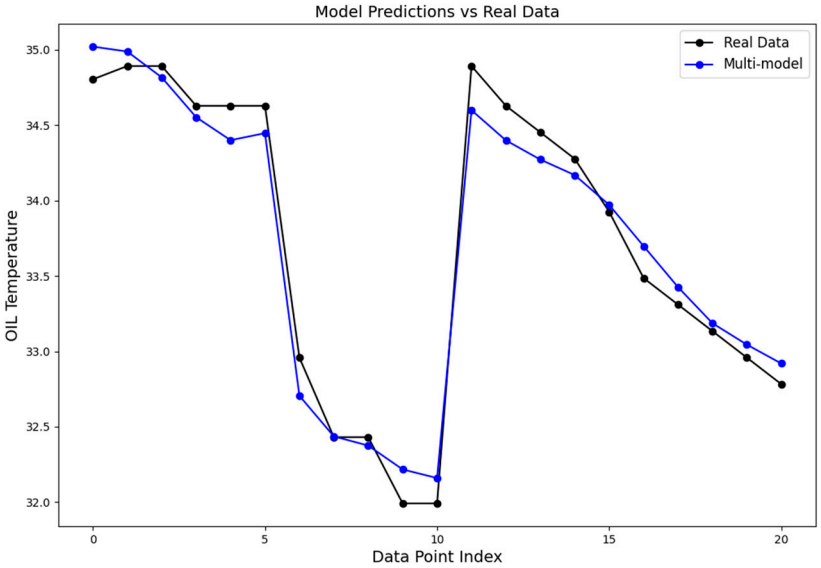


Figure 9. Results of low-load, small-scale data prediction in winter.

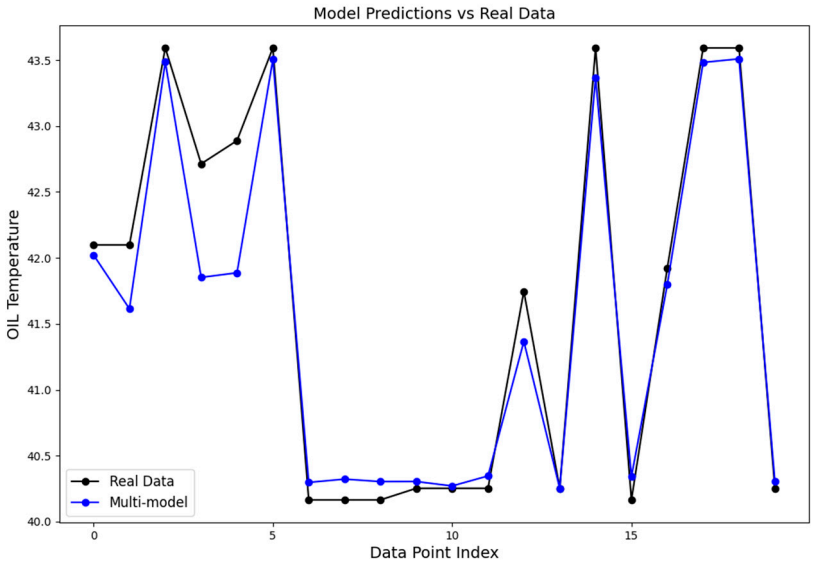


Figure 10. Results of high-load, small-scale data prediction in winter.

When experiments are conducted for larger datasets, i.e., when Experiment 2, Experiment 4, Experiment 6, and Experiment 8 are conducted, the prediction results are presented in a tabular form, taking into account the large number of predicted sample points.

Parameter settings for the larger dataset were determined experimentally. The number of sparrow populations was set to 50; the number of sparrow search iterations was set to 100; the optimisation parameter dimension was 2-dimensional; the warning value ST was set to 0.8, and the percentage of discoverers was 20%. The number of nodes in the hidden layer of the BP neural network was set to 10; the activation function selected the ReLU function; the learning rate was set to 0.001; the number of iterations was set to 1000, and the loss function was selected as MSE.

The prediction results of the multi-model fusion prediction method on larger datasets are shown in Table 9.

Table 9. Predicting results for larger datasets.

	RMSE	**MAPE**
Experiment 2	0.6479	0.95%
Experiment 4	0.5770	0.93%
Experiment 6	0.3694	0.71%
Experiment 8	1.0877	1.58%

As can be seen from Table 8, the multi-model fusion method also has good results in the 2000-group data size case, which proves the feasibility of the method for larger datasets as well. The prediction error is minimised in the case of high load in winter, with an RMSE of 0.3694 and an MAPE of 0.71%. The prediction error is highest in the case of low load in winter, with an RMSE of 1.0877 and an MAPE of 1.58%. From the experimental results, it can be obtained that the error between the predicted value and the actual value of the multi-model fusion prediction method is about one degree, and the purpose of the transformer oil temperature prediction is to evaluate the transformer operation status or obtain early warning of abnormal data according to the predicted value, and the error of one degree or less is basically in line with the actual requirements of the project.

4.4. Comparison Experiment

The GA-BP prediction model and the SSA-BP prediction model are established from the same sample dataset. The GA-BP prediction model represents the genetic algorithm-optimised BP neural network [32], and the SSA-BP prediction model represents the sparrow search algorithm-optimised BP neural network [33]. The effects of different prediction methods on the same dataset are explored through comparative experiments, and the three models are trained on the summer high-load, small-scale dataset and the summer high-load, larger-scale dataset, respectively, and the effects are verified. The prediction results of the three models on the summer high-load, small-scale dataset are shown in Figure 11.

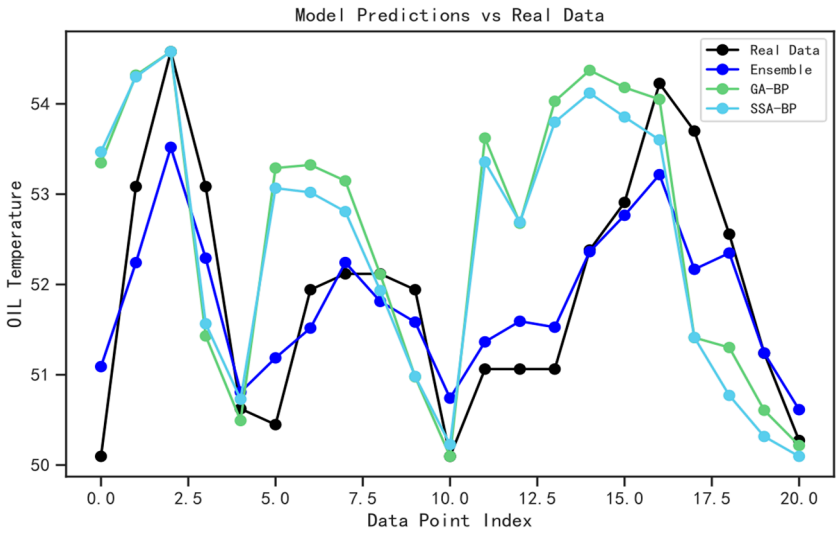


Figure 11. Comparison of prediction curves of different models.

By comparing and analysing the prediction curves of different methods in the figure, we find that the prediction effect of the multi-model fusion method is better than the other prediction methods. The root mean square errors of the multi-model fusion prediction method, the GA-BP model, and the SSA-BP model on the summer high-load small-scale dataset are 0.5581, 1.8437, and 1.7184, respectively, and the experimental results proved that the multi-model fusion method has the smallest error, and the trend of the prediction curves is the same as the real value.

The prediction results of the three models on the larger-scale dataset with high load in summer are shown in Table 10, from which it can be seen that the multi-model fusion method used in this paper has a root-mean-square error of 0.6479 and an average absolute percentage error of 0.95%, which is the smallest of the three prediction model errors, further validating the feasibility of the method.

Table 10. Different models predict different results with the larger datasets.

Prediction Model	RMSE	MAPE
Multi-model fusion method	0.6479	0.95%
SSA-BP model	1.6659	2.91%
GA-BP model	1.4059	2.22%

5. Conclusions

In this paper, a transformer oil temperature prediction method based on data-driven and multi-model fusion is proposed, and the innovation of this method has two main points:

1. Data processing and experimental validation based on larger-scale real operating data can, on the one hand, be used to study the abnormal data detection method for real operating data, and the premise of the data-driven approach is to ensure the standardisation and accuracy of the data, taking into account the complexity of the real data. Therefore, the normality and accuracy of datasets are improved through data analysis and processing. On the other hand, for a large amount of various types of real operating data under different conditions, the prediction effect of the multi-model fusion method is investigated by constructing eight groups of different types of datasets, and the prediction effect of the multi-model fusion method is investigated on datasets with different seasons, loads, and sizes to validate the method's generalisation performance for the real data. 2. The multi-model fusion prediction method is established for transformer oil temperature prediction, and the initial prediction results of different machine learning methods are used as model inputs to the TSSA-BP neural network. The multi-model fusion prediction method uses different base models for initial prediction to improve the generalisation ability of the model, and then uses TSSA to optimise the parameters of the BP neural network, and it finally learns the features of the base model prediction results and the real values via TSSA-BP, which improves the generalisation ability and prediction accuracy of the prediction model.

The experimental results show that the maximum root mean square error and the mean absolute percentage error of this method on different datasets are 1.0877 and 1.58%. The results of the verification experiments show that the error of the multi-model fusion method for oil temperature prediction on the real transformer operation and inspection data is about one degree, which verifies the practicability of the method on transformer oil temperature prediction. Comparison experiment results show that the prediction accuracy of the multi-model fusion prediction method proposed in this paper is better than that of other prediction methods, which verifies the feasibility of the method in transformer oil temperature prediction.

Multi-model fusion prediction methods also have corresponding limitations, mainly in the base model parameter optimisation problem and computational cost considerations; different base models have different model parameters, and time-consuming parameterization of different models is needed to avoid the occurrence of overfitting or underfitting. The multi-model fusion approach increases the computational cost, which needs to be considered in real-time prediction applications for large-scale data. Although the method proposed in this paper achieves good prediction results on large-scale datasets, the actual modelling data in engineering may be larger and more complex, so subsequent research will model the prediction for larger-scale datasets to improve the generalisation ability and prediction performance of the model as much as possible.

Author Contributions: L.Y. is responsible for the establishment of the multi-model fusion method and the writing of the paper manuscript. L.C. is responsible for experimental verification and result analysis. F.Z. is responsible for data processing and analysis. S.M. is responsible for the establishment of the comparison test model. Y.Z. and S.Y. are responsible for format revision and grammar checking of the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by research project of State Grid Sichuan Electric power company (52199723003P).

Data Availability Statement: Restrictions apply to the datasets. The datasets presented in this article are not readily available because the datasets used in this paper belong to the confidential documents of the power system.

Conflicts of Interest: Authors Liang Chen, Shen Ma, Yang Zhang and Sixu Yang were employed by the company State Grid Meishan Power Supply. 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. The company had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Guo, Y.; Chang, Y.; Lu, B. A review of temperature prediction methods for oil-immersed transformers. *Measurement* **2025**, *239*, 115383. [CrossRef]
2. Yang, C. Study on Temperature Characteristics and Hot Spot Temperature Prediction of Oil-Immersed Transformer. Master's Thesis, Nanjing University of Posts and Telecommunications, Nanjing, China, 2023.
3. Liao, C.B.; Ruan, J.J.; Wei, C.; Lu, Y.; Wu, P.; Li, J.S. Research on transformer real-time hot spot temperature prediction method based on improved support vector machine. *High Volt. Appar.* **2018**, *54*, 174–179. [CrossRef]
4. Cicek, A.; Şengör, I.; Guner, S.; Karakus, F.; Erenoglu, A.K.; Erdinc, O.; Shafie-Khah, M.; Catalao, J.P.S. Integrated rail system and EV parking lot operation with regenerative braking energy, energy storage system and PV availability. *IEEE Trans. Smart Grid* **2022**, *13*, 3049–3058. [CrossRef]
5. Pradhan, M.K.; Ramu, T.S. Prediction of hottest spot temperature (HST) in power and station transformers. *IEEE Trans. Power Deliv.* **2003**, *18*, 1275–1283. [CrossRef]
6. Li, M.L. Research on Transformer Winding Hot Spot Temperature Prediction Method Based on BP Neural Network. Master's Thesis, Chongqing University, Chongqing, China, 2010.
7. Li, S.Q.; Chen, D.; Chou, Q.H.; Shi, J.L.; Xu, W.M.; Song, X. Transformer winding hot spot temperature prediction based on grey neural network. *Autom. Instrum.* **2017**, *4*, 116–118. [CrossRef]
8. Jiang, B.; Yang, C.; Yang, Y.T.; Chao, Y.F. Transformer hot spot temperature prediction based on ACO optimized BP neural network. *J. Electron. Meas. Instrum.* **2022**, *36*, 235–242. [CrossRef]
9. Zhang, Z.; Kong, W.; Li, L.; Zhao, H.; Xin, C. Prediction of Transformer Oil Temperature Based on an Improved PSO Neural Network Algorithm. *Recent Adv. Electr. Electron. Eng. (Former. Recent Pat. Electr. Electron. Eng.)* **2024**, *17*, 29–37. [CrossRef]
10. Lee, K.R. A Study on SVM-Based Speaker Classification Using GMM-supervector. *J. IKEEE* **2020**, *24*, 1022–1027.
11. Chen, W.G.; Teng, L.; Liu, J.; Peng, S.Y.; Sun, C.X. Transformer winding hot spot temperature prediction model based on support vector machine optimized by genetic algorithm. *Trans. China Electrotech. Soc.* **2014**, *29*, 44–51. [CrossRef]
12. Xi, Y.; Lin, D.; Yu, L.; Chen, B.; Jiang, W.; Chen, G. Oil temperature prediction of power transformers based on modified support vector regression machine. *Int. J. Emerg. Electr. Power Syst.* **2022**, *24*, 367–375. [CrossRef]
13. Liu, C.; Lu, Y.J.; Liu, H.Y.; Xiang, X.; Wang, L.W. Transformer winding hot spot temperature prediction based on improved PSO-LSSVM model. *Inn. Mong. Electr. Power* **2021**, *39*, 94–97. [CrossRef]
14. Liang, F.; Yang, X.; Le, X.W.; Jia, P.F.; Jin, H.; Liu, Y. Hot spot temperature prediction method of oil-immersed air-cooled transformer based on ACO-SVM model. *Transformer* **2023**, *60*, 6–12. [CrossRef]
15. Liu, J.K.; Xiao, T.; Wei, Y.; Liu, Y.F.; Liu, C.; Chen, H.X. Hot spot temperature prediction of transformer winding based on PCA-IHHO-LSSVM. *Shandong Electr. Power* **2024**, *51*, 59–66.
16. Mikhak-Beyranvand, M.; Faiz, J.; Rezaeealam, B. Thermal analysis and derating of a power transformer with harmonic loads. *IET Gener. Transm. Distrib.* **2020**, *14*, 1233–1241. [CrossRef]
17. Xu, L.; Ma, H.; Ren, D. Reliability analysis of tractor multi-way valves based on the improved weighted grey relational method. *J. Eng.* **2019**, *2019*, 86–92. [CrossRef]
18. Côté, P.O.; Nikanjam, A.; Ahmed, N.; Humeniuk, D.; Khomh, F. Data cleaning and machine learning: A systematic literature review. *Autom. Softw. Eng.* **2024**, *31*, 54. [CrossRef]
19. Singh, B.; Khan, R.U.; Khan, M.A. Characterizations of q–Weibull distribution based on generalized order statistic. *J. Stat. Manag. Syst.* **2019**, *22*, 1573–1595. [CrossRef]
20. Bazilevskiy, M.P. Multi-criteria approach to pair-multiple linear regression models constructin. *Izv. Saratov University. Math. Mech. Inform.* **2021**, *21*, 88–99.
21. Choi, S.H.; Jung, H.Y.; Kim, H. Ridge Fuzzy Regression Model. *Int. J. Fuzzy Syst.* **2019**, *21*, 2077–2090. [CrossRef]
22. Pavithraa, S.G. Analysis and Comparison of Prediction of Heart Disease Using Novel Support Vector Machine and Logistic Regression Algorithm. *Cardiometry* **2022**, *25*, 783–787. [CrossRef]
23. Ren, S.P.; Li, J.; Wang, J.R.; Yue, K. Lncrna-disease association prediction method based on ensemble regression decision tree. *Comput. Sci.* **2022**, *49*, 265–271.
24. Gangwar, A.K.; Shaik, A.G. k-Nearest neighbour based approach for the protection of distribution network with renewable energy integration. *Electr. Power Syst. Res.* **2023**, *220*, 109301. [CrossRef]
25. Poojary, S. A Hybrid Regression Model for improving prediction accuracy. *Electron. J. Appl. Stat. Anal.* **2023**, *16*, 784–801.
26. Zheng, Y.Y.; Li, X.; Chen, T.X.; Zhao, Y.N. Short-term wind power forecasting method for extreme weather based on Stacking multi-model fusio. *High Volt. Eng.* **2024**, *50*, 3871–3882. [CrossRef]

27. Zhang, Z.; He, R.; Yang, K. A bioinspired path planning approach for mobile robots based on improved sparrow search algorithm. *Adv. Manuf.* **2022**, *10*, 114–130. [CrossRef]
28. Shi, W.Y.; Wei, J.Q.; Zhao, Y.H. Transmission line fault diagnosis based on improved sparrow algorithm-support vector machine. *Zhejiang Electr. Power* **2021**, *40*, 38–45.
29. Zheng, Y.; Li, L.; Qian, L.; Cheng, B.; Hou, W.; Zhuang, Y. Sine-SSA-BP ship trajectory prediction based on chaotic mapping improved sparrow search algorithm. *Sensors* **2023**, *23*, 704. [CrossRef]
30. Li, H. Network traffic prediction of the optimized BP neural network based on Glowworm Swarm Algorithm. *Syst. Sci. Control. Eng.* **2019**, *7*, 64–70. [CrossRef]
31. Li, N.; Wang, W. Prediction of mechanical properties of thermally modified wood based on TSSA-BP model. *Forests* **2022**, *13*, 160. [CrossRef]
32. Wang, X.T.; Zou, Y.; Yu, C.Y. Research on transformer winding hot spot temperature prediction based on GA-BP neural network. *Northeast. Electr. Power Technol.* **2021**, *42*, 1–4+8.
33. Wang, J.; Zeng, G.; Xu, M.; Wan, X.; Wang, K.; Mou, J.; Hua, C.; Fan, C.; Han, P. SSA-BP Neural Network Model for Predicting Rice-Fish Production in China. *J. Appl. Ichthyol.* **2024**, *2024*, 573996. [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.