


Article

A Heat Load Prediction Method for District Heating Systems Based on the AE-GWO-GRU Model

Yu Yang^{1,2}, Junwei Yan^{1,2,3}  and Xuan Zhou^{1,2,3,*}

¹ School of Mechanical and Automotive Engineering, South China University of Technology, Guangzhou 510640, China; 202120101672@mail.scut.edu.cn (Y.Y.); mmjwyan@scut.edu.cn (J.Y.)

² Guangzhou Institute of Modern Industrial Technology, Guangzhou 511458, China

³ Artificial Intelligence and Digital Economy Guangdong Provincial Laboratory (Guangzhou), Guangzhou 511442, China

* Correspondence: zhouxuan@scut.edu.cn

Abstract: Accurate prediction of the heat load in district heating systems is challenging due to various influencing factors, substantial transmission lag in the pipe network, frequent fluctuations, and significant peak-to-valley differences. An autoencoder—grey wolf optimization—gated recurrent unit (AE-GWO-GRU)-based heat load prediction method for district heating systems is proposed, employing techniques such as data augmentation, lag feature extraction, and input feature extraction, which contribute to improvements in the model's prediction accuracy and heat load control stability. By using the AE approach to augment the data, the issue of the training model's accuracy being compromised due to a shortage of data is effectively resolved. The study discusses the influencing factors and lag time of heat load, applies the partial autocorrelation function (PACF) principle to downsample the sequence, reduces the interference of lag and instantaneous changes, and improves the stationary characteristics of the heat load time series. To increase prediction accuracy, the GWO algorithm is used to tune the parameters of the GRU prediction model. The prediction error, measured by RMSE and MAPE, dropped from 56.69 and 2.45% to 47.90 and 2.17%, respectively, compared to the single GRU prediction approach. The findings demonstrate greater accuracy and stability in heat load prediction, underscoring the practical value of the proposed method.

Keywords: heat load prediction; data augmentation; GRU neural network; grey wolf optimization algorithm; parameter optimization



Citation: Yang, Y.; Yan, J.; Zhou, X. A Heat Load Prediction Method for District Heating Systems Based on the AE-GWO-GRU Model. *Appl. Sci.* **2024**, *14*, 5446. <https://doi.org/10.3390/app14135446>

Academic Editor: Maria Vicidomini

Received: 23 May 2024

Revised: 15 June 2024

Accepted: 20 June 2024

Published: 23 June 2024



Copyright: © 2024 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

1.1. Background

District heating systems (DHSs) utilize local heat sources to distribute heat to buildings via heating pipes. Primarily, they serve to fulfill the heating requirements of residential, commercial, and industrial premises [1]. DHSs also offer benefits such as energy conservation and centralized management convenience. China's district heating industry has experienced rapid growth in recent years, witnessing significant expansions in heating coverage and capacity, along with increased adoption. As shown in Figure 1, the urban district heating area nearly doubled over a decade, with an average annual growth rate of 7.7%, increasing from 5.72 billion square meters in 2013 to 11.13 billion square meters in 2022 [2].

The expansion of the district heating area continues unabated, inevitably leading to substantial energy consumption. Over the past decade, China's civil buildings have accounted for between 20% and 25% of the nation's total energy consumption. By the end of 2019, district heating in northern cities and towns consumed over 200 million tce of energy annually, constituting 21% of the energy utilized for building operations [3]. This underscores the significant potential for energy savings through district heating.

Enhancing the operation of DHSs holds paramount importance in mitigating air pollution and reducing carbon dioxide emissions [4].

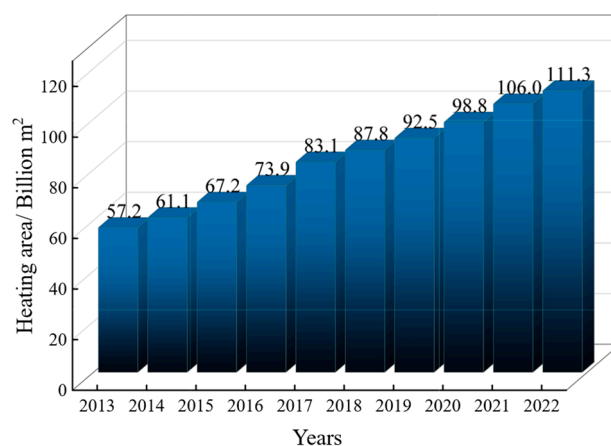


Figure 1. The change trend in urban district heating area from 2013 to 2022.

As a pivotal technology for district heating, heat load prediction plays a crucial role in ensuring the safe, stable, economical, and environmentally friendly operation of the DHS [5]. By employing heat load prediction, the heat source can align with the heat supply according to the heat load demand to fulfill indoor thermal comfort requirements. The varying heat demands of heat users across different time periods coupled with the influence of transmission distance render heat load prediction challenging due to the characteristics of time variation, significant lag, and nonlinearity.

1.2. Related Works

In the early stages of heat load prediction research, statistical analysis techniques were primarily used, leveraging statistical and mathematical expertise to construct prediction models. Their straightforward form offers significant advantages in modeling and prediction speed [6]. Some widely used statistical analysis prediction models include recursive least squares (RLSs) [7], multiple linear regression (MLR) [8], auto-regressive (AR) [9], autoregressive with extra inputs model (ARX) [10], autoregressive and moving average model (ARMA) [11], and autoregressive integrated moving average model (ARIMA) [12]. Despite their lower modeling complexity and shorter modeling times, these statistical models typically only consider time factors, neglecting other external elements that can influence the prediction model. Multi-dimensional heat load sequence data have nonlinear relationships that are difficult to capture, making these models less reliable in complex heat load prediction scenarios.

With the rapid advancement of industrial hardware and artificial intelligence technology, models based on machine learning and deep learning have demonstrated significant capabilities for learning and mapping nonlinear objects, offering new potential for the development of heat load prediction. Examples of popular machine learning regression techniques include support vector regression (SVR) [13], decision tree (DT) [14], random forest (RF) [15], and neural network models [16]. Li et al. [17] developed a heat load prediction model for heating systems using a BP neural network by quantifying temperature and date types. The genetic algorithm optimized the neural network's connection weights and thresholds, resulting in accurate 24-h heat load predictions. Xu et al. [18] determined the optimal parameters of the multilayer perceptron (MLP) through different optimization algorithms to establish a heat load prediction model. The RMSE, R^2 , and MAE of the MLP model optimized by the biogeography-based optimization algorithm are 2.82, 0.92, and 2.15, respectively. Despite their advantages of speed and high precision for small data sets, these traditional machine learning algorithms are ineffective for handling large-scale data sets and often neglect the feature expression of heat load data in time series.

Recurrent neural networks (RNNs) [19] have gained popularity in time series prediction due to their superior ability to remember historical data compared to other algorithms. Long short-term memory (LSTM) [20] introduces a gated unit structure based on RNNs, enabling it to learn sequence dependencies in prediction problems. The gated recurrent unit (GRU) [21] simplifies the LSTM model structure, ensuring prediction accuracy without overfitting and offering better adaptability. Xu et al. [22] developed a GRU-based heat load prediction model that fully leverages GRU's ability to memorize long-term time series data, resulting in enhanced prediction accuracy. However, GRU networks that rely solely on empirical parameters can be more prone to randomness and may fall into local optima due to their complex parameters, potentially impacting final heat load prediction values. The grey wolf optimization (GWO) algorithm [23], a swarm intelligence optimization method that simulates the predation behavior of grey wolves, has been recently proposed. Its advantages include easy implementation, few parameters, and strong convergence. Compared to other popular algorithms such as particle swarm optimization (PSO) and genetic algorithm (GA) [24], the GWO algorithm offers higher convergence speed and superior search capabilities. Consequently, it is frequently employed in model parameter optimization and adjustment [25,26].

The GRU model requires a substantial amount of data for training. Inadequate training data can lead to overfitting issues and reduce the model's accuracy. A common method to address this issue is data augmentation. Inspired by the emerging Autoencoder (AE) generation model, the AE-based data augmentation method [27–29] can enhance the performance of deep learning models.

1.3. Structure

Based on the above research and fully considering the influence of various parameters and meteorological factors, this paper proposes an AE-GRU heat load prediction model optimized by the GWO algorithm, using a DHS as a case study. Firstly, to address the issue of short system running time and insufficient training data, AE improves the model's generalization performance by increasing the amount of effective data. Secondly, the GWO algorithm is utilized to optimize the GRU network's model parameters. The GRU model is then trained with the optimal parameter combination to predict the heat load of the DHS.

2. Methodology

2.1. The Proposed Method

In this paper, a heat load prediction method for a DHS based on AE-GWO-GRU is designed and implemented. The overall framework of the system consists of two main parts: data analysis and processing, and model training and prediction. The overall framework is shown in Figure 2.

- (1) The data analysis and processing part mainly includes four steps: data acquisition, data cleaning, data analysis, and data processing.
 - a. Data Acquisition: Multi-type sensors in the DHS acquire system operating parameters and outdoor meteorological factors during each cycle. These data are then uploaded and stored in the database.
 - b. Data Cleaning: Outliers, abnormal zeros, and missing values in the original data set are processed to obtain a complete and high-quality multi-dimensional heat load time series.
 - c. Data Analysis: The analysis includes examining the impact of system lag on prediction modeling and determining the prediction period. Additionally, the correlation between the influencing factors of heat load is investigated to determine the input features and dimensions for the prediction model.
 - d. Data Processing: The multi-dimensional heat load time series data are augmented and then divided into training, validation, and test sets in a ratio of 6:2:2. The data are normalized to a form suitable for model training and prediction.

- (2) The model training and prediction component includes basic model training and model parameters optimization.
 - a. Basic Model Training: The GRU basic prediction model is trained using the default model parameters.
 - b. Model Parameters Optimization: The GWO algorithm optimizes the key model parameters and determines the optimal parameter combination.
 - c. Online Prediction: The GRU model is trained and used for prediction with the optimal parameter combination.

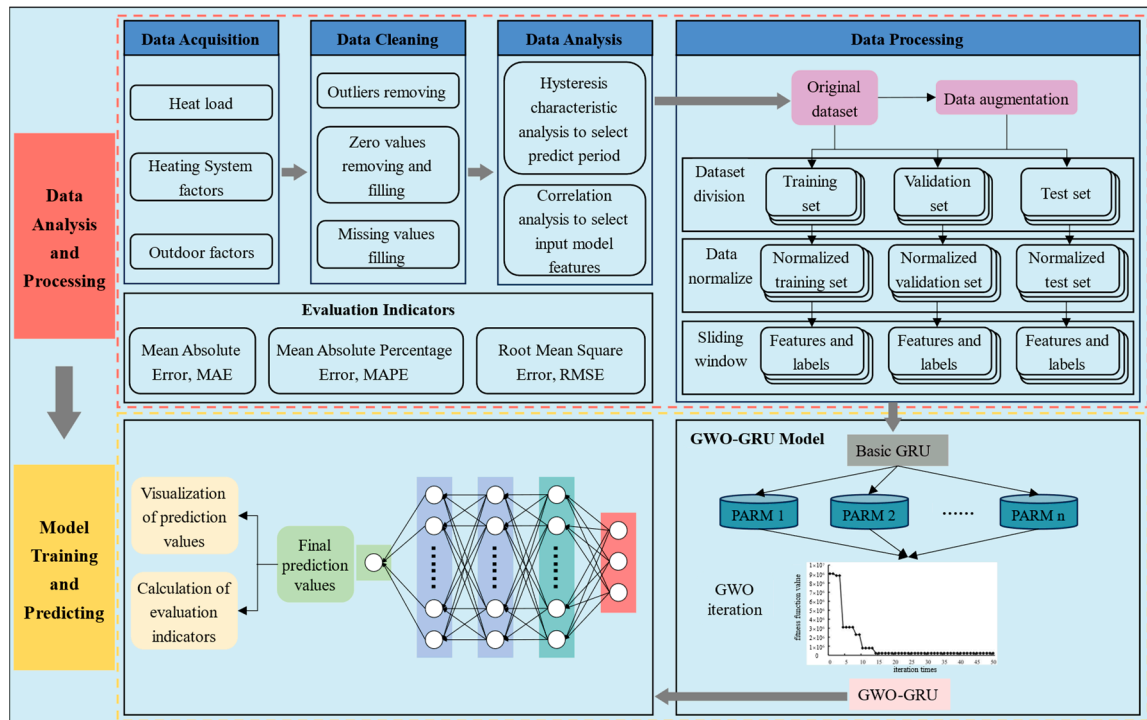


Figure 2. Overall system framework diagram.

2.2. GRU Prediction Model Based on GWO

The determination of GRU model parameters is often based on artificial experience, leading to problems such as long model adjustment times and convergence to local optimal solutions. To improve the accuracy of the heat load prediction model, this paper employs GWO to optimize and adjust the GRU model parameters. The number of GRU layer neurons, the number of hidden layer neurons, the learning rate, the dropout rate, and the training batch size in the GRU network are used as the positions of the wolf group. By calculating the fitness function, the positions of the wolf group are updated to obtain the optimal GRU network model parameters, which are then used to construct the heat load prediction model.

2.2.1. Gated Recurrent Unit

The GRU neural network is an improved version of the LSTM neural network. Compared to the standard LSTM network, the GRU reduces the gating mechanisms from three to two: the reset gate and the update gate. This simplification enhances the ability to capture the correlation between time and information [30]. GRU facilitates input updates for long data sequences and mitigates the vanishing gradient problem, improving iteration efficiency and prediction speed without sacrificing accuracy. The GRU network structure is shown in Figure 3.

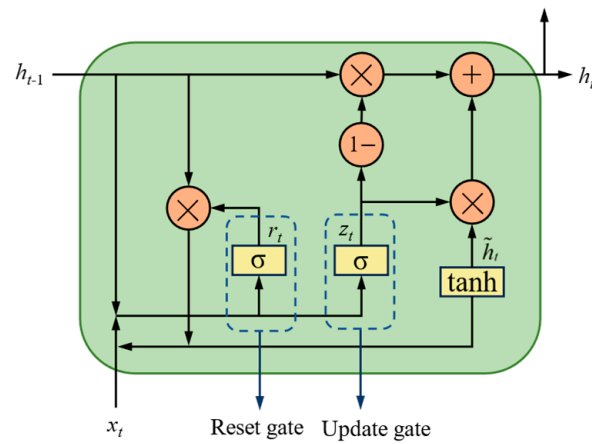


Figure 3. GRU model diagram.

In the figure, r_t represents the reset gate, z_t represents the update gate, \tilde{h}_t represents the candidate hidden layer, h_t represents the hidden layer, and σ represents the Sigmoid function. The reset gate r_t controls the amount of candidate hidden layer information from time $t-1$ that is passed into time t . The update gate z_t controls the amount of state information updated from time $t-1$ to time t . The candidate hidden layer \tilde{h}_t calculates the candidate hidden state at time t , and then combined with the reset gate to determine the final hidden state. The hidden layer h_t stores information from time $t-1$ and passes it to time t . The GRU network update formulas are as follows:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{1}$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{2}$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]) \tag{3}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{4}$$

where W_r , W_z and $W_{\tilde{h}}$ represent the weight matrix of reset gate, update gate and candidate hidden layer, respectively. “[]” denotes the connection of two matrices, “*” denotes the matrix product, and “.” denotes the point multiplication.

2.2.2. Grey Wolf Optimization

The GWO is a swarm intelligence optimization algorithm that simulates the hierarchy and hunting mechanism of gray wolves. It offers high precision and strong convergence, making it suitable for optimizing GRU parameters. During predation operations, the gray wolf population follows a strict social hierarchy that influences the behavior and role assignment of individuals. The hierarchy consists of four roles: leader (α), sub-leader (β), bottom wolf (δ) and candidate wolf (ω). In each iteration, the best three wolves (α , β , δ) in the current population are retained, and the positions of other bottom wolves are updated based on their position information. The position update process of the wolves corresponds to the optimization of GRU parameters, as shown in Figure 4.

In the figure, D_α , D_β and D_δ represent the distance between α , β , δ , and other gray wolves. The optimization process of GWO starts from initialization, then updates the positions of the gray wolves through iterations, and finally converges to the approximate optimal solution. The basic optimization process is shown in Figure 5.

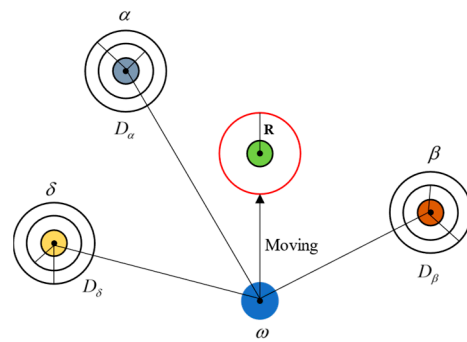


Figure 4. Candidate wolves position update diagram.

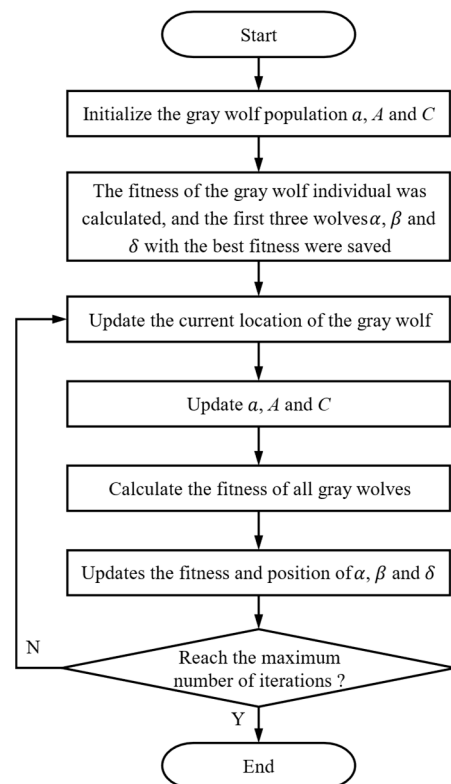


Figure 5. GWO algorithm flowchart.

3. Case Study

3.1. Research Object

In this paper, the DHS of a multifunctional comprehensive region is taken as the research object. The total heating building area in this region is 43,000 m², including offices buildings, apartments, public restaurants, and clinics. The buildings have an ordinary brick–concrete structure with good thermal insulation performance. The DHS comprises three parts: the heat source, heating pipe network, and heat users, as shown in Figure 6. The heating system is equipped with a data acquisition system that collects and stores operation parameters and outdoor meteorological parameters in the database. These parameters include the temperature and pressure of the hot water supply and return, the hot water flow, outdoor temperature, and outdoor relative humidity. Due to the short running time of the data acquisition system and the insufficient amount of stored data, a data augmentation method needs to be discussed later to improve the model prediction performance. In this experiment, data collected from December 2023 to February 2024 were selected from the database, with measurements taken every 15 min, resulting in a total of 7464 data sets.

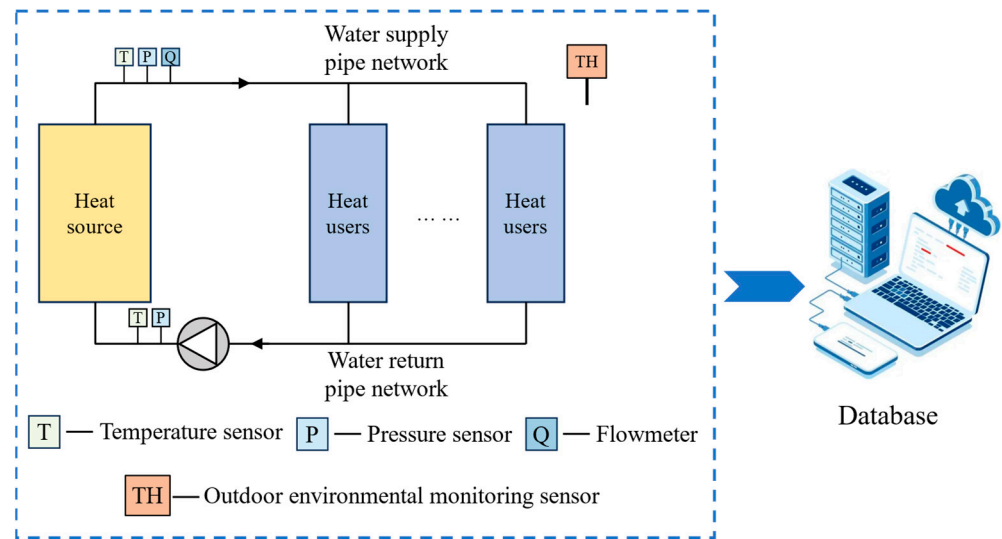


Figure 6. Data acquisition system.

The heating objects of the system include diverse building types, varying heating times, and complex heating characteristics. The heat load is defined as the heat supplied by the heat source system per unit time to maintain the indoor calculated temperature under outdoor meteorological conditions. In this paper, it is assumed that the heating capacity of the system is approximately equal to the heat load, so the heating capacity is used as a proxy for the heat load. The original operational data of the heating objects from 8 February to 14 February 2024, are selected to further analyze the heat load characteristics, as shown in Figure 7.

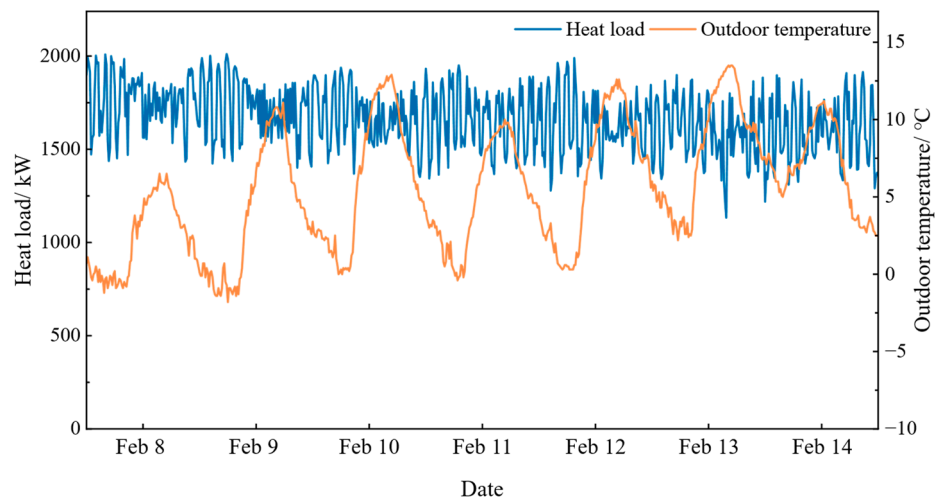


Figure 7. One-week heat load change trend diagram (acquisition cycle is 15 min).

The comprehensive region, which serves both office and residential functions, requires 24-h heating during the heating season. As shown in Figure 7, the general trend of daily heat load changes is evident, generally decreasing as the outdoor temperature increases. The heat load data acquisition and control period of the original control system is 15 min. Due to the large area of the region, the physical distance between the heat users and the heat source is significant, resulting in a large time lag in heat consumption due to the transmission time of the hot water. The original simple feedback control strategy for heating return water, shown in Figure 8, controls the heat source load rate and adjusts the water supply temperature based on the deviation between the actual and set values of the heating return water temperature. However, this strategy struggles to meet the control

requirements of systems with large time lags. When heating demand changes, the system cannot adjust in time to meet the new demand. Additionally, using heating capacity as a proxy for heat load results in large short-term fluctuations due to the influence of the control strategy, significantly complicating accurate load prediction.

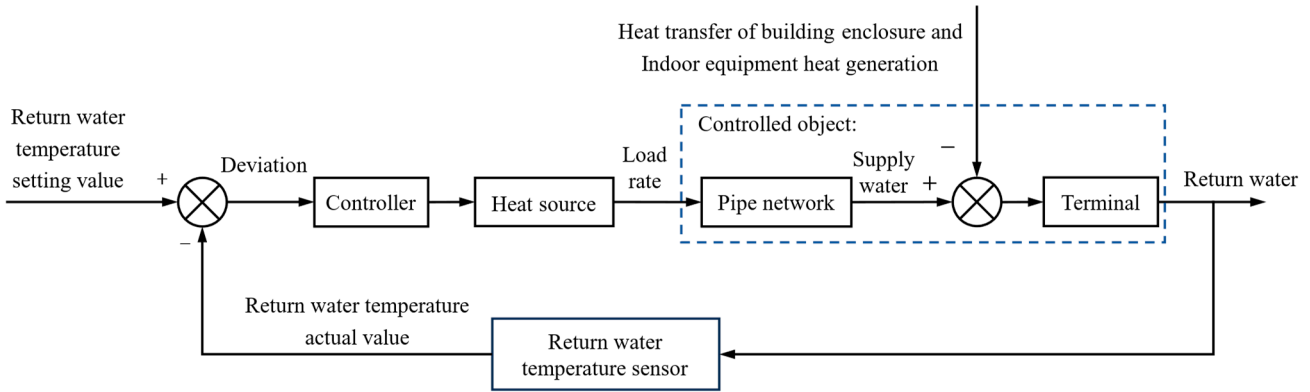


Figure 8. System block diagram.

3.2. Data Preprocessing

Due to data transmission interruptions, data acquisition may be intermittent and incomplete. Data cleaning is a crucial step to ensure data quality and modeling accuracy. In this paper, the 3σ criterion is used to identify outliers. Abnormal zero values and outlier values are deleted and treated as missing values. For single missing values, the average of the observed values at the preceding and following time points is used for interpolation. For multiple consecutive missing values, the date, time, and meteorological attributes of the missing period are used to match similar patterns, and an equal amount of data is selected for filling. Each parameter has different measurement units and ranges, leading to dimensional inconsistency. To eliminate these differences, the maximum–minimum normalization method is employed to normalize the data [31].

3.3. Hysteresis Feature Extraction

3.3.1. Partial Autocorrelation Analysis

The partial autocorrelation function (PACF) is a crucial tool for time series analysis, describing the autocorrelation within a time series. Similar to the autocorrelation function (ACF), PACF measures the correlation between different lag orders in time series data [32]. However, PACF excludes the influence of other lag orders during the calculation, thereby more accurately reflecting the correlation between a specific lag order and the current value. PACF is frequently used to identify and analyze characteristics and trends in time series data and to determine the appropriate lag order for modeling.

For the time series Q_t , the k -order autoregressive model is expressed as follows:

$$Q_t = \varphi_{k1}Q_{t-1} + \varphi_{k2}Q_{t-2} + \dots + \varphi_{kk}Q_{t-k} + a_{kt} \tag{5}$$

where φ_{ki} represents the regression coefficient of the Q_{t-i} ; a_{kt} represents the error term of the autoregressive model, φ_{kk} represents the k -order partial autocorrelation coefficient of time series, $\varphi_{kk}Q_{t-k}$ represents the relationship between Q_t and Q_{t-k} after removing the effects of $Q_{t-1}, Q_{t-2}, \dots, Q_{t-(k-1)}$. The partial autocorrelation coefficients of each order of the time series constitute the PACF, which is calculated by the following formulas:

$$\left\{ \begin{array}{l} \varphi_{11} = \gamma_1, k = 1 \\ \varphi_{kk} = \frac{\gamma_k - \sum_{i=1}^{k-1} \varphi_{k-1,i} \cdot \gamma_{k-i}}{1 - \sum_{i=1}^{k-1} \varphi_{k-1,i} \cdot \gamma_{k-i}}, k = 2, 3, \dots \\ \varphi_{ki} = \varphi_{k-1,i} - \varphi_{kk} \varphi_{k-1,k-i}, i = 1, 2, \dots, k-1 \end{array} \right. \tag{6}$$

where γ_k represents the lagged k -order autocorrelation coefficient.

3.3.2. Heat Load Hysteresis Feature Extraction

This paper analyzes the PACF of the heat load data and selects the 95% confidence interval to determine the significance of the lag order. Specifically, the lag order where the partial autocorrelation coefficient is greater than or equal to 0.05 is selected as the optimal historical period of the heat load sequence, as shown in Figure 9.

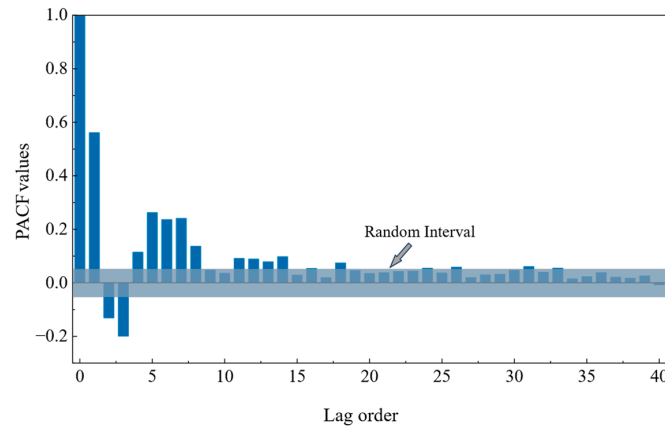


Figure 9. Partial autocorrelation diagram of heat load sequence.

The partial autocorrelation diagram shows that the sample coefficient of the sequence falls within the random interval for the first time after a lag order of 8. However, at a lag order of 11, it deviates from the random interval again, displaying a tailing effect. Considering the characteristics of various data acquisition cycles and other factors, a lag order of 8 is selected. Consequently, the data with a collection frequency of 15 min is downsampled to a 2-h interval. The specific method involves using a 2-h data window and taking the arithmetic average of the eight data points within this window as the characteristic data for that moment. Each data point includes heating pipe network parameters and outdoor meteorological parameters, resulting in the data volume being reduced to one-eighth of the original.

After data downsampling, the prediction period changes from 15 min to 2 h, and the downsampled heat load sequence is shown in Figure 10. This process smooths short-term fluctuations, reduces the impact of instantaneous changes such as noise, and retains significant variations. By focusing on the overall trend of the system, this method enhances the tracking and prediction accuracy of heat load fluctuations. It also meets the requirements of practical engineering applications.

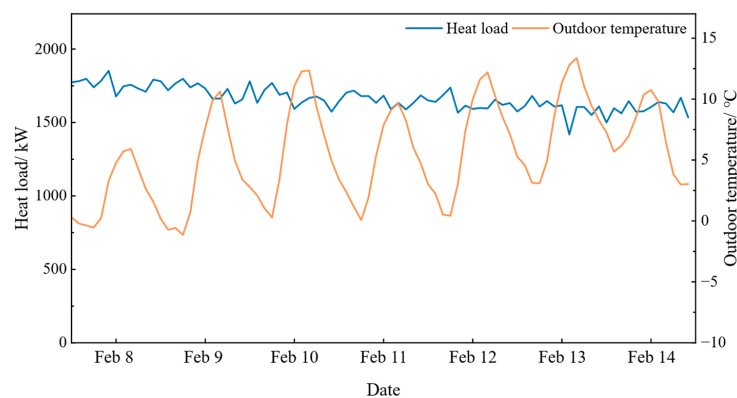


Figure 10. One-week heat load change trend diagram (acquisition cycle is 2 h).

3.4. Analysis of Influencing Factors

Based on the outdoor meteorological data and historical operational data of the heating system during the heating season in the comprehensive area, a correlation analysis of the factors influencing heat load was conducted using a Pearson correlation coefficient. Table 1 shows the degree of influence of each factor on the heat load.

Table 1. The degree of influence of various factors on heat load.

Influence Factors	Correlation Coefficient	Influence Factors	Correlation Coefficient
Heat load at the previous moment	0.863	Pressure difference of the heating pipe at the previous two moments	−0.196
Outdoor temperature at the previous moment	−0.621	Heat load at the previous three moments	0.836
Outdoor relative humidity at the previous moment	0.037	Outdoor temperature at the previous three moments	−0.551
Pressure difference of the heating pipe at the previous moment	−0.194	Pressure difference of the heating main pipe at the previous three moments	−0.191
Heat load at the previous two moments	0.834	Heat load at the previous four moments	0.823
Outdoor temperature at the previous two moments	−0.590	Outdoor temperature at the previous four moments	−0.517
Outdoor relative humidity at the previous two moments	0.018	Pressure difference of the heating main pipe at the previous four moments	−0.189

To avoid an excessive input parameter set, only factors with an absolute correlation coefficient greater than 0.5 were selected as input features for the heat load prediction model. Consequently, the input features for the model are determined as follows: heat load at the previous moment, outdoor temperature at the previous moment, heat load at the previous two moments, outdoor temperature at the previous two moments, heat load at the previous three moments, outdoor temperature at the previous three moments, heat load at the previous four moments, outdoor temperature at the previous four moments. In total, there are eight input features, and the prediction period is 2 h.

Based on the correlation analysis results, a window size of 4 was selected. The data were then reconstructed according to the sequence shape (of samples, timesteps, features) to create a three-dimensional dataset for input into the GRU model. In this configuration, timesteps is 4 and features is 2, as shown in Figure 11.

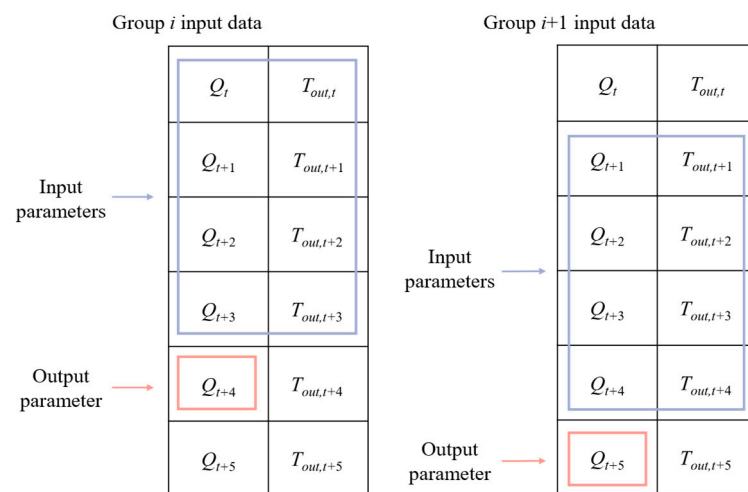


Figure 11. Input and output parameter sliding window selection.

3.5. AE-Based Data Augmentation Method

3.5.1. Autoencoder

AE is an unsupervised learning network architecture used for data dimensionality reduction, feature extraction, and data reconstruction. It consists of a basic encoder and decoder three-layer structure [33]. The encoder compresses the input data into a low-dimensional coding representation, samples the low-dimensional space, and then reconstructs the data sample through the decoder. During the training process, the aim is typically to minimize the reconstruction loss. Data augmentation is performed using AE to generate new synthetic data samples, thereby expanding the original training dataset, enhancing the model’s generalization ability, and reducing the risk of overfitting. The AE model is shown in Figure 12.

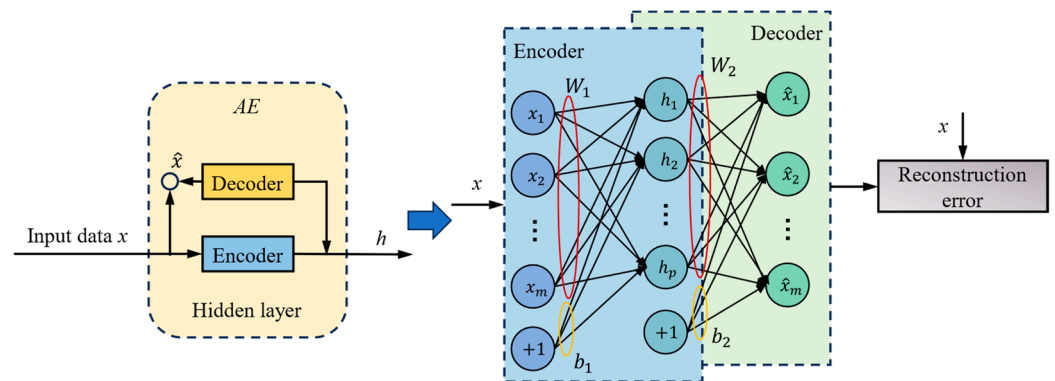


Figure 12. AE model diagram.

In the figure, x_1, x_2, \dots, x_m represents the input data, h_1, h_2, \dots, h_p represents the feature vector, and $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m$ represents the reconstructed data. The encoder uses a nonlinear activation function to obtain the feature vector. The feature vector is reconstructed by the decoder and mapped back to the original data space to obtain the reconstructed data. The update formulas for the AE network are as follows:

$$h = \sigma(W_1x + b_1) \tag{7}$$

$$\hat{x} = \sigma(W_2h + b_2) \tag{8}$$

where W_1 and W_2 represent the weight matrix, b_1 and b_2 represent the bias term, and σ represents the activation function.

3.5.2. Heat Load Time Series Data Augmentation

Due to the limited duration of the system operation, a challenge arises from the insufficient training data available for establishing the GRU prediction model. The AE algorithm exhibits notable efficacy in data augmentation, offering a promising avenue for enhancing the performance of the training model. In this study, the AE algorithm is employed to augment the multi-dimensional heat load time series data. The data set undergoes augmentation at a 1:1 ratio, expanding it through the generation of additional data. Figure 13 illustrates that the trend observed in the original data aligns closely with that of the generated data. A total of 40% of the original dataset is designated as the test set, another 40% serves as the validation set, and the remaining 20%, along with all the augmented data, comprises the training set. This allocation results in an overall ratio of 6:2:2 for the training set, validation set, and test set, respectively, facilitating both the training and testing phases of the prediction model.

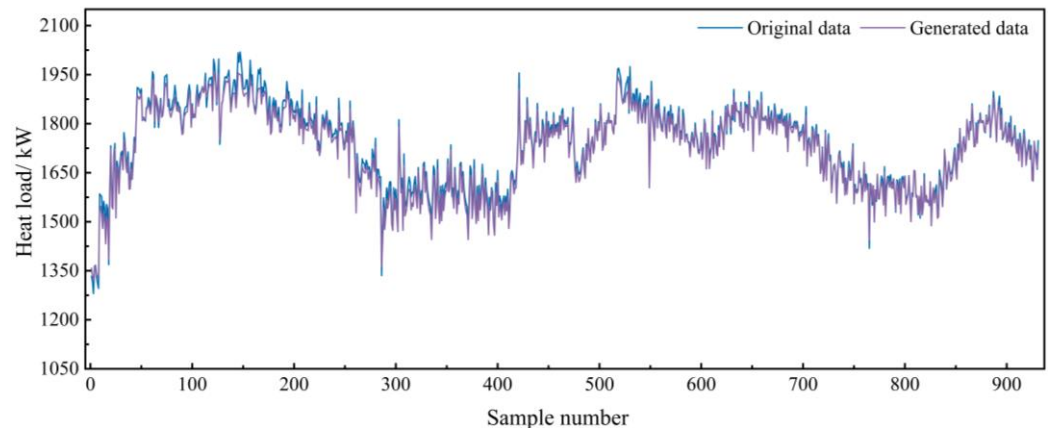


Figure 13. Comparison of original data and generated data.

3.6. Evaluation Indicators

To evaluate the prediction performance of the proposed AE-GWO-GRU model, this paper employs three performance indicators: Mean absolute error (MAE), mean absolute percentage error (MAPE), root mean square error (RMSE), and standard deviation of absolute percentage error (SDAPE). These indicators are calculated using the following formulas. Smaller values of MAE, MAPE, and RMSE signify more accurate prediction results from the model. The smaller the SDAPE value is, the higher is the prediction stability of the model.

$$MAE = \frac{100\%}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

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

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

$$SDAPE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\left| \frac{y_i - \hat{y}_i}{y_i} \right| - MAPE \right)^2} \quad (12)$$

where y_i represents the actual value of the heat load of the i th data sample, \hat{y}_i represents the heat load prediction value of the i th data sample, n represents the total number of predicted samples.

3.7. Results

3.7.1. Model Parameter Optimization

This experiment, implemented in Python 3.10, uses the Keras 2.10 deep learning library within the Anaconda 3 environment for model construction. The GWO-GRU network model comprises five layers: an input layer, two GRU layers, a hidden layer, and an output layer. The Adam algorithm trains the internal network parameters of the GRU. The number of neurons in the GRU layers, the number of neurons in the hidden layer, the learning rate, the dropout rate, and the training batch size in the GRU network serve as the position coordinates of the wolf pack. The parameter ranges are as follows: the number of neurons in the GRU and hidden layers ranges from 16 to 256, the learning rate ranges from 1×10^{-4} to 1×10^{-1} , the dropout rate ranges from 0.2 to 0.5, and the training batch size ranges from 8 to 128. The initial parameters of the grey wolf optimization algorithm are set as follows: the population size of the wolf group is 30, the search space dimensionality is 6, the maximum number of iterations is 200, and the initial positions of the grey wolves are randomly generated.

The iterative evolution process of the AE-GWO-GRU model during training is shown in Figure 14. As the number of iterations increases, the fitness value gradually decreases until it stabilizes, achieving convergence after the 125th iteration. The optimized hyperparameters are detailed in Table 2, while the network structure is shown in Figure 15.

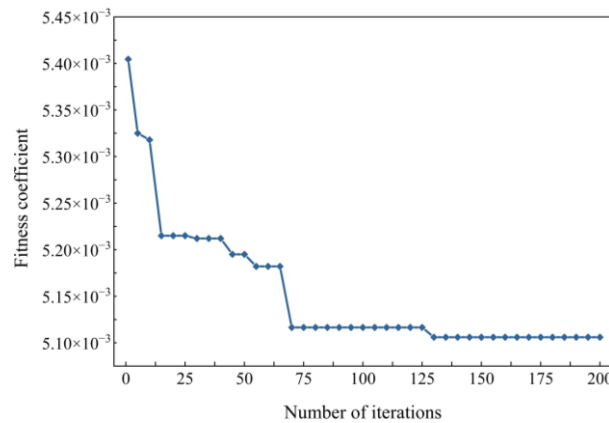


Figure 14. GWO iterative fitness curve.

Table 2. The value of GRU model parameters.

Parameter	Value Ranges	Optimal Value
Number of neurons in the first layer GRU layer	16~256	121
Number of neurons in the second layer GRU layer	16~256	156
Dropout rate	0.2~0.5	0.2
Number of neurons in the hidden layer	16~256	71
Learning rate	$1 \times 10^{-4} \sim 1 \times 10^{-1}$	0.007
Batch size	8~128	24

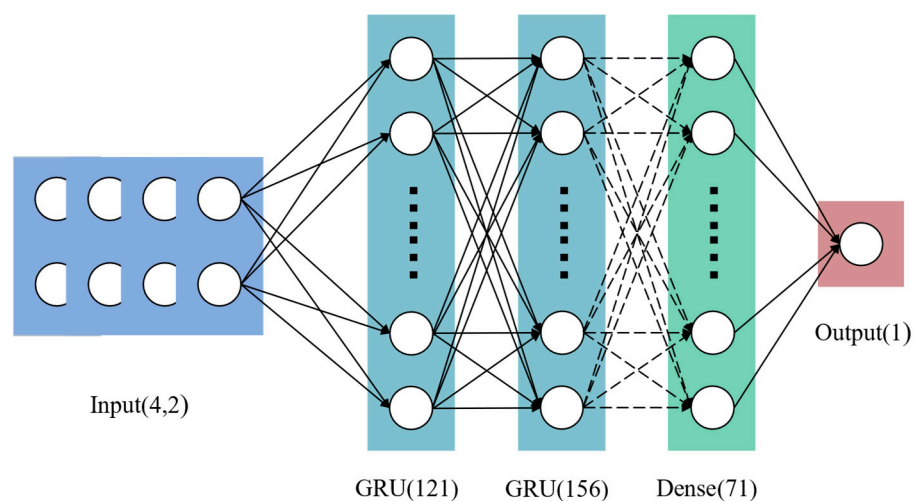


Figure 15. GRU model network structure.

The GRU model includes an input layer, a first GRU layer with 121 neurons, a second GRU layer with 156 neurons, a hidden layer with 71 neurons, and an output layer, as shown in Figure 15. The dotted line represents the dropout strategy.

To evaluate the performance of the AE-GWO-GRU prediction model proposed in this paper, four models were selected for comparison: standard GRU, standard BP, AE-GRU,

and AE-BP. The parameters, including the number of neurons and the learning rate of the GRU prediction model, are kept consistent with those of the BP prediction model. Additionally, the input and output parameters remain the same across all models.

3.7.2. Heat Load Prediction Based on AE-GWO-GRU

The GWO algorithm is employed to optimize the GRU model, resulting in the determination of optimal parameters: 121 neurons for the first GRU layer, 156 neurons for the second GRU layer, 71 neurons for the hidden layer, a learning rate of 0.007, a dropout rate of 0.2 and a training batch size of 24. The GRU model is retrained using these optimal parameters, and heat load prediction is conducted alongside comparison models. The relative errors of the predictions from different models are shown in Figure 16, and the MAPE and SDAPE are shown in Figure 17.

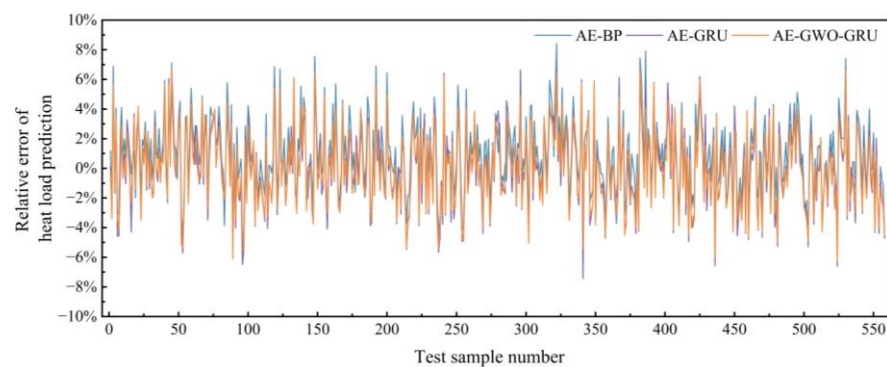


Figure 16. The relative errors of heat load prediction of different models.

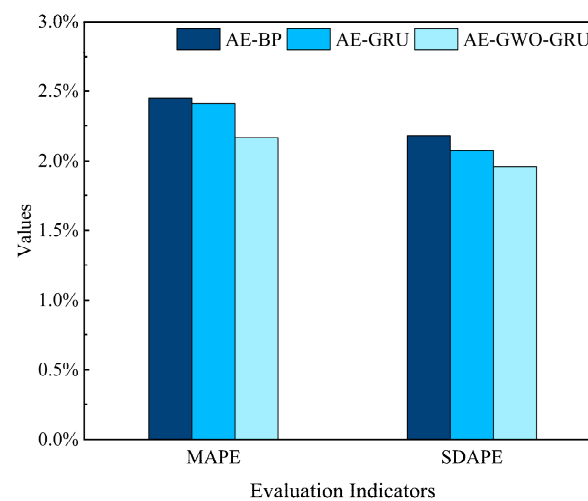


Figure 17. The MAPE and SDAPE of different heat load prediction models.

Through experimental analysis, the prediction errors of AE-BP and AE-GRU are relatively large, while the prediction error of AE-GWO-GRU is the smallest, with maximum relative errors of 8.40%, 7.43%, and 6.73%, respectively. The AE-GWO-GRU model has the smallest SDAPE, indicating that the prediction errors are more concentrated. This demonstrates the model's superior stability and better generalization ability. To further validate the model's prediction accuracy, the results of 24 test samples are examined. The prediction curves of different models, including AE-BP, AE-GRU, and AE-GWO-GRU, are shown in Figure 18. The accuracy analysis values of heat load prediction results under different models are presented in Table 3.

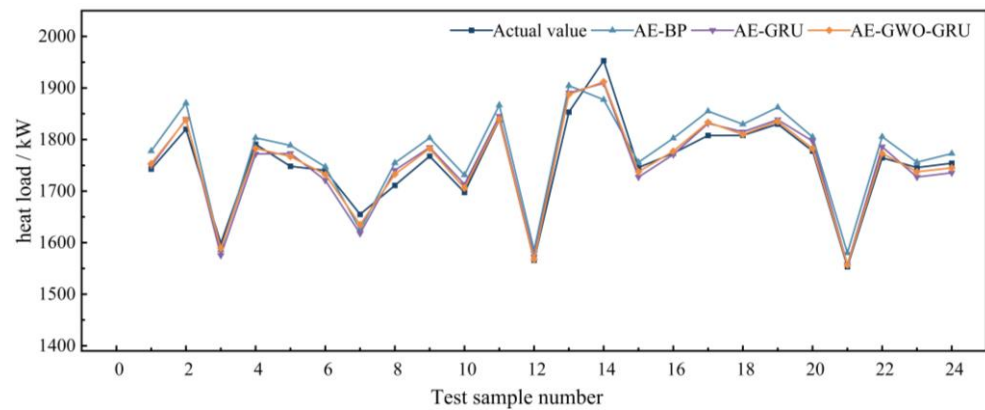


Figure 18. The prediction results of different models.

Table 3. Comparison results of different models.

Model	RMSE	MAE	MAPE	SDAPE
BP	57.86	42.75	2.51%	2.45%
AE-BP	54.83	41.46	2.45%	2.18%
GRU	56.69	41.77	2.45%	2.38%
AE-GRU	53.98	41.18	2.41%	2.07%
AE-GWO-GRU	47.90	36.90	2.17%	1.96%

Comparing the prediction results, the AE-GWO-GRU model yields values closest to the actual data, demonstrating the best predictive performance, as shown in Figure 18. Compared with the BP, AE-BP, GRU, and AE-GRU models, the AE-GWO-GRU model exhibits lower error values in terms of RMSE, MAE, and MAPE, with values of 47.90, 36.90, and 2.17%, respectively. This demonstrates the superiority of the AE-GWO-GRU model, achieving the expected effectiveness.

As shown in Table 3, a significant error exists between the predicted and actual values before using AE for data augmentation. After data augmentation, the errors of the AE-BP and AE-GRU models are reduced. Specifically, the RMSE and MAE of the BP model decrease by 3.03 and 1.29, respectively, while those of the GRU model decrease by 2.71 and 0.59, respectively. The SDAPE of BP and GRU both decrease after data augmentation. This demonstrates that utilizing AE to augment data in scenarios with insufficient data volume effectively reduces the prediction error and enhances the prediction stability. Moreover, compared with AE-GRU, the AE-GWO-GRU model's predicted values are closer to the real values, resulting in a higher prediction accuracy. The RMSE and MAE are reduced by 6.08 and 4.28, respectively, indicating that using GWO to optimize GRU parameters enables adjustments that best match the heat load data, thereby further enhancing prediction accuracy.

4. Conclusions

To improve the accuracy of heat load prediction in DHSs, a novel AE-GWO-GRU model is proposed, integrating an AE network, the GWO algorithm, and a GRU network. By optimizing GRU model parameters and establishing the prediction model, along with a comprehensive case study, the following conclusions are drawn:

- (1) The AE model augments data derived from a limited number of original samples, ensuring an adequate volume of samples for training the GRU model. This augmentation process significantly enhances the model's prediction accuracy and stability.
- (2) The GWO algorithm tunes the GRU model parameters, addressing issues such as inadequate model fitting, low prediction accuracy, and prolonged parameter adjustment times associated with manual selection. Compared to BP, AE-BP, GRU, and AE-GRU models, the AE-GWO-GRU model demonstrates superior prediction accuracy.

In summary, heat load prediction enables the estimation of future heat load values, serving as a fundamental component for optimal heat control strategy research. This predictive capability aids control systems in adjusting the water supply flow or temperature based on anticipated changes in the heat load, facilitating on-demand heating, and promoting energy efficiency and emission reduction efforts.

Author Contributions: Conceptualization, Y.Y. and J.Y.; methodology, Y.Y.; software, Y.Y.; validation, Y.Y.; formal analysis, X.Z. and Y.Y.; investigation, Y.Y. and J.Y.; resources, J.Y.; data curation, Y.Y. and J.Y.; writing—original draft preparation, J.Y., X.Z. and Y.Y.; writing—review and editing, X.Z. and Y.Y.; visualization, X.Z. and Y.Y.; supervision, J.Y. and X.Z.; project administration, J.Y.; funding acquisition, J.Y. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Guangdong Province, grant number 2022A1515011128.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the confidentiality requirements of the project, the research data cannot be disclosed.

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

References

- Henchoz, S.; Chatelan, P.; Maréchal, F.; Favrat, D. Key energy and technological aspects of three innovative concepts of district energy networks. *Energy* **2016**, *117*, 465–477. [CrossRef]
- Comprehensive Finance Department of the Ministry of Housing and Urban-Rural Development. *2022 Urban and Rural Construction Statistical Yearbook*, 1st ed.; China Architecture & Building Press: Beijing, China, 2022. (In Chinese)
- China Building Energy Conservation Association. 2019 China Building Energy Consumption Research Report. *Building* **2020**, *67*, 30–39. (In Chinese)
- President Xi Voices Confirmation in Implementing Paris Agreement, Improving Global Climate Governance. Available online: https://english.www.gov.cn/news/topnews/202012/13/content_WS5fd56f5dc6d0f72576941cbb.html (accessed on 20 April 2024).
- Tan, X.; Zhu, Z.J.; Sun, G.X.; Wu, L.F. Room thermal load prediction based on analytic hierarchy process and back-propagation neural networks. *Build. Simul.* **2022**, *15*, 1989–2002. [CrossRef]
- Yang, R.; Liu, H.; Nikitas, N.; Duan, Z.; Li, Y.; Li, Y. Short-term wind speed forecasting using deep reinforcement learning with improved multiple error correction approach. *Energy* **2022**, *239*, 122128. [CrossRef]
- Vogler-Finck, P.J.C.; Bacher, P.; Madsen, H. Online short-term forecast of greenhouse heat load using a weather forecast service. *Appl. Energy* **2017**, *205*, 1298–1310. [CrossRef]
- Hua, P.M.; Wang, H.C.; Xie, Z.C.; Lahdelma, R. District heating load patterns and short-term forecasting for buildings and city level. *Energy* **2024**, *289*, 129866. [CrossRef]
- Dahl, M.; Brun, A.; Andresen, G.B. Using ensemble weather predictions in district heating operation and load forecasting. *Appl. Energy* **2017**, *193*, 455–465. [CrossRef]
- Sarwar, R.; Cho, H.; Cox, S.J.; Mago, P.J.; Luck, R. Field validation study of a time and temperature indexed autoregressive with exogenous (ARX) model for building thermal load prediction. *Energy* **2017**, *119*, 483–496. [CrossRef]
- Li, Z.; Huang, G. Re-evaluation of building cooling load prediction models for use in humid subtropical area. *Energy Build.* **2013**, *62*, 442–449. [CrossRef]
- Fang, T.T.; Lahdelma, R. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Appl. Energy* **2016**, *179*, 544–552. [CrossRef]
- Xue, G.; Zhang, Y.; Yu, S.-a.; Song, J.; Bian, T.; Gao, Y.; Yan, W.; Guo, Y. Daily residential heat load prediction based on a hybrid model of signal processing, econometric model, and support vector regression. *Therm. Sci. Eng. Prog.* **2023**, *43*, 102005. [CrossRef]
- Zhou, Y.; Liu, Y.F.; Wang, D.J.; Liu, X.J. Comparison of machine-learning models for predicting short-term building heating load using operational parameters. *Energy Build.* **2021**, *253*, 111505. [CrossRef]
- Küçüktopcu, E. Comparative Analysis of Data-Driven Techniques to Predict Heating and Cooling Energy Requirements of Poultry Buildings. *Buildings* **2023**, *13*, 142. [CrossRef]
- Huang, S.; Zuo, W.D.; Sohn, M.D. A Bayesian Network model for predicting cooling load of commercial buildings. *Build. Simul.* **2018**, *11*, 87–101. [CrossRef]
- Li, Q.; Zhao, F. Improved BP neural network of heat load forecasting based on temperature and date type. *J. Syst. Simul.* **2018**, *30*, 1464–1472. [CrossRef]

18. Xu, Y.; Li, F.; Asgari, A. Prediction and optimization of heating and cooling loads in a residential building based on multi-layer perceptron neural network and different optimization algorithms. *Energy* **2022**, *240*, 122692. [[CrossRef](#)]
19. Yuan, T.L.; Jiang, D.S.; Huang, S.Y.; Hsu, Y.Y.; Yeh, H.C.; Huang, M.N.L.; Lu, C.N. Recurrent Neural Network Based Short-Term Load Forecast with Spline Bases and Real-Time Adaptation. *Appl. Sci.* **2021**, *11*, 5930. [[CrossRef](#)]
20. Cui, M.S. District heating load prediction algorithm based on bidirectional long short-term memory network model. *Energy* **2022**, *254*, 124283. [[CrossRef](#)]
21. Jung, S.; Moon, J.; Park, S.; Hwang, E. An Attention-Based Multilayer GRU Model for Multistep-Ahead Short-Term Load Forecasting. *Sensors* **2021**, *21*, 1639. [[CrossRef](#)]
22. Xu, J.Q.; Bao, K.Q.; Cai, Z.P.; Tang, H. Heat load prediction based on GRU unit and MC error correction. In Proceedings of the 2022 4th International Conference on Electrical Engineering and Control Technologies (CEEECT 2022), Shanghai, China, 16–18 December 2022; pp. 364–368.
23. Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey wolf optimizer. *Adv. Eng. Softw.* **2014**, *69*, 46–61. [[CrossRef](#)]
24. Jamshidi, V.; Nekoukar, V.; Refan, M.H. Real time UAV path planning by parallel grey wolf optimization with align coefficient on CAN bus. *Cluster Comput.* **2021**, *24*, 2495–2509. [[CrossRef](#)]
25. Du, M.M.; Zhao, Y.Q.; Liu, C.J.; Zhu, Z. Lifecycle cost forecast of 110 kV power transformers based on support vector regression and gray wolf optimization. *Alex. Eng. J.* **2021**, *60*, 5393–5399. [[CrossRef](#)]
26. Ge, L.J.; Xian, Y.M.; Wang, Z.G.; Gao, B.; Chi, F.J.; Sun, K. Short-term Load Forecasting of Regional Distribution Network Based on Generalized Regression Neural Network Optimized by Grey Wolf Optimization Algorithm. *CSEE J. Power Energy Syst.* **2021**, *7*, 1093–1101. [[CrossRef](#)]
27. Nakhwan, M.; Duangsoithong, R. Comparison Analysis of Data Augmentation using Bootstrap, GANs and Autoencoder. In Proceedings of the 2022-14th International Conference on Knowledge and Smart Technology (KST 2022), Krabi, Thailand, 28 February 2024; pp. 18–23.
28. Lee, H.; Kim, C.; Jeong, D.H.; Lee, J.M. Data-driven fault detection for chemical processes using autoencoder with data augmentation. *Korean J. Chem. Eng.* **2021**, *38*, 2406–2422. [[CrossRef](#)]
29. Li, L.; Zhang, Z.; Yang, L. Influence of Autoencoder-Based Data Augmentation on Deep Learning-Based Wireless Communication. *IEEE Wirel. Commun. Lett.* **2021**, *10*, 2090–2093. [[CrossRef](#)]
30. Guo, Z.; Yang, C.; Wang, D.; Liu, H. A novel deep learning model integrating CNN and GRU to predict particulate matter concentrations. *Process Saf. Environ. Protect.* **2023**, *173*, 604–613. [[CrossRef](#)]
31. Karaman, Ö.A. Prediction of Wind Power with Machine Learning Models. *Appl. Sci.* **2023**, *13*, 11455. [[CrossRef](#)]
32. Khanarsa, P.; Sinapiromsaran, K. Automatic SARIMA Order Identification Convolutional Neural Network. *Int. J. Mach. Learn. Comput* **2020**, *10*, 685–691. [[CrossRef](#)]
33. Yang, L.X.; Zhang, Z.J. A Conditional Convolutional Autoencoder-Based Method for Monitoring Wind Turbine Blade Breakages. *IEEE Trans. Ind. Inform.* **2021**, *17*, 6390–6398. [[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.