Energy 278 (2023) 127911

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

A multivariate time series graph neural network for district heat load forecasting

Zhijin Wang [a,*], Xiufeng Liu [b,*], Yaohui Huang [c], Peisong Zhang [d], Yonggang Fu [a]

[a] College of Computer Engineering, Jimei University, Yinjiang Road 185, 361021 Xiamen, China
[b] Department of Technology, Management and Economics, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark
[c] College of Electronic Information, Guangxi Minzu University, Daxue East Road 188, 530006 Nanning, China
[d] School of Science, Jimei University, Yinjiang Road 185, 361021 Xiamen, China

ARTICLE INFO

Keywords:
District heating
Prediction
Graph neural network
Multivariate time series
Meteorological factors

ABSTRACT

Heat load prediction is essential for energy efficiency and carbon reduction in district heating systems. However, heat load is influenced by many factors, such as building characteristics, consumption behavior, and climate, making its prediction challenging. Traditional methods based on physical models are complex and insufficiently accurate, whereas most data-driven statistical methods ignore customer energy consumption behaviors and their correlation, and do not account for the temporal inertia of consumption. This paper proposes a graph ambient intelligence (GAIN) method for heat load prediction, which classifies customers based on their load profiles and uses collaborative attention on temporal graphs to capture their associations and the weather impact on heat loads. GAIN also incorporates recursive and autoregressive methods to model the temporal inertia of consumption. The proposed method is evaluated on four metrics and compared with fifteen baseline methods. The results show that GAIN achieves the lowest daily forecasting errors in terms of RMSE, MAE, and CV-RMSE, with values of 6.972, 4.442, and 0.191, respectively. Besides, the proposed method achieves a maximum reduction of 25%, 29%, and 25% in RMSE, MAE, and CV-RMSE, respectively, compared to other methods when taking meteorological factors into account.

1. Introduction

Today, the energy crisis and environmental issues have become global concerns. The building sector accounts for 30%–40% of energy consumption worldwide, and in Europe, it accounts for 40%–50%, with thermal energy demand comprising 13% of this figure [1]. District heating systems have been strongly advocated in numerous countries, becoming essential in achieving energy-saving and emission-reduction targets. These targets include the European Union's net-zero carbon emission goal by 2050 [2] and China's aim to reach peak carbon emissions by 2030 and carbon neutrality by 2060 [3].

District heating systems serve as essential infrastructure components, generating and supplying heat to buildings in urban areas by employing a diverse range of energy sources, such as biomass, natural gas, solar energy, industrial waste heat, and geothermal energy. Geothermal energy, as a renewable and clean source of heat and power, can be harnessed to drive combined heat and power (CHP) systems for various applications, including district heating. This results in improved overall system efficiency by utilizing combined heat and power [4], which can contribute to reduced emissions and increased energy security in urban areas. The main advantage of district heating systems lies in their ability to harness combined heat and power, thus enhancing the overall efficiency of the system [5]. Fig. 1 illustrates a simplified schematic of a district heating system, which consists of a CHP plant, supply and return pipes, and buildings. Heat from the CHP plant is distributed to different end-users, such as residential households, via supply pipes, while water is returned to the CHP plant for reheating through the return pipes. The difference between the supply and return water generally reflects the efficiency of the heating devices within a building. The heating network typically comprises a hierarchical structure, including primary and secondary networks, and heat exchangers facilitate heat transfer between the two networks [6]. In most existing district heating systems, the heating temperature and supply are manually controlled based on outdoor temperatures and subjective human judgment or experience, which can lead to considerable heat waste, especially during sudden changes in weather conditions or the emergence of new heat demands.

* Corresponding authors.
E-mail addresses: zhijinecnu@gmail.com (Z. Wang), xiuli@dtu.dk (X. Liu), yhhuang5212@gmail.com (Y. Huang), pencil007123@gmail.com (P. Zhang), yonggangfu@jmu.edu.cn (Y. Fu).

https://doi.org/10.1016/j.energy.2023.127911
Received 25 November 2022; Received in revised form 21 April 2023; Accepted 21 May 2023
Available online 25 May 2023

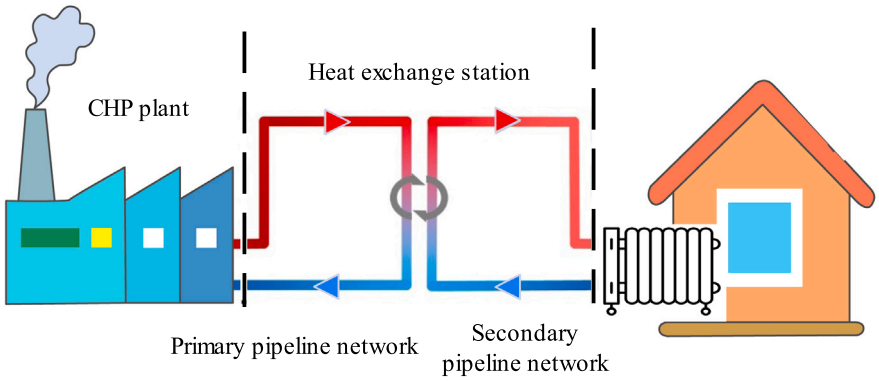


Fig. 1. Schematic diagram of district heating system.

For district heating systems, optimal control should aim for efficient operation, which can only be achieved by matching heat production to the actual demand of end users [7]. Estimating heat load is crucial and is usually a prerequisite for the operations of a district heating system. In particular, short-term predictions, ranging from a few hours to a few days, are used to predict the heat load. Accurate forecasting allows heat production (i.e., heating temperature and flow rate) to be matched to actual demand, thereby supplying heat on demand, reducing production costs and distribution losses, and lowering return temperatures. This is particularly important for CHP plants to improve coordination between the plant and the grid [8]. In recent years, with the digitization of energy systems, intelligent systems have realized fine-grained control over major thermal devices, which help to provision energy supply and improve energy efficiency. For example, Adamski et al. [9] propose a predictive control system that achieves 20.4% thermal energy savings in the studied building compared to the traditional weather-based control method. Kapalo et al. [10] point out that an accurate prediction of heat load is essential for the district heating system in optimizing the supply and demand structure, and also to improve automatic boiler control systems, such as timely and efficient control of each unit to ensure efficiency and reliability. In addition, accurate load forecasting is essential for demand-side management. Recent studies [11] show that with the implementation of demand-side management, including short-term demand forecasting (2–3 h), the thermal load of individual buildings can be reduced by an average of 25%. Therefore, forecasting should be a key component for district heating management.

Heat load prediction is at the core of the district heating system, and accurate prediction plays a key role in energy efficiency and carbon reduction [12]. However, heat load can be affected by many factors, such as the physical characteristics of buildings, consumption behavior, and climate, and its prediction is a significant challenge [13]. Traditional prediction methods based on thermodynamic models are complex and insufficiently accurate, whereas most data-driven statistical methods ignore the temporal inertia of consumption itself, nor do they capture the potential associations between similar customers or the impact of weather factors on heat loads. All of these limit the prediction capability. To address these issues, this paper proposes a deep learning-based graph ambient intelligence (GAIN) method for heat load prediction. Graph neural networks (GNNs) are a class of deep learning models that can learn from graph-structured data, such as molecules, knowledge and social networks [14]. However, there is still a big lack of application and research of GNNs for time series prediction, especially for district heat load forecasting. GNNs are suitable for this problem because they can model the complex relationships between customers, heat load and meteorological factors in a graph representation. Moreover, they can overcome the limitations of existing methods by capturing both local and global patterns in time series data. The main contributions and significance of this paper are as follows:

- We propose a novel GAIN model for district heat load prediction. This model takes into account the intrinsic correlation between customers and the causal relationship between time steps in a time series.
- To integrate the external factors affecting the prediction performance, we design another neural network branch for extracting patterns from meteorological external factors and a collaborative temporal graph attention to fuse heat load and meteorological observations. This novel fusion mechanism has a powerful feature selection capability that leads to a significant improvement in accuracy.
- We conduct a comprehensive evaluation for the proposed model over four prediction time horizons using a real district heating dataset with correspondent meteorological data. The results show the efficiency and desirable performance of the proposed model, which outperforms other state-of-the-art methods by 25%, 29%, and 25% in terms of RMSE, MAE, and CV-RMSE, respectively.

The remainder of this paper is organized as follows. Section 2 discusses related work. Section 3 describes the dataset. Section 4 presents the proposed model. Section 5 conducts the experiments to evaluate the model. Section 6 concludes the paper and presents future work.

2. Related work

2.1. Physical models for heat load prediction

Research themes on heat load prediction can be divided into two categories: thermodynamic and data-driven models. Thermodynamic modeling is a traditional approach, also known as “white box” modeling, which relies on complex mathematical construction methods. It requires a large number of physical parameters for heating, such as flow rate, temperature, and volume, as well as building parameters, such as material, area, and orientation. For example, Zhang et al. [15] estimate the heat load of a residential building using the white-box model that takes into account the parameters including indoor climate, building characteristics, and solar radiation. The model is calibrated using the particle swarm optimization (PSO) algorithm to search for the optimal combination of the parameters in the heat load estimation. The second physical model is also called “gray box” model, which often used to address the difficulty in determining optimal parameters for physical models [16]. Therefore, grey-box models typically use physical knowledge to define the model structure and then use historical data to estimate model parameters [17]. For example, Thilker et al. [18] create a non-linear gray box model to estimate the heat demand of a water-heated school building in Denmark, where meteorological weather observations were used as input. However, a significant limitation of the above physical models is their sensitivity to the physical building parameters [19], which can render them less robust and necessitate more frequent recalibration or updates.

2.2. Feature selection for data-driven heat load prediction models

Data-driven models refer to machine learning or deep learning-based models, which require a large amount of data for training [20]. In addition to historical heat load data, meteorological data is the most used auxiliary dataset for heat load prediction [21]. The selection of appropriate exogenous input variables is crucial to the accuracy of heat load prediction, which has been extensively studied. Among others, Song et al. [22] conducted a Pearson correlation study on different available variables, and selected the positively correlated variables, heat supply, and return temperatures; and the negatively correlated outdoor temperature as features for training the deep learning model. Gong et al. [6] used Pearson and least absolute shrinkage and selection operator (LASSO) methods to optimize the feature set, including system parameters, meteorological parameters, and time steps; and resulted in four feature sets, a total of 28 features as the final input. The meteorological parameters, including outdoor temperature, solar radiation intensity, wind speed, and humidity, have been widely used for studying the impact on heat load (e.g., [23,24]). The early study [25] shows that the outside temperature can affect the heat load of a building by 60%, the wind speed by 1% to 4%, and the solar radiation intensity by 1% to 5%. Moreover, the study in [26] reveals that the higher the intensity of solar radiation, the greater the indoor temperature. Therefore, the intensity of solar radiation can affect the heat consumption for those heat users with control equipment. In our study, we use the available meteorological data to perform correlation analysis and select four meteorological parameters as additional features of our model. Additionally, we evaluate the individual contribution to the model performance through a feature ablation study.

2.3. Data-driven heat load prediction models

Currently, significant efforts have been made to data-driven heat load prediction, mainly due to their excellence in non-linear modeling capabilities and the availability of fine-grained heat load smart meter data [27]. The most popular data-driven models for heat load predictions can be grouped into three categories: traditional statistical methods, classical machine learning methods, and advanced machine learning methods.

Traditional statistical methods have gained widely employed in predicting heat load due to their high interpretability and computational efficiency [28]. Bujalski et al. [29] proposed a data-driven model for hourly heat load predictions based on the Generalized Additive Model during the off-season. Jagait et al. [30] include autoregressive integrated moving average (ARIMA) for online load prediction and alleviate the concept drift problem. However, most of these methods rely on stringent assumptions that may not be feasible in practical scenarios [20], consequently leading to worse performance.

Classical machine learning methods have superiority in capturing non-linear relationships over traditional statistical methods. Cui [31] verify the effectiveness of bidirectional long short-term memory network (BiLSTM) for short-term heat load forecasting. Zhao et al. [17] proposed a residential district heating prediction model based on an improved convolution neural network (CNN). Ding et al. [32] designed multi-input artificial neural network, which achieve the well performance in long and short-term heat load prediction. Song et al. [33] proposed a model combining a convolutional neural network and a long short-term memory algorithm, namely CNN-LSTM, for heat load prediction. They found that the combined model can achieve better prediction accuracy in district heating systems with thermal inertia problems than others, including SVM and ensemble learning algorithms. However, due to the limitation in capturing the global and long-term dependencies, these may not achieve promising accuracy in complex scenarios.

Advanced machine learning methods, such as transformer and graph neural networks, possess a remarkable ability to capture complex associations and have been highly effective in energy prediction tasks [24]. Gong et al. [34] have designed a novel framework for heating load forecasting based on the Informer architecture, which is a variant of the Transformer that employs self-attention mechanisms to capture long-term dependencies [35]. Wang et al. [36] proposed a multi-task multi-energy load forecasting model based on Decoder and Transformer component. This model further verifies the effectiveness and generalization capability of self-attention mechanisms. Hu et al. [37] put forward a spatiotemporal graph convolutional network to track building energy consumption. The results demonstrate that the GNN enables modeling the interdependency of features and capturing dynamic nonlinear representations. In [38], an improved graph neural network is employed for controlling wind energy and shows good performance. However, most advanced machine learning methods neglect the diversity of customer behavior patterns in energy consumption prediction, which can limit the accuracy of prediction models. Moreover, the potential of graph neural networks with attention mechanisms in heat load prediction has not yet been verified.

Most of the above studies are considered conventional methods for heat load prediction, and some of them have applied the current deep neural network-based approaches, which aim to improve the prediction performance. However, most of these methods ignore the customer energy consumption behaviors and their correlation and do not account for the temporal inertia of consumption itself. In contrast, in this study, we design a GNN-based model structure that captures the intrinsic correlation between customer, heat load, and meteorological data, as well as the temporal dependencies in time series. Moreover, few studies have considered the economic analysis of the district heating system and its integration with solar desalination. A comprehensive framework for assessing the economic feasibility and sustainability of water-related interventions, including desalination, has been proposed by [39,40], which considers the costs and benefits from multiple perspectives such as financial, environmental, social, and institutional.

3. Study materials

The data used in this study include district heating consumption data, related building registration data, and meteorological data. The heating consumption data were from Danish residential buildings in Aalborg [41], which are publicly available on the Zenodo repository at https://doi.org/10.5281/zenodo.6563114. The original dataset consists of data from 3127 smart heat meters for the period 2018-01-01 to 2020-12-31, with hourly resolution. The data were anonymized, cleaned, and provided as comma-separated CSV files. The background data includes dwelling type, building year of construction, and energy level. Heat load data were collected for dwelling types including apartments, townhouses, single-family houses, and nonresidential buildings, numbered 88, 474, 2460, and 8. In this study, we focused on heat load prediction for single-family houses. The meteorological data were obtained from https://confluence.govcloud.dk, which were collected from weather stations located near the district heating areas in Aalborg.

The heat load of a building can be affected by a variety of factors, including customer activities, indoor and outdoor climate, and building characteristics. Therefore, in a data-driven model, it is necessary to include appropriate external variables as model inputs in addition to historical heat load data to improve accuracy. As mentioned earlier, considerable research has considered meteorological factors for improving heat load prediction accuracy. Similarly, in this study, we select appropriate meteorological variables based on the following analysis. We first conduct a correlation analysis for the nine meteorological factors in the dataset and obtain the results shown in Fig. 2.

As can be seen, grass temperature and outdoor temperature are two temperature features that demonstrate a high correlation with heat load observations and a strong inter-correlation with each other. To reduce the redundancy of temperature information in the predictive model, we choose the latter in our study. This is reinforced by prior studies that highlight outdoor temperature as the foremost meteorological

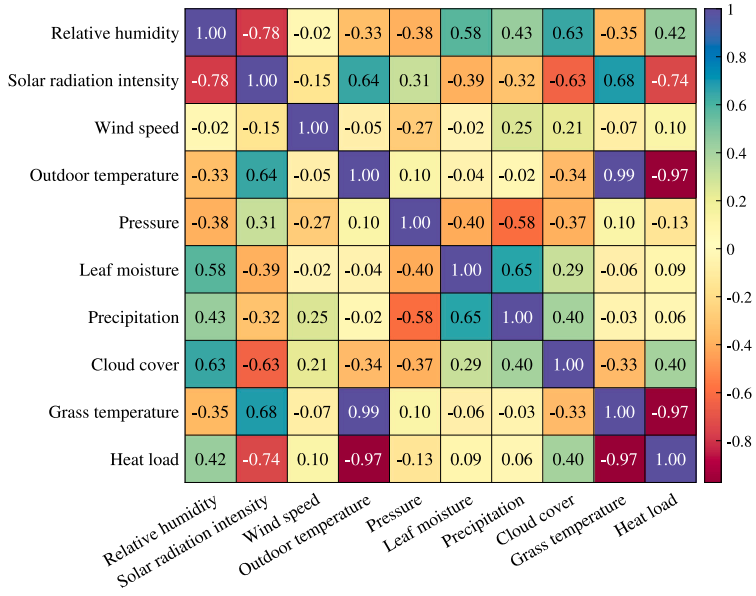


Fig. 2. The Pearson correlation coefficient between meteorological factors and heat load.

factor impacting building heating loads, e.g., [42,43]. Besides, solar radiation intensity is also a critical factor that can impact the indoor environment, e.g., if the building has an opaque envelope, the solar radiation can increase the building surface temperature, thus reducing heat loss; if the building has a translucent envelope, the sunlight entering the interior can increase the indoor temperature, thus reducing the heating load demand [43]. Therefore, we have included solar radiation intensity as a meteorological factor in our study. In addition, Fig. 2 highlights a robust correlation between relative humidity and cloud cover, but these two factors exhibit a relatively minor impact on heat load. Since the well-established connection between relative humidity and building heat transfer [44], we have chosen relative humidity as the third meteorological factor. The wind speed, air pressure, leaf moisture, and precipitation have weak correlations with heat load data. To improve the diversity of meteorological features without being overly inclusive, we have also included wind speed as a factor, as it is relatively less correlated with the other three selected factors.

Therefore, outdoor temperature, relative humidity, solar radiation intensity, and wind speed have been chosen as the final additional features for our modeling. This selection is appropriate because while additional variables can typically improve model accuracy, they can also increase model complexity, potentially leading to overfitting.

Fig. 3 shows four selected outdoor meteorological parameters, and Fig. 4 shows the percentiles of the heat load for the entire year 2018, respectively. From the overall pattern of heat consumption, we can visually observe a negative correlation between outdoor temperature (and the correlated solar radiation intensity) and heat load. In other words, a higher outdoor weather temperature corresponds to a lower heat load, for example in summer, while a lower outdoor temperature corresponds to a higher heat load, for example in winter. As for the other two factors, wind speed and humidity, a slight correlation with the heat load can be visually observed. The quantitative results of the correlation studies between the heat load and the four meteorological parameters, as well as the basic statistics of the heat load and meteorological data are presented in Table 1.

4. Methodology

This section formulates the problem definition of heat load prediction and illustrates the detailed techniques of the proposed GAIN. The main notations used in this paper are listed in Table 2.

4.1. Problem formulation

The prediction problem can be formulated using historical heat load observations, and exogenous observations, meteorological data, as the input to predict the future load with a specific time horizon. The district heat load observations are denoted by $\boldsymbol{Y} = \{Y_1, Y_2, \dots, Y_L\} \in \mathbb{R}^{L\times D'}$, where L is the number of time steps and D' is the number of districts. The meteorological observations are represented as $\boldsymbol{M} = \{M_1, M_2, \dots, M_L\} \in \mathbb{R}^{L\times D^{\#}}$, where $D^{\#}$ is the number of factors (i.e., solar radiation intensity, outdoor temperature, humidity and wind speed in this study). At a time step t, the heat load observations and the meteorological observations can be formulated as $Y_t = \{y_{t,1}, y_{t,2}, \dots, y_{t,D'}\} \in \mathbb{R}^{1\times D'}$ and $M_t = \{m_{t,1}, m_{t,2}, \dots, m_{t,D^{\#}}\} \in \mathbb{R}^{1\times D^{\#}}$, respectively. To learn the short-term temporal dependencies of time series, we introduce a sliding window with a size of T to generate the input of the model, representing the consecutive observations of a period. The resulting heat load and the meteorological time series can be formulated as $\boldsymbol{Y}_{t+1:t+T} \in \mathbb{R}^{T\times D'}$ and $\boldsymbol{M}_{t+1:t+T} \in \mathbb{R}^{T\times D^{\#}}$, respectively. Therefore, the prediction can be seen as a multivariate time series prediction problem, and the learning processing with the two time series can be formulated as follows:

$$\hat{Y}_{t+h} \leftarrow \mathcal{F}([\boldsymbol{Y}_{t-T:t}; \boldsymbol{M}_{t-T:t}]), \tag{1}$$

where $\hat{Y}_{t+h} \in \mathbb{R}^{1,D'}$ are the prediction values of the hth time step ahead of t; $\boldsymbol{Y}_{t-T:t} \in \mathbb{R}^{T\times D'}$ and $\boldsymbol{M}_{t-T:t} \in \mathbb{R}^{T\times D^{\#}}$ represent using historical observations with a window size of T for the prediction; $[;]$ represents the concatenation operation; and $\mathcal{F}(\cdot)$ is the linearity/non-linearity mapping function that will be learned in this study.

Since the original time series may contain noise, such as information redundancy and nonstationary fluctuations, this will impair the predictive stability and accuracy of the model. To mitigate this effect, we first use a representation learning function to extract the key patterns from the original time series, and then use the extracted key patterns as the input for prediction. Thus, the prediction model is further formulated as follows:

$$\hat{Y}_{t+h} \leftarrow \mathcal{F}(\mathcal{R}^1(\boldsymbol{Y}_{t-T:t}), \mathcal{R}^2(\boldsymbol{M}_{t-T:t})), \tag{2}$$

where $\mathcal{R}^1(\cdot)$ and $\mathcal{R}^2(\cdot)$ are the representation learning functions to capture key patterns from the heat load and the meteorological time series, respectively.

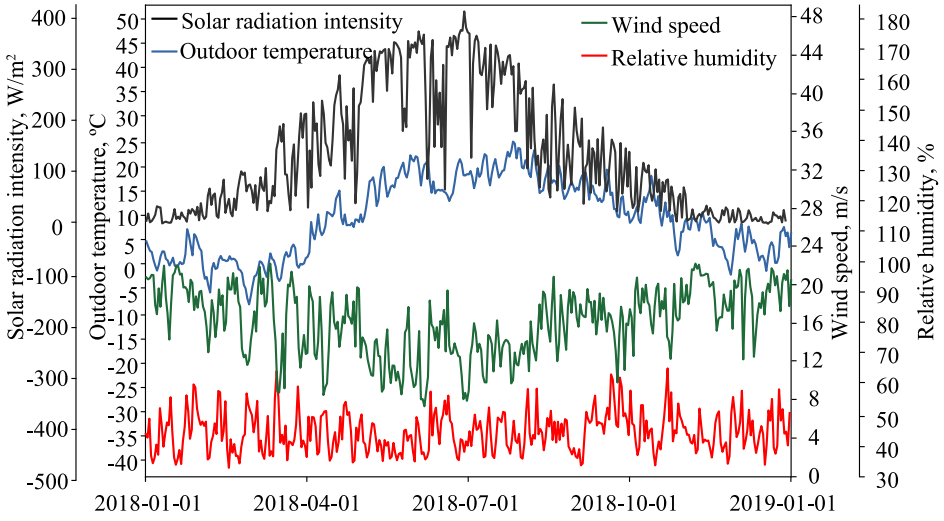


Fig. 3. Outdoor meteorological parameters.

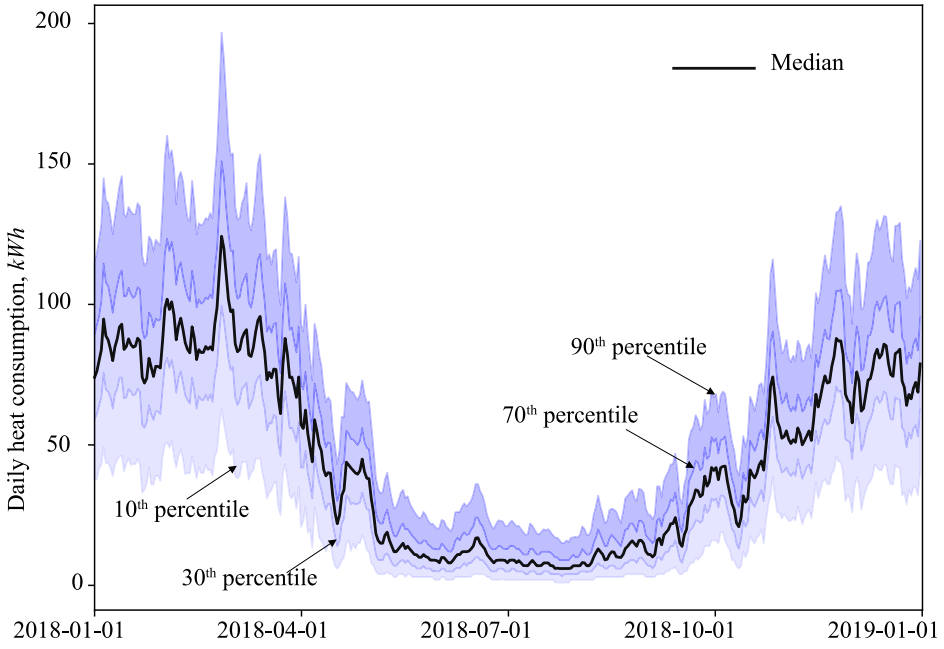


Fig. 4. Percentiles of heat load.

Table 1
The statistics of heat load and meteorological factors.

Symbol	Number of time series	Min	Max	Medium	Mean	STD	PCC	SCC
Relative humidity	1	45.758	99.892	84.252	82.422	10.930	0.369***	0.424***
Solar radiation intensity	1	1.804	391.483	94.046	126.140	106.669	−0.708***	−0.737***
Wind speed	1	1.150	12.829	4.685	4.996	2.059	0.104***	0.103***
Outdoor temperature	1	−7.883	23.712	8.450	9.159	6.138	−0.954***	−0.966***
Single family house observations	2459	0	462.328	42.975	49.685	37.341	–	–

"STD" denotes standard deviation. Pearson correlation coefficient (PCC) and Spearman correlation coefficient (SCC) are calculated by meteorological factors and the median value of heat load.
***p-value < 0.001.

To incorporate the effect of the correlation between exogenous meteorological variables and heat load, we add an additional representation learning component $\mathcal{R}^3(\cdot)$ to the input of the model and obtain the following prediction function:

$$\hat{Y}_{t+h} \leftarrow \mathcal{F}(\mathcal{R}^1(\boldsymbol{Y}_{t-T:t}), \mathcal{R}^2(\boldsymbol{M}_{t-T:t}), \mathcal{R}^3([\boldsymbol{Y}_{t-T:t}; \boldsymbol{M}_{t-T:t}])). \quad (3)$$

4.2. Proposed model

In this subsection, we will describe in detail the proposed GAIN model and the metrics for the model evaluation.

Table 2
Symbols and semantics.

Symbol	Semantic
t	The tth time steps
h	The horizon of prediction
L	The length of input time steps
D	The dimension of input data
U	The number of clusters
T	The length of look-back window size
$\boldsymbol{C}$	The candidate barycenter of K-Means
$\boldsymbol{M}$	The observation of meteorological factors.
$\hat{\boldsymbol{Y}}_T + h$	The prediction of the next h days
$\boldsymbol{Y}_T + h$	The actual value of the next h days
$[;]$	Concatenation operation
Ω	The dataset used for training or testing
$\mathcal{F}(\cdot)$	The mapping function
$\mathcal{R}(\cdot)$	The representation learning component

4.2.1. Overview

Fig. 5 presents an overview of the model, which is a deep neural network trained from historical heat load data and meteorological data. Note that in our experiments later in Section 5, we use GAIN(+) to denote the model trained from both types of data and GAIN to denote the model trained from heat load data only. The upper left corner of the figure indicates the processing of the heat load data, while the lower left corner indicates the processing of the weather data. For heat load data processing, we first perform global clustering to obtain different clusters of customers with similar consumption behaviors, mainly to allow the model to capture the potential relationship between the customers within each cluster. Subsequently, we perform normalization and apply a convolution operation on the obtained results to extract key features from the heat load observations (the convolution operation is adept at local feature engineering [45]). The convolutional output of the heat load and meteorological observations is fed to a collaborative temporal graph attention component to learn the local associations between time steps. An enhanced temporal representation (ETR) with a recursive component is then used to establish the relationship between heat load and weather observations and to improve the ability to learn long-term temporal dependencies. Multiple ETR outputs are combined and then processed by a linear layer, where we add autoregressions of both linear representations of the heat load and meteorological data to adjust the output of the linear layer. Lastly, the final prediction results are obtained by de-normalization (see the lower right corner of the figure). In the following, we describe the components of the proposed neural network structure in more detail.

4.2.2. Global clustering

The living habits of the inhabitants can lead to different preferences in the heat load. To learn the association between the behavior of the inhabitants and the fluctuations in the heat load, we first employ a *K-means* algorithm to cluster the customers into C groups based on the heat load observations in the training set.

Heat load observations are inherently time series samples, which cannot be learned directly by the clustering algorithm. To address the limitation, we utilize the DTW Barycenter Averaging (DBA) [46] approach to average a set of time series and find centroids for the *K-means* method. Dynamic Time Warping (DTW) [47] is the core of the DBA approach and is widely used as a similarity measure between time series. The optimization process of DTW can be expressed as:

$$\mathrm{DTW}(\boldsymbol{Y}_i, \boldsymbol{Y}_j) = \min_{\pi} \sqrt{\sum_{(\lambda^1,\lambda^2)\in\pi} \mathrm{distance}(y_{i,\lambda^1}, y_{j,\lambda^2})^2}, \tag{4}$$

where $\pi = [\pi_0, \ldots, \pi_Q]$ is the optimal alignment between two time series; λ^1 and λ^2 both represent the time step; The π satisfies $\pi_q = (\lambda^1_q, \lambda^2_q)$ with $0 \le \lambda^1_q, \lambda^2_q \le L$, $\pi_0 = (0,0)$ and $\pi_Q = (L-1, L-1)$ where L is the length of the input samples; $y_i \in \boldsymbol{Y}_i$ and $y_j \in \boldsymbol{Y}_j$ denote the heat load observations.

The DBA designs an averaging algorithm to calculate the distance between each time series and the candidate barycenter. The *K-means* algorithm based on DBA is designed to minimize the following objective function:

$$\mathcal{L}(\boldsymbol{Y}, \boldsymbol{C}) = \min_{\boldsymbol{C}} \sum_{i\in N} \sum_{j\in U} \boldsymbol{W}^u_{i,j}\ \mathrm{DTW}(\boldsymbol{y}_i, c_j)^2, \tag{5}$$

where y_i represents the observations of the ith household belonging to Y and N represents the number of households; $\boldsymbol{C}$ denotes the candidate barycenter, and $c_j \in \boldsymbol{C}$ indicates the jth candidate barycenter; $\boldsymbol{W}^u_{i,j}$ is a binary variable denoting if the time series $\boldsymbol{y}_i$ belongs to the jth cluster, $j \in \{1, 2, \ldots, U\}$, U representing the number of clusters.

4.2.3. Time series transformation

Time series transformation is a preprocessing step of the GAIN model. Due to the different magnitudes of heat consumption, this will lead to learning bias and reduce the accuracy of prediction. Therefore, in order to eliminate the impact of numerical differences, it is necessary to apply normalization to scale the input samples.

The *z-score* and *min–max* are two main normalization techniques. The z-score normalization is suitable for processing the data with extreme values. But, the relative difference between the samples is adjusted after z-score normalization. The min–max normalization maintains the relative difference and scales the inputs in a range of $[0, 1]$. Due to the relatively stable fluctuations of the heat load and meteorological observations, we employ *min–max normalization* in the proposed GAIN to normalize the original input data. The normalization processing and recovery processing are formulated as follows:

$$\tilde{\boldsymbol{X}} = \frac{\boldsymbol{X} - \min(\boldsymbol{X})}{\max(\boldsymbol{X}) - \min(\boldsymbol{X})}, \tag{6}$$

$$\boldsymbol{X} = \tilde{\boldsymbol{X}} \cdot (\max(\boldsymbol{X}) - \min(\boldsymbol{X})) + \min(\boldsymbol{X}), \tag{7}$$

where $\boldsymbol{X} \in \mathbb{R}^{L\times D}$ denotes the input data, L is the number of input samples, and D is the dimension of input; $\tilde{\boldsymbol{X}}$ is the normalized data, $\max(\cdot)$ and $\min(\cdot)$ are the maximal and minimal values of $\boldsymbol{X}$, respectively.

After normalization, the heat load observations are adjusted as $\tilde{\boldsymbol{Y}}$, and the meteorological observations are adjusted as $\tilde{\boldsymbol{M}}$. Then, the normalized data are split into multiple pairs for the prediction. That is, using the normalized values of $[t, t+T]$ to predict the values of the $(t+h)$th time step. The split process is also referred as the *h-ahead-step split* [48].

4.2.4. Collaborative temporal graph attention (CTGAT)

In the GAIN, we introduce the Collaborative Temporal Graph Attention (CTGAT) module, which is the core design in the model. The CTGAT contains two multi-head graph attention (GAT) components, which can learn the temporal dependencies from heat load and meteorological observations in parallel (represented as the multiple layers in Fig. 6). The learning process of multi-head GAT is plotted in Fig. 6. This design considers a potential correlation between time steps within a look-back window, and the relationship between time steps is formulated as a complete graph, as shown to the left in Fig. 6.

To obtain representative patterns and reduce information redundancy, CTGAT first employs a convolutional layer in the variable dimension to highlight the key features of heat load observations. The kernels of the convolutional layer with a size of η in the variable dimension. The convolutional operation of the ith filter can be formulated as:

$$\boldsymbol{H}_{i,j} = \boldsymbol{W}^{\epsilon}_{i,j} * \tilde{\boldsymbol{Y}}_j + \boldsymbol{b}^{\epsilon}_{i,j}, \tag{8}$$

where $\boldsymbol{H}_{i,j} \in \mathbb{R}^{B\times T\times 1}$ denotes the output of the ith filter from the jth cluster group, and the final convolution output of the jth group is $\boldsymbol{H}_j \in \mathbb{R}^{B\times T\times \eta}$; $\tilde{\boldsymbol{Y}}_j$ is the heat load observation of the jth cluster group; The symbol $*$ represents the convolutional operation, $\boldsymbol{W}^{\epsilon}$ is the learnable weighting matrix, and $\boldsymbol{b}^{\epsilon}$ is the bias term.

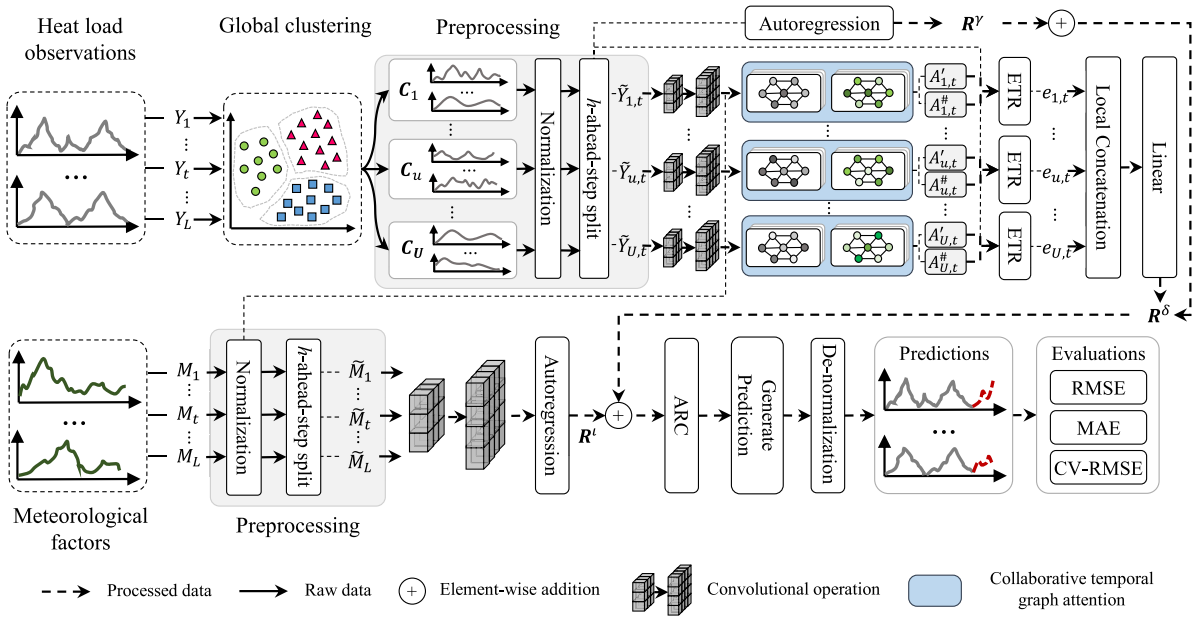


Fig. 5. Overview of the proposed GAIN model.

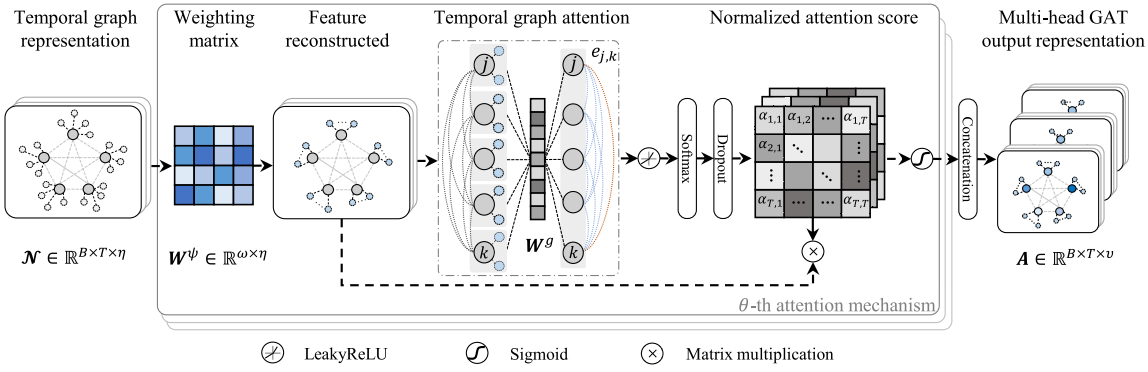


Fig. 6. The multi-head graph attention architecture in the proposed GAIN.

Subsequently, we construct two graph structures by the outputs of the convolutional layer and original meteorological observations, respectively, and then these graph structures are delivered into the GAT layer to learn the potential association between heat loads, and between heat load and meteorological factors. The graph has T nodes, described as $\mathcal{N}_i = \{n_{i,1}, n_{i,2}, \ldots, n_{i,T}\}$, where $n_{i,t} \in \mathbb{R}^{\eta}$ denotes the feature representation of the ith cluster group at time step, t. The adjacent nodes include all of the other time steps in a look-back window.

The correlation between the features of two nodes, j and k, can be expressed as:

$$e_{j,k} = \text{LeakyReLU}(\boldsymbol{W}^{g} \cdot [\boldsymbol{W}^{\psi} \cdot n_j; \boldsymbol{W}^{\psi} \cdot n_k]), \tag{9}$$

where $e_{j,k}$ denotes the importance of node k to node j; LeakyReLU indicates a nonlinear activation function; $\boldsymbol{W}^{g} \in \mathbb{R}^{2\times\omega}$ and $\boldsymbol{W}^{\psi} \in \mathbb{R}^{\omega\times\eta}$ are both learnable weight matrix, and ω is the output dimension of node features.

After obtaining the correlation between nodes, the softmax function is employed to normalize the coefficients and obtain the attention score α by:

$$\alpha_{j,k} = \text{softmax}(e_{j,k}) = \frac{\exp(e_{j,k})}{\sum_{q\in\mathcal{N}} \exp(e_{j,q})}, \tag{10}$$

where $q \in \mathcal{N}$ represents the adjacent nodes for node j, $\alpha_{j,k}$ indicates the normalized attention score of node k to node j, and $\exp(\cdot)$ denotes the exponential function.

Then, the attention score is utilized to transform the input vectors and obtain the final output features for every node:

$$\tilde{n}_{j,\theta} = \sigma\Big(\sum_{k\in\mathcal{N}} \alpha_{j,k,\theta} \cdot \boldsymbol{W}^{\psi} \cdot n_k\Big), \tag{11}$$

where $\tilde{n}_{j,\theta} \in \mathcal{R}^{B\times 1\times\omega}$ represents the weighted feature of node j in the θth attention mechanism, $\boldsymbol{W}^{\psi}$ is the learnable weighting matrix used to linearly transform the input graph, and $\sigma(\cdot)$ is the sigmoid activation function. Finally, the outputs of multiple attention mechanisms are concatenated to generate the final representation:

$$\boldsymbol{A}_j = [\tilde{n}_{j,1}; \tilde{n}_{j,2}; \ldots; \tilde{n}_{j,\Theta}], \tag{12}$$

where $\boldsymbol{A}_j \in \mathbb{R}^{B\times 1\times\upsilon}$ is the weighted features for node j, Θ represents the number of attention mechanism, and υ is the output feature dimension of node j; $\boldsymbol{A}' \in \mathbb{R}^{B\times T\times\upsilon}$ and $\boldsymbol{A}^{\#} \in \mathbb{R}^{B\times T\times\upsilon}$ is the final output of the CTGAT, which represent heat load observations and meteorological observations, respectively.

4.2.5. Enhanced temporal representation (ETR)

As mentioned earlier, temporal graph attentions have superiority in capturing the short-term temporal dependencies, but the long-term time dependencies are commonly ignored. To enhance the learning ability for long-term temporal characteristics, we feed the outputs of CTGAT into the gated recurrent unit (GRU). The heat load observations

are also integrated into GRU to capture the temporal dynamics between attention representation and raw heat load observations. The hidden representation $\boldsymbol{e}_t$ of GRU can be calculated based on the prior $\boldsymbol{e}_{t-1}$, CTGAT outputs, and the heat load observations:

$$\boldsymbol{r}_t = \sigma(\boldsymbol{W}^r \cdot [\boldsymbol{e}_{t-1}; \boldsymbol{A}'; \tilde{\boldsymbol{Y}}; \boldsymbol{A}^{\#}] + \boldsymbol{b}^r), \tag{13}$$

$$\boldsymbol{z}_t = \sigma(\boldsymbol{W}^z \cdot [\boldsymbol{e}_{t-1}; \boldsymbol{A}; \tilde{\boldsymbol{Y}}; \boldsymbol{A}^{\#}] + \boldsymbol{b}^z), \tag{14}$$

$$\boldsymbol{s}_t = \sigma(\boldsymbol{W}^s \cdot [\boldsymbol{r}_t \odot \boldsymbol{e}_{t-1}; \boldsymbol{A}; \tilde{\boldsymbol{Y}}; \boldsymbol{A}^{\#}] + \boldsymbol{b}^s), \tag{15}$$

$$\boldsymbol{e}_t = (1 - \boldsymbol{z}_t) \odot \boldsymbol{s}_t + \boldsymbol{z}_t \odot \boldsymbol{e}_{t-1}, \tag{16}$$

where $\boldsymbol{r}_t$, $\boldsymbol{z}_t$ and $\boldsymbol{s}_t$ represent the reset gate, update gate, and cell state at timestamp t, respectively; $\odot$ is the Hadamard product, and $\sigma(\cdot)$ denotes the sigmoid activation function; $\boldsymbol{W}^r$, $\boldsymbol{W}^z$, and $\boldsymbol{W}^s$ are weights; $\boldsymbol{b}^r$, $\boldsymbol{b}^z$, and $\boldsymbol{b}^s$ are the corresponding biases.

Then, the GRU layer productions of U groups are concatenated, and subsequently make a linear transformation to obtain the final temporal representation:

$$\boldsymbol{R}^\delta = \boldsymbol{W}^\delta \cdot [\boldsymbol{e}_{1,t}; \boldsymbol{e}_{2,t}; ...; \boldsymbol{e}_{U,t}] + \boldsymbol{b}^\delta, \tag{17}$$

where $\boldsymbol{R}^\delta$ is the output temporal representation, generated by extracting the short and long-term temporal dependencies; $\boldsymbol{W}^\delta$ and $\boldsymbol{b}^\delta$ are the learnable weighting matrix and biases parameters, respectively.

4.2.6. Autoregressive representation concatenation (ARC)

The nonlinear learning components have the capability to extract high-level potential relationships, but excessive non-linearity may lead to the neural network outputs being insensitive. To improve the predictive robustness of the proposed GAIN, two autoregression components are used to capture the linear characteristics of heat load and meteorological observations.

For the heat load observations, the autoregression component makes a linear mapping for the temporal dimension, which can be expressed as:

$$\boldsymbol{R}^\gamma = \sum_T \boldsymbol{W}^\gamma \cdot \tilde{\boldsymbol{Y}} + \boldsymbol{b}^\gamma, \tag{18}$$

where $\boldsymbol{R}^\gamma$ is the output linear representation of target time series, and T_γ is the length of the look-back window for autoregression component; $\boldsymbol{W}^\gamma$ and $\boldsymbol{b}^\gamma$ are both the learnable parameters.

For the meteorological observations, we first use a convolutional layer to upscale the dimension of features. This process not only enhances non-linearity by replacing input features with nonlinear combinations, but also improves the expressiveness of neural networks. Then, similar to the processing for target time series, an autoregression component is used to capture the temporal information linearly. Finally, the output dimension of the meteorological observations is made consistent with the dimension of the heat load observations through a linear transformation:

$$\boldsymbol{R}^\iota = \boldsymbol{W}^{\iota^3} \cdot (\sum_T \boldsymbol{W}^{\iota^2} \cdot (\boldsymbol{W}^{\iota^1} * \tilde{\boldsymbol{X}})) + \boldsymbol{b}^\iota, \tag{19}$$

where $\boldsymbol{R}^\iota$ denotes the output linear representation of exogenous time series; $\boldsymbol{W}^{\iota^3}$, $\boldsymbol{W}^{\iota^2}$, and $\boldsymbol{W}^{\iota^1}$ are the weighting matrices; and $\boldsymbol{b}^\iota$ is the bias term.

The final prediction of the proposed GAIN is obtained by integrating the ETR and ARC outputs:

$$\boldsymbol{O}_{t+h} = \boldsymbol{R}^\delta + \boldsymbol{R}^\gamma + \boldsymbol{R}^\iota, \tag{20}$$

where $\boldsymbol{O}_{t+h} \in \mathbb{R}^{B \times 1 \times D'}$ is the output of proposed GAIN, D' is the input dimension of district heat load data, $\hat{\boldsymbol{Y}}_{t+h}$ is the final prediction results after de-normalization.

The Mean Square Error (MSE) is adopted as the loss function in the training process, which can be formulated as:

$$\mathcal{L}(\boldsymbol{Y}_{t+h}, \hat{\boldsymbol{Y}}_{t+h}) = \frac{1}{L} \sum_{i=1}^{L} \sum_{j=1}^{D'} (y_{t+h,i,j} - \hat{y}_{t+h,i,j})^2, \tag{21}$$

where L is the length of input time steps in the training process. The $y_{t+h,i,j}$ denotes the ith training sample's actual heat load of jth district at $t+h$ time step. $\hat{y}_{t+h,i,j}$ is the predictive values.

4.3. Evaluation metrics

The root mean square error (RMSE), mean absolute error (MAE), and coefficient of variation of RMSE (CV-RMSE) are combined to measure the prediction performance. Each metric has a different purpose: MAE and RMSE are both scale-related metrics, but MAE depends on the absolute errors, while RMSE is based on squared errors. CV-RMSE is a scale-independent metric commonly used in energy prediction, which eliminates the RMSE's dependence on the scale of the data. These metrics are formulated as follows:

- Root Mean Squared Error:

$$\text{RMSE} = \sqrt{\mu(\sum_{(i,t) \in \Omega_{Test}} (Y_{i,t} - \hat{Y}_{i,t})^2)}, \tag{22}$$

- Mean Absolute Error:

$$\text{MAE} = \mu(\sum_{(i,t) \in \Omega_{Test}} |Y_{i,t} - \hat{Y}_{i,t}|), \tag{23}$$

- Coefficient of Variation of Root Mean Square Error:

$$\text{CV-RMSE} = \frac{\sqrt{\mu(\sum_{(i,t) \in \Omega_{Test}} (Y_{i,t} - \hat{Y}_{i,t})^2)}}{\mu(\hat{\boldsymbol{Y}}_{i,:})}, \tag{24}$$

where $Y_{i,t}$ and $\hat{Y}_{i,t}$ represent real and predictive heat load value at time t of the i household, respectively; Ω_{Test} represents the test set and $\mu(\cdot)$ is used to calculate the mean value. The smaller the values of RMSE, MAE and CV-RMSE, the better the performance.

5. Experiments

This section carries out experiments to evaluate the proposed model and presents experimental settings, model implementations, data, and results.

5.1. Experimental settings and data

The GAIN model was implemented using the deep learning framework, Pytorch v1.12.1. All experiments were carried out on a server equipped with an Intel(R) Xeon(R) Gold 6226R CPU (2.90 GHz) with 128G memory and were accelerated by two NVIDIA RTX A6000 GPUs. The data used for the experiments were described in Section 3. Due to the difference in the data distribution in different types of households, we select the heat load observations of the single-family household to conduct experiments, which has the largest number of samples in the dataset.

5.2. Baseline methods

To evaluate the proposed GAIN model, we select 15 multivariate time series prediction methods as the baselines for comparison. Their brief descriptions are listed as follows.

- Global Autoregression (GAR) [28] uses an autoregressive component to capture global features.
- Long and Short-term Memory (LSTM) [49] captures the temporal dependence by cycling four gate units.

- Gated Recurrent Unit (GRU) [50] is a variant of the recurrent neural network with a more concise gate architecture.
- Encoder–Decoder (ED) [50] employs LSTM components in the encoding stage and the decoding stage, respectively.
- Convolution Neural Network (CNN) [51] is a two-layer learning structure using one-dimensional convolution operations.
- Convolution Recurrent Neural Network (CRNN) [52] combines the recurrent neural network with the CNN component.
- Convolution Recurrent Neural Network with Residual (CRNN-Res) [52] uses residual components to avoid the loss of detailed information caused by convolution operations.
- TPA-LSTM [53] uses the recurrent neural network with an attention mechanism to capture nonlinear interdependencies between time steps and series.
- LSTNet [54] captures long-term dependencies and periodic patterns by redesigned convolutional and recurrent structures.
- MTNet [55] employs memory components, encoders, and autoregressive components to model complex temporal patterns or dependencies.
- Multivariate Shapelet Learning (MSL) [48] learns multiple crucial subsequences from historical observations.
- Dense [32] is a non-linear artificial neural network with fewer parameters and performs well in district heating load prediction.
- BiLSTM [31] is a variant of the LSTM that has been used to predict the district heating load.
- Hybrid CNN-LSTM (HCLSTM) [33] is a spatio-temporal prediction algorithm that has shown promising performance in short-term heating load prediction.
- Informer [35] is a variant of vanilla Transformer architecture that had been used in district heat load prediction [34].

5.3. Model configurations

We use the grid search method to adjust the hyperparameters for each method. Due to the chronological order nature of the heat load time series, we use five repeat experiments instead of cross-validation. The Adam optimizer [56] is used to obtain the training models, and the mean squared error (MSE) is selected as the loss function. The training epoch and the learning rate are tuned to the optimal states for each method, while the other training-related constant parameters are kept the same. Table 7 describes the details of the possible hyperparameters of the baseline methods.

5.4. Results and analysis

This subsection presents the experimental results for the evaluation of hyperparameters, comparison with the baseline methods, correlation analysis of prediction, model ablation study and feature ablation study, respectively.

5.4.1. Hyperparameter evaluation

The sliding window size indicates the association between the past T days and the future $T+h$ days. An appropriate window size can help the prediction model to better capture potential energy consumption patterns. It is worth noting that, in order to explore the temporal relationship between time steps, the window size should be greater than 1. Therefore, to determine the optimal window size for the proposed model, we make predictions for a time horizon of $h = 1$, and keep the other parameters constant while increasing the window size from 2 to 21. To make fair comparisons, we repeat each experiment five times to avoid the effect of accidental values. The experimental results are shown in Fig. 7.

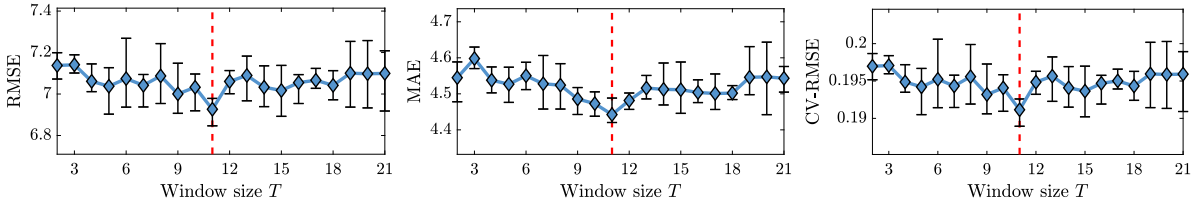
As shown in Fig. 7, the model achieves the best performance for a window size of 11, while for small or large window sizes, the model shows sub-optimal performance. For example, when the window size is set to 1–5 days, it is difficult to achieve good accuracy. This is due to the greater randomness of small windows, including many unexpected events, inflection points, and small fluctuations, which have an impact on short-term energy and thus on prediction performance. When using a relatively large window size, e.g. 15–21 days, although the large window contains more temporal information, it has more serial fluctuations. Thus, relatively small or large window sizes can affect key temporal information extraction capability and increase prediction errors. Therefore, for the rest of the experiments, we will use the medium window size $T = 11$, which has a good capability of capturing important fluctuations while reducing the effect of stochastic fluctuations, with less parameter complexity.

Furthermore, the number of clusters, U, is another key hyperparameter that determines the number of heat load groups and the graph learning components. To find an optimal number of clusters, we conduct the experiments by varying the number of clusters from 2 to 6 under the optimal window size, a prediction time horizon of $h = 1$, and fixing other uncorrelated structural parameters. The results in Fig. 8 show that when the number of clusters is set to 3, the overall performance is relatively superior.

The obtained results can be explained by the fact that the number of clusters selected for the model plays a crucial role in the accuracy and efficiency of the proposed method. Specifically, if the number of clusters is set too low, the model may not be able to capture sufficient details in heat load observations, which results in excessive heterogeneity within the clusters. Conversely, if the number of clusters is set excessively high, it may cause some data points that should be grouped together to be separated into different clusters, leading to excessive homogeneity between the clusters. Both of these scenarios can have a significant impact on the accuracy and efficacy of the proposed method. When the number of clusters U is set to 3, the prediction model can effectively extract the potential behavior patterns of customers while avoiding excessive model complexity.

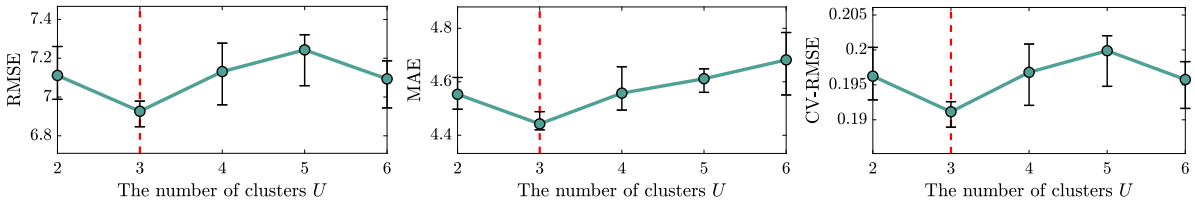
5.4.2. Comparison with baselines

Table 3 summarizes the performance results of all methods based on three metrics and four time horizons, $h = 1, 3, 5, 7$, for short-term prediction. The best results are highlighted in bold font. In general, the performance of all methods decreases with increasing time horizons. First, when meteorological factors are not considered, we can observe the following findings. The proposed GAIN model demonstrates superior performance for all three evaluation metrics across all four time horizons, with particularly strong results for low horizons. Specifically, GAIN achieves a maximum reduction of 2%, 3%, and 2% in RMSE, MAE, and CVRMSE, respectively. This result demonstrates the good generalization capability of the proposed GAIN. The Informer model exhibits superior performance in general and achieves the second-best results for forecasting horizons of 5 and 7. This is attributed to its utilization of multiple attention mechanisms, which enable the model to capture multi-scale temporal representations. MTNet has the second best performance when $h = 1$ and $h = 3$, which capture the temporal dependencies with a recurrent learning component and local attention mechanism. However, with the increase of horizon, merely relying on the information of local relationships for MTNet is insufficient to make an accurate prediction. When h is set to 1 and 3, GAR outperforms most of the RNN variants (i.e., LSTM, GRU, ED, and BiLSTM) and hybrid RNN models (i.e., CRNN, CRR-Res, TPA-LSTM, LSTNet, and HCLSTM). This phenomenon highlights the importance of linear representation in short-term daily heat load prediction. The trend and fluctuation of heat load observation at this data granularity are relatively stable. The CRNN model outperforms the vanilla RNN or CNN architecture in low horizon forecasting tasks. This suggests the potential benefits of integrating a CNN component to encode input features and enhance prediction accuracy. The combination of convolution and RNN is the core component of TPA-LSTM, LSTNet, and MTNet, but these methods further incorporate the auto-regressive linear characteristics and exhibit more stable performance than traditional CRNN architectures,



(a) RMSE (b) MAE (c) CV-RMSE

Fig. 7. The GAIN performance by varying the window size T in terms of three metrics. For each metric, the optimal value is found at red dash line.



(a) RMSE. (b) MAE. (c) CV-RMSE.

Fig. 8. The GAIN performance with varying the number of clusters U in terms of three metrics.

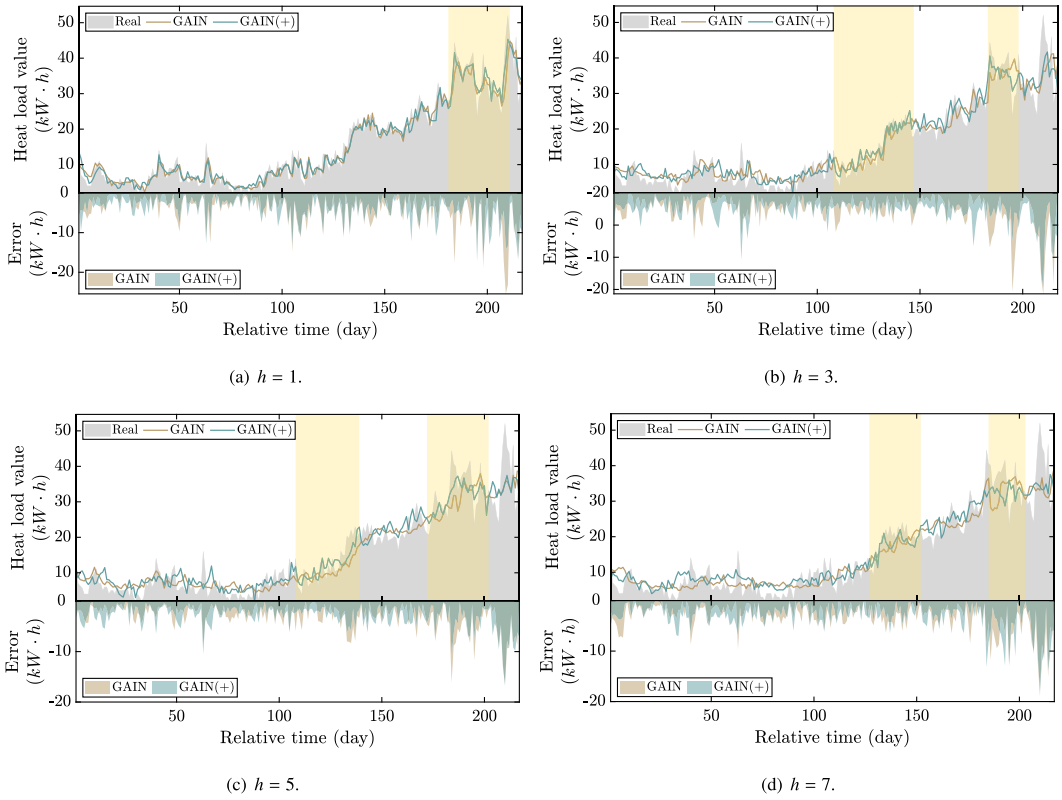
Table 3
Performance comparison based on three metrics and four time horizons in heat load prediction.

Model	$h = 1$			$h = 3$			$h = 5$			$h = 7$		
	RMSE	MAE	CVRMSE	RMSE	MAE	CVRMSE	RMSE	MAE	CVRMSE	RMSE	MAE	CVRMSE
GAR	7.256	4.601	0.200	9.187	6.145	0.254	10.321	7.093	0.285	11.197	7.800	0.309
LSTM	9.363	6.307	0.258	10.396	7.136	0.287	10.489	7.296	0.289	10.748	7.524	0.297
GRU	9.463	6.415	0.261	10.457	7.238	0.289	10.880	7.584	0.300	11.195	7.868	0.309
ED	9.321	6.269	0.257	10.391	7.166	0.287	10.635	7.404	0.293	10.703	7.423	0.295
CNN	9.784	6.759	0.270	11.023	7.682	0.304	11.407	8.110	0.315	11.495	8.253	0.317
CRNN	9.297	6.290	0.257	10.378	7.160	0.286	10.912	7.634	0.301	11.110	7.910	0.307
CRNN-Res	10.019	6.919	0.276	11.187	7.845	0.309	11.680	8.263	0.322	12.050	8.591	0.333
TPA-LSTM	7.148	4.622	0.197	9.355	6.307	0.258	9.839	6.887	0.272	10.168	7.102	0.281
LSTNet	7.103	4.605	0.196	9.231	6.382	0.255	10.036	6.888	0.277	10.510	7.278	0.290
MTNet	7.072	4.559	0.195	8.873	6.012	0.245	9.790	6.843	0.270	10.607	7.536	0.293
MSL	7.265	4.631	0.200	9.123	6.182	0.252	10.248	7.080	0.283	11.106	7.760	0.306
Dense	7.202	4.612	0.199	9.151	6.093	0.253	10.277	7.012	0.284	11.128	7.715	0.307
BiLSTM	9.259	6.197	0.256	10.287	7.086	0.284	10.668	7.446	0.294	10.935	7.679	0.302
HCLSTM	9.254	6.199	0.255	10.219	7.091	0.282	10.781	7.587	0.298	10.868	7.660	0.300
Informer	9.012	6.148	0.249	9.169	6.286	0.253	9.457	6.534	0.261	9.962	6.998	0.275
GAR(+)	10.922	7.552	0.301	11.065	7.735	0.305	11.231	7.841	0.310	11.684	8.174	0.322
LSTM(+)	9.338	6.279	0.258	10.423	7.176	0.288	10.760	7.542	0.297	10.854	7.600	0.300
GRU(+)	9.493	6.439	0.262	10.456	7.225	0.289	10.918	7.700	0.301	10.997	7.777	0.303
ED(+)	9.275	6.195	0.256	10.412	7.127	0.287	11.309	8.026	0.312	10.980	7.695	0.303
CNN(+)	9.581	6.563	0.264	10.681	7.474	0.295	10.745	7.549	0.297	11.312	8.106	0.312
CRNN(+)	9.245	6.220	0.255	10.342	7.117	0.285	11.567	8.206	0.319	11.129	7.874	0.307
CRNN-Res(+)	9.958	6.882	0.275	11.104	7.795	0.306	10.805	7.508	0.298	11.831	8.493	0.327
Dense(+)	9.081	6.067	0.251	10.348	7.154	0.286	10.886	7.686	0.300	11.211	7.938	0.309
HCLSTM(+)	9.169	6.168	0.253	10.191	7.043	0.281	10.677	7.523	0.295	10.710	7.523	0.296
Informer(+)	8.998	6.122	0.248	9.281	6.334	0.256	9.585	6.622	0.265	10.084	7.134	0.278
GAIN	6.927	4.442	0.191	8.764	5.970	0.242	9.384	6.437	0.259	9.897	6.966	0.273
GAIN(+)	**6.720**	**4.329**	**0.185**	**8.520**	**5.730**	**0.235**	**9.364**	**6.432**	**0.258**	**9.678**	**6.863**	**0.267**

The best result for each metric is highlighted in bold. "(+)" denotes the prediction method that integrates meteorological factors. Unit of h: day.

e.g., HCLSTM, CRNN, and CRNN-Res. The MSL and Dense models exhibit relatively good performance in the low horizon forecasting task. However, they are unable to effectively extract consecutive temporal dependencies and display significant performance fluctuations as the forecasting horizon h increases.

Then, in the presence of meteorological factors, we compared our proposed method, GAIN(+), against 10 benchmark models, excluding TPA-LSTM, LSTNet, MSL, and BiLSTM, as they do not take exogenous factors into account in their prior studies. According to the experimental results presented in Table 3, GAIN(+) outperforms all other benchmark methods in all four time horizons, achieving a maximum reduction of 25%, 29%, and 25% in RMSE, MAE, and CVRMSE, respectively. Notably, GAIN(+) also achieves a maximum reduction of 3%, 4%, and 3% in RMSE, MAE, and CVRMSE compared to GAIN, highlighting the effectiveness and necessity of considering meteorological factors in our model for short-term heat load prediction tasks. In our design, we utilize the CTGAT module to extract the key information of heat load and meteorological observations, while the linear autoregressive properties of meteorological factors are also taken into account, which also improves the prediction performance. To further compare GAIN and GAIN(+), we randomly select the heat load of a household for prediction over different time horizons and obtain the results shown in Fig. 9. Intuitively, GAIN(+), which takes into account meteorological factors, fits the real heat load observations better than GAIN, especially when the time horizon is set to 1 or 3. This is because, with consideration of meteorological factors, GAIN(+) can be more effective



(a) $h = 1$. (b) $h = 3$.

(c) $h = 5$. (d) $h = 7$.

Fig. 9. The visualizations among real values, GAIN predictions, and GAIN(+) predictions and prediction error; Unit of h: day.

in capturing the fluctuation of the peak load. During the highlight period, the prediction accuracy of GAIN(+) is significantly better than GAIN. We also note that as the time horizon increases, the prediction performance of the two methods gets closer. Several fluctuations are observed during periods of high volatility. This indicates that there is uncertainty in the effect of meteorological factors on heat load prediction for large time horizons.

Regarding the benchmark methods, the major observations are summarized as follows. The prediction accuracy of GAR(+) and Dense(+) is significantly worse than that without weather factors. This result suggests that these models have limitations in extracting potential representations from exogenous inputs, which can cause issues in the weight allocation of these features, ultimately affecting the prediction performance. HCLSTM(+) benefited from the weather factors and shows better performance than HCLSTM under the four time horizons. The phenomenon reflects the potential of the CRNN architecture in enhancing the accuracy of predictions by integrating exogenous factors, while also demonstrating the effectiveness of weather factors in heat load prediction. Other methods, such as LSTM, ED, CNN, CRNN, CRNN-Res, and Informer, also show slight improvement when h is set as 1 or 3h. This result further emphasizes the important role of weather factors in short-term heat load prediction tasks.

5.4.3. Analysis of predictions

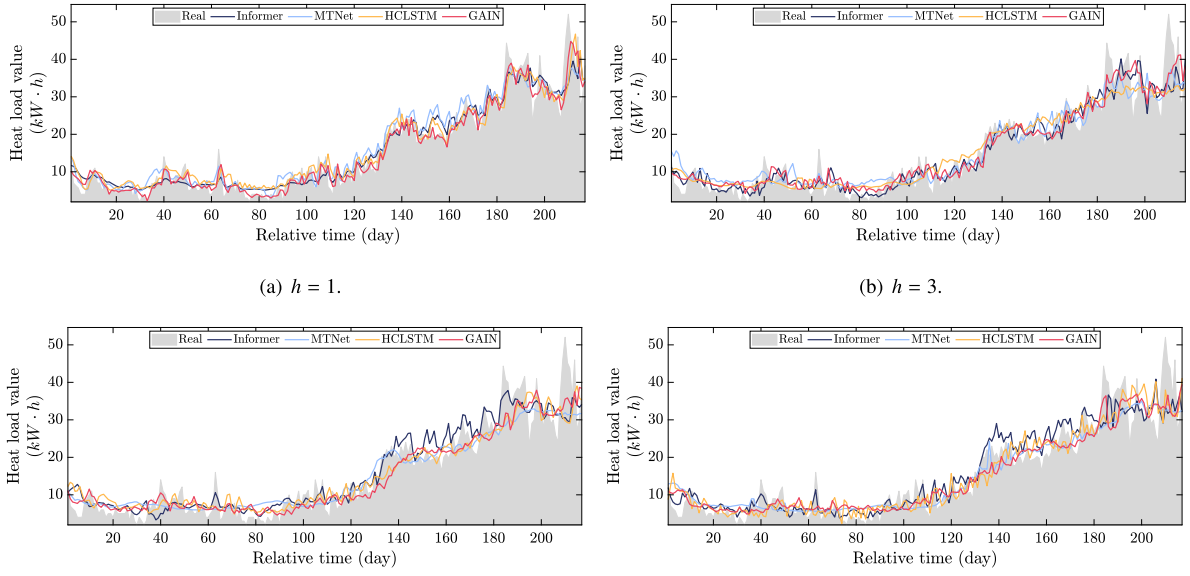
Fig. 10 shows the comparison of the actual and predicted heat load values by GAIN, Informer, HCLSTM, and MTNet. Among the four methods, GAIN performs the best in tracking the heat load variations, especially when the prediction horizon h is short (1 or 3). However, as h increases, all the methods tend to underestimate the peak values and lose accuracy. Informer, which mainly uses the attention mechanism, shows more fluctuations in the forecasting than MTNet, HCLSTM, and GAIN, which incorporate RNN. This suggests that RNN is more effective in capturing the temporal dependencies and reducing forecasting errors. GAIN leverages the graph neural network framework to capture the complex and nonlinear relationships between the customers, heat load and meteorological factors, which enables it to produce more accurate and realistic predictions.

Fig. 11 shows the normalized results and Pearson correlation (*PCC*) between the actual and predicted values by GAIN. The prediction error increases as the observation size increases, which is consistent with the previous figure. This indicates the difficulty of predicting the heat load accurately during high fluctuation periods, especially for long-term forecasting. Most of the data points are below the diagonal line, which means that the prediction model tends to lag behind the actual values when they reach the peak. This could be due to the instability of the heat load data during high fluctuation periods, which makes it hard for the prediction model to find stable patterns. It could also be due to the prediction model's limitation in capturing the sudden changes in the data, which leads to a delayed reaction in forecasting the peak values. Nevertheless, GAIN still outperforms other methods in terms of PCC for all four forecasting horizons.

The effect of weather factors on the prediction performance is shown in Figs. 12 and 13. Informer(+) exhibits unstable and fluctuating results, especially for $h = 7$. This could be due to its complex learning architecture, which may not be able to filter out the noise from the external data that affects the prediction target. HCLSTM(+) shows stable but inaccurate results, especially for the peak values. On the contrary, GAIN(+) performs the best among the three methods in terms of fitting and *PCC* values for all four horizons, especially for $h = 1$. However, adding weather features also increases the uncertainty and worsens the prediction during high volatility periods, as seen in Figs. 13(c) and 13(d). This confirms the challenge of predicting the heat load with meteorological factors for long-term forecasting.

5.4.4. Model ablation study

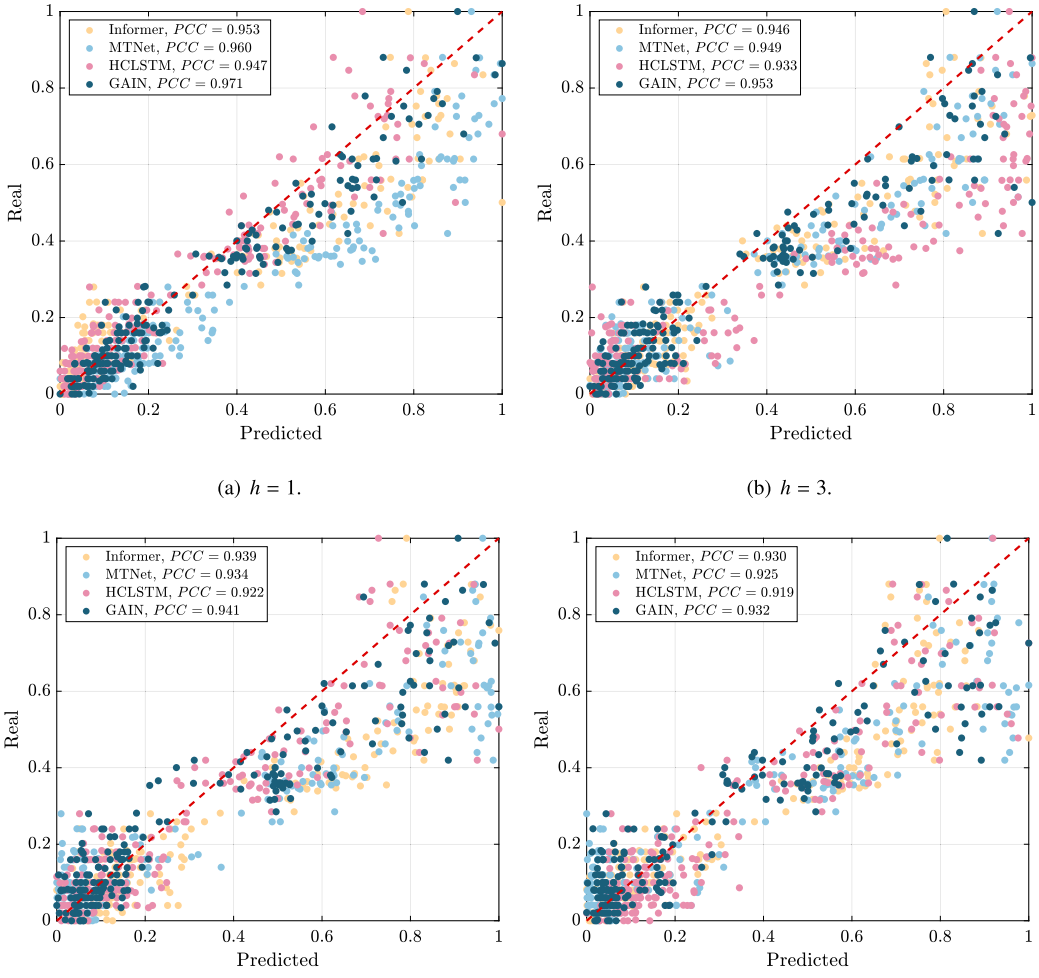
To better understand the contribution of each GAIN component to the accuracy of the final model, we performed ablation experiments for four time horizons and obtained the results presented in Table 4.



(a) $h = 1$. (b) $h = 3$.

(c) $h = 5$. (d) $h = 7$.

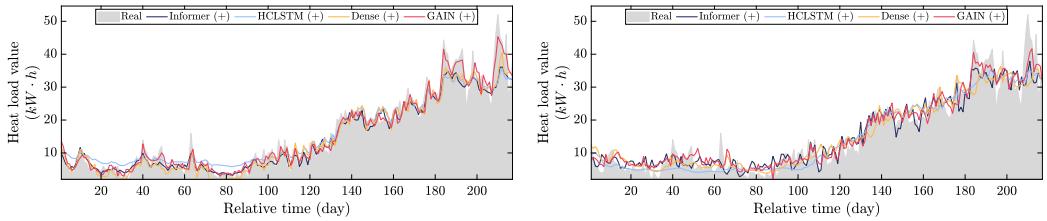
Fig. 10. The visualizations comparing the real values with the GAIN predictions and the four other benchmark predictions; Unit of h: day.



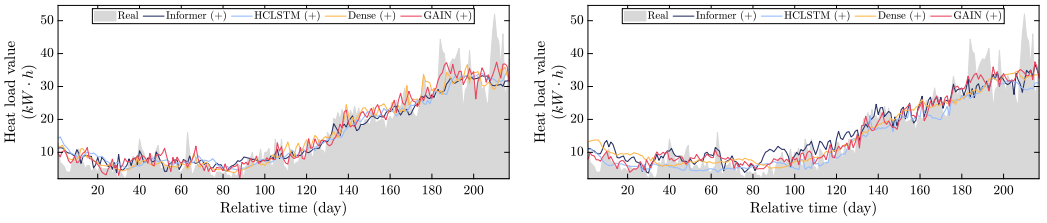
(a) $h = 1$. (b) $h = 3$.

(c) $h = 5$. (d) $h = 7$.

Fig. 11. The correlation visualizations of GAIN predictions and four other benchmark predictions; Unit of h: day.

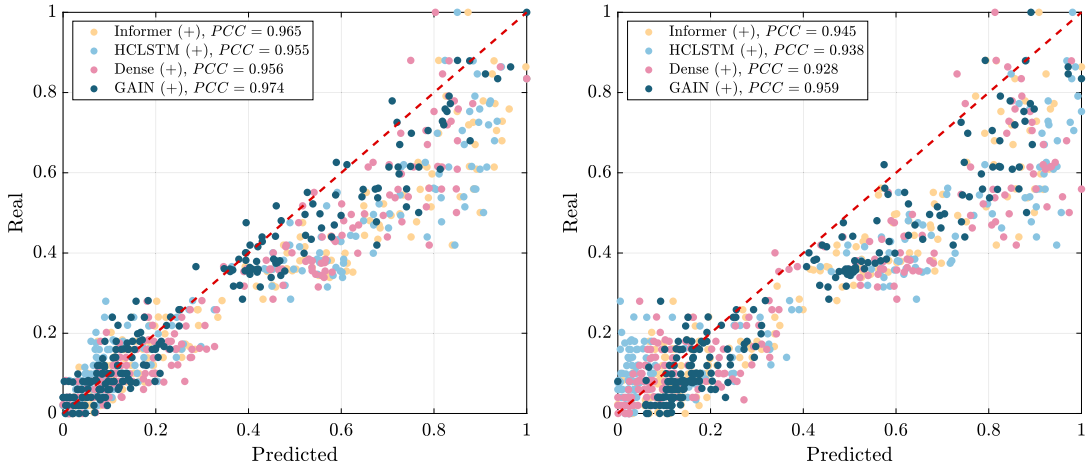


(a) $h = 1$. (b) $h = 3$.



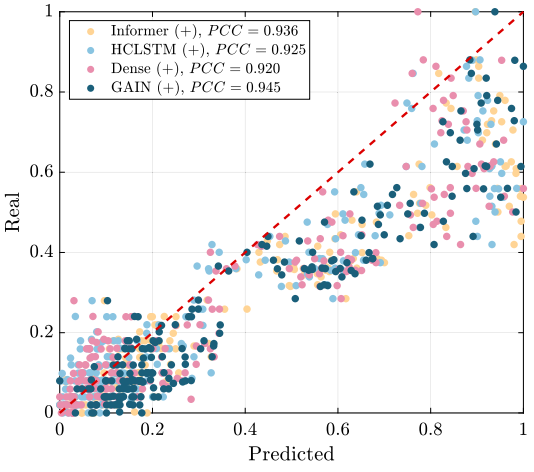
(c) $h = 5$. (d) $h = 7$.

Fig. 12. The visualizations comparing the real values with the GAIN(+) predictions and the four other benchmark predictions; Unit of h: day.

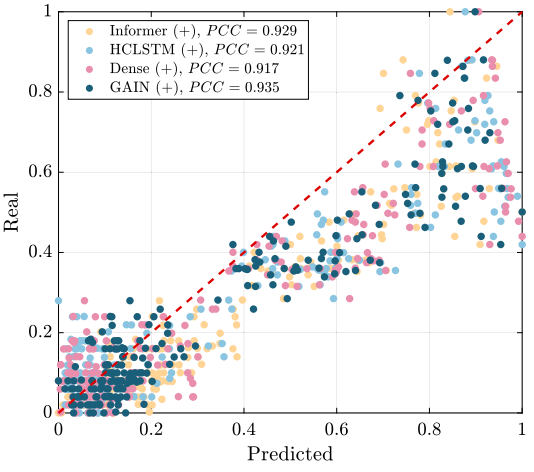


(a) $h = 1$.

(b) $h = 3$.



(c) $h = 5$.



(d) $h = 7$.

Fig. 13. The correlation visualizations of GAIN(+) predictions and four other benchmark predictions; Unit of h: day.

Table 4
Model ablation study. The best results are shown in bold, and the worst results in wavy lines; Unit of h: day.

Model	h = 1			h = 3			h = 5			h = 7		
	RMSE	MAE	CVRMSE	RMSE	MAE	CVRMSE	RMSE	MAE	CVRMSE	RMSE	MAE	CVRMSE
GAIN	**6.927**	**4.442**	**0.191**	**8.764**	**5.970**	**0.242**	**9.384**	**6.437**	**0.259**	**9.897**	**6.966**	**0.273**
w/o K-means	7.258	4.616	0.200	9.178	6.142	0.253	10.266	7.033	0.283	11.123	7.696	0.307
w/o convolution	7.148	4.572	0.197	8.853	5.995	0.244	9.387	6.557	**0.259**	9.987	7.080	0.276
w/o TGAT	6.956	4.473	0.192	8.842	6.098	0.244	9.706	6.874	0.268	10.212	7.254	0.282
w/o ETR	6.942	4.520	0.192	8.790	6.025	0.243	9.458	6.684	0.261	10.004	7.057	0.276
w/o ARC	9.540	6.442	0.263	10.438	7.211	0.288	10.741	7.527	0.296	11.104	7.772	0.306

Table 5
Feature ablation study. The best results are marked in bold and the second-best results are underlined. Unit of h: day.

Model	h=1		h=3		h=5		h=7	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
GAIN(+)	**6.720**	**4.329**	**8.520**	**5.730**	**9.364**	**6.432**	**9.678**	**6.863**
w/o outdoor temperature (w/o OT)	7.153	4.661	8.834	6.016	9.673	6.784	9.858	7.031
w/o solar radiation intensity (w/o SRI)	6.893	4.586	8.705	6.019	9.577	6.637	10.067	7.002
w/o wind speed (w/o WS)	6.776	4.503	8.611	5.944	9.437	6.674	9.743	6.927
w/o relative humidity (w/o RH)	6.813	4.551	8.638	5.973	9.473	6.645	9.756	6.974

First, we remove the K-means structure (`w/o K-means`), and then feed the unclustered data directly into CTGAT. The results show that `w/o K-means` does not perform well, especially for relatively large time horizons. This is because the clustering component allows the prediction model to learn potential similarities and regularities of the load profiles for each cluster. Second, we remove the convolution component in CTGAT (`w/o convolution`), and the results show considerable prediction error. Convolution filters can incorporate or separate correlations among different dimension features and discover local dependency patterns [57]. When the convolution component was removed, the ability to learn local heat load patterns will be compromised. Local patterns can typically affect more on the accuracy of the prediction of small time horizons. Third, we remove TGAT (`w/o TGAT`) and feed the convolution output directly into the ETR structure. The results show that the `w/o TGAT` performs worse than the original GAIN, especially for large time horizons. Temporal graph attention treats each time step as an individual node in a time window and establishes correlations between different time steps [58]. The correlations of local time steps can capture the temporal dynamics between load profiles. For large time horizon predictions, the key ingredient is to capture the causal relationship between time steps. TGAT has superiority in capturing global temporal dependencies and enhancing the stability of the model with large time horizons [59]. Similarly, to verify the validity of the recurrent layer, we use the GAR component instead of the ETR component (`w/o ETR`). Interestingly, the prediction accuracy of `w/o ETR` decreases only slightly for small time horizons, while the prediction error becomes prominent for large time horizons. This suggests that recurrent neural networks are more suitable for establishing temporal dependencies over large time horizons, while short-term local dependencies can be obtained with the TGAT component. Finally, we remove the ARC component (`w/o ARC`) and the performance shows a significant drop. This implies the importance of the autoregressive linearity property. The fluctuations of daily heat load observations are more stable than hourly observations. Therefore, simple autoregression can make a proper linear adjustment for prediction [54].

In conclusion, the ablation study validates the efficacy of the GAIN model structure design, which takes into account not only the heat load and meteorological influences, but also the short-term and long-term time dependence in the time series and the local correlation between time steps. In addition, linear features are incorporated into the proposed model design.

5.4.5. Feature ablation study

The proposed GAIN(+) considers four meteorological factors, including outdoor temperature, solar radiation intensity, wind speed, and relative humidity. To investigate their contribution to model performance, we conducted a feature ablation study and present the results in Table 5.

First, we remove the outdoor temperature factor (`w/o OT`), which is the dominant ingredient related to the heat load [60]. The GAIN(+) exhibits the worst prediction accuracy in all four time horizons. A potential explanation is that the outdoor temperature is the most perceptible by the human body. Prediction models can easily establish the causal relationship between outdoor temperature and heat load. Without outdoor temperature, the prediction model lacks an important causal relationship, leading to an increase in forecast error. Second, we remove the feature of solar radiation intensity (`w/o SRI`), and the prediction accuracy slightly decreases. The intensity of solar radiation is one of the causes of temperature variations, which also indirectly influences relative humidity [61]. As shown in Fig. 3, the trends in the intensity of solar radiation and outdoor temperature are similar, indicating their causal relevance. Therefore, ignoring solar radiation does not have a significant effect on prediction accuracy. Third, we remove the feature of wind speed (`w/o WS`), and the obtained model has the second-best performance in terms of RMSE in the four time horizons. This result suggests that wind speed has some contribution to heat load prediction accuracy but is much lower than the outdoor temperature. This may be due to the fact that wind speed is an air property that is indirectly related to weather temperature. Previous studies [36] have verified that wind speed is valuable auxiliary information. Lastly, when the relative humidity (`w/o RH`) is not considered, the performance is better than `w/o SRI` but worse than `w/o WS`. Previous research [62] has shown that relative humidity plays an important role in both the indoor environment and energy conservation.

The above ablation studies of the four meteorological factors show that their contributions to the model prediction performance are of varying validity, which has been fully considered in our study.

6. Conclusions and future work

Heat load prediction plays a pivotal role in the operations of energy stations and assists with energy demand side management. In this paper, we propose a graphical ambient intelligence algorithm for district heating load forecasting. We first applied clustering to identify different customer groups based on their load profiles. Then, we designed a collaborative temporal graph attention mechanism for extracting features from heat load and meteorological observations, enabling the discovery of causal relationships between time steps. To improve the capability of capturing temporal dependencies, we introduced a recurrent neural network into the proposed GAIN structure, allowing it to

Table 6
Performance comparison based on two metrics and four time horizons in three category collections on the JERICHO-E-usage dataset. The best results are marked in bold and the second best results are underlined; Unit of h: day. Note that the RMSE and MAE values are with an increase of 6 orders of magnitude, i.e., $\times 10^6$. For the commerce and industry datasets, the number of clusters U in GAIN is set to 3, while for the residential dataset, U is set to 2.

Category	h	Metrics	Dense	BiLSTM	HCLSTM	Informer	MTNet	GAIN
Commerce	1	RMSE	2.180	2.391	2.475	**1.867**	2.230	2.164
		MAE	**1.320**	1.482	1.539	1.334	1.360	1.332
	3	RMSE	2.805	2.696	2.567	**2.412**	2.523	2.471
		MAE	1.905	1.723	1.627	1.711	1.720	**1.610**
	5	RMSE	2.990	2.682	2.736	2.562	2.682	**2.556**
		MAE	1.998	1.768	1.767	1.803	1.770	**1.761**
	7	RMSE	3.029	2.758	2.741	2.650	2.696	**2.603**
		MAE	2.016	1.780	1.807	1.837	1.761	**1.644**
Industry	1	RMSE	0.505	0.583	0.559	0.486	0.553	**0.466**
		MAE	0.292	0.336	0.332	0.321	0.314	**0.290**
	3	RMSE	0.704	0.689	0.656	0.645	0.655	**0.640**
		MAE	0.443	0.418	0.387	0.419	0.396	**0.386**
	5	RMSE	0.762	0.699	0.652	**0.642**	0.650	0.644
		MAE	0.470	0.429	0.392	0.405	0.422	**0.391**
	7	RMSE	0.770	0.748	0.698	0.656	0.742	**0.653**
		MAE	0.473	0.450	0.431	0.429	0.455	**0.399**
Resident	1	RMSE	**12.302**	12.745	13.171	12.666	12.912	12.492
		MAE	**7.661**	8.670	9.072	9.479	8.770	8.209
	3	RMSE	15.525	14.497	14.895	**14.362**	15.624	14.421
		MAE	10.494	10.091	10.509	10.604	11.495	**9.935**
	5	RMSE	16.369	14.989	15.369	**14.167**	16.769	14.820
		MAE	11.153	10.676	11.017	10.378	11.317	**9.916**
	7	RMSE	17.215	15.446	17.346	14.903	16.731	**14.849**
		MAE	11.791	10.957	12.200	10.895	11.673	**10.067**

correlate heat load and meteorological data in the temporal dimension. Moreover, we considered linear characteristics of time series in our model to increase the diversity of feature representation and to enhance the robustness of the model. Finally, we conducted comprehensive experiments to evaluate our model, including comparisons with fifteen baseline methods, correlation analysis of predictions, model structure, and feature ablation studies. The experimental results have shown that the proposed model outperforms all the baseline methods and demonstrate the effectiveness of our model design.

Several directions for future work exist. First, the interpretability of the proposed GAIN model will be further investigated. Second, further consideration will be given to the application of graph neural networks to multiple energy loads and household types. Third, we will extend the prediction from short-term to long-term time horizons and improve generalization performance.

CRediT authorship contribution statement

Zhijin Wang: Project administration, Conceptualization, Formal analysis, Software, Funding acquisition, Writing – review & editing, Supervision. **Xiufeng Liu:** Data acquition, Conceptualization, Funding acquisition, Writing – review & editing, Supervision. **Yaohui Huang:** Data curation, Methodology, Visualization, Investigation, Validation, Writing – original draft. **Peisong Zhang:** Conceptualization, Methodology, Writing – review & editing. **Yonggang Fu:** Conceptualization, Methodology, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Table 7
Hyper-parameter settings.

Model	Parameter	Option range
LSTM GRU ED Bi-LSTM Dense	Hidden size	$\{2^4, 2^5, 2^6\}$
CNN	Kernel size	3–9 (2 per step)
	Out Channels	$\{2^2, 2^3, 2^4, 2^5, 2^6\}$
CRNN CRNN-Res HCLSTM	Kernel size	3–9 (2 per step)
	CNN out channels	$\{2^2, 2^3, 2^4, 2^5, 2^6\}$
	Hidden size	$\{2^4, 2^5, 2^6\}$
CRNN-Res	Residual window size	1–5 (1 per step)
TPA-LSTM	Kernel size	3–9 (2 per step)
	CNN out channels	$\{2^2, 2^3, 2^4, 2^5, 2^6\}$
LSTNet	GRU hidden size	$\{2^4, 2^5, 2^6\}$
MTNet	The number of GRU layers	1–3 (1 per step)
	Highway window size	1–10 (1 per step)
LSTNet	Skip window size	1–3 (1 per step)
	Skip GRU hidden size	$\{2^4, 2^5, 2^6\}$
MTNet	Block size	1–10 (1 per step)
MSL	Shapelet size	$\{2^2, 2^3, 2^4, 2^5, 2^6\}$
Informer	Encoder layers	1–3 (1 per step)
	Decoder layers	1–3 (1 per step)
	The numbers of heads	$\{2^2, 2^3, 2^4\}$
	The label length	1–10 (1 per step)
	The dimension of the model	$\{2^4, 2^5, 2^6\}$
GAIN	GAT hidden size	$\{2^4, 2^5, 2^6\}$
	The number of heads of GAT	$\{2^0, 2^1, 2^2, 2^3, 2^8\}$
	GRU hidden size	$\{2^4, 2^5, 2^6\}$
	Kernel size	3–9 (2 per step)
	CNN out channels	$\{2^2, 2^3, 2^4, 2^5, 2^6\}$
	Highway window size	1–10 (1 per step)

Table 8
Abbreviations and meanings.

Abbreviation	Notation
ARC	Autoregressive Representation Concatenation
ARIMA	Autoregressive Integrated Moving Average
ARX	Autoregressive Model with Exogenous Inputs
CHP	Combined Heat And Power Generation
CNN	Convolution Neural Network
CRNN	Convolution Recurrent Neural Network
CTGAT	Collaborative Temporal Graph Attention
CV-RMSE	Coefficient of Variation of Root Mean Square Error
DBA	DTW Barycenter Averaging
DNN	Deep Neural Network
DTW	Dynamic Time Warping
ED	Encoder–Decoder
ETR	Enhanced Temporal Representation
FNN	Feedforward Neural Network
GNN	Graph Neural Network
GRU	Gated Recurrent Unit
IPSO	Improved Particle Swarm Optimization
LASSO	Least Absolute Shrinkage and Selection Operator
LSTM	Long Short-term Memory
MAE	Mean Absolute Error
MLR	Multiple Linear Regression
MSE	Mean Square Error
MSL	Multivariate Shapelet Learning
OT	Outdoor Temperature
PSO	Particle swarm optimization
PSO-NN	Particle Swarm Optimization Neural Network
RH	Relative Humidity
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RT	Regression Trees
SRI	Solar Radiation Intensity
SVM	Support Vector Machine
TGAT	Temporal Graph Attention
WS	Wind Speed
XGBoost	Extreme Gradient Boosting

Acknowledgments

This research was supported in part by the Fujian Province Natural Science Foundation of Fujian Province (CN) (nos. 2021J01857, 2021J01859, and 2022J01335), and the EMB3Rs project (no. 84712) funded by the European Union Horizon 2020 research and innovation programme.

We thank the editors and anonymous reviewers for their valuable comments and suggestions, which greatly benefited this paper.

Appendix

A.1. Additional experiments

The effectiveness of the proposed method was also validated using the JERICHO-E-usage dataset [63], which available at https://doi.org/10.6084/m9.figshare.c.5245457.v1. This dataset comprises hourly energy consumption patterns from 38 regions in Germany throughout the year of 2019, with district space heating consumption data categorized by different energy types (e.g., residential, industrial, commerce) used for experiments. For an additional experiment, we selected five prediction methods for heat load forecasting, including Dense, BiLSTM, HCLSTM, Informer, and MTNet, as comparable models. The first four methods have been used in previous studies for heat load prediction. The data resolution was aggregated from hourly to daily in consistent with the district heating data in Aalborg. We separate 70% data for training, 10% data for validation, and the remaining 20% data for test.

The experimental results are presented in Table 6, that evaluates the proposed method across four different time horizons using MAE and RMSE. As can be seen, Informer demonstrates more accurate performance than other benchmark methods, suggesting the robustness of Informer in datasets with a small number of samples. Dense achieves promising accuracy in all three category datasets when $h = 1$. However, as the horizon increased, the performance of Dense decreased significantly. This result reflects the limitation of Dense in capturing the temporal dynamics across time steps. The methods BiLSTM, HCLSTM, and MTNet exhibit lower accuracy compared to the other methods. This can potentially be attributed to the fact that these models may capture insufficient temporal representations due to the limited amount of data, which can lead to poor fitting performance. GAIN outperforms other comparable methods in most cases, which further demonstrates the effectiveness and generalization performance of GAIN on different heat load prediction tasks.

A.2. Hyper-parameter setting

The parameters setting of the proposed method and benchmarks are listed in Table 7.

A.3. Abbreviation

The meanings of abbreviations are listed in Table 8.

References

[1] Gholamian E, Zare V, Javani N, Ranjbar F. Dynamic 4E (energy, exergy, economic and environmental) analysis and tri-criteria optimization of a building-integrated plant with latent heat thermal energy storage. Energy Convers Manage 2022;267:115868. http://dx.doi.org/10.1016/j.enconman.2022.115868.

[2] Forster PM, Maycock AC, McKenna CM, Smith CJ. Latest climate models confirm need for urgent mitigation. Nature Clim Change 2020;10(1):7–10. http://dx.doi.org/10.1038/s41558-019-0660-0.

[3] Li W, Zhang S, Lu C. Exploration of China's net CO2 emissions evolutionary pathways by 2060 in the context of carbon neutrality. Sci Total Environ 2022;831:154909. http://dx.doi.org/10.1016/j.scitotenv.2022.154909.

[4] Guan Z, Zhao P, Wang J, Sun Q, Guo Y, Lou J, et al. Dynamic performance and control strategy of a combined heat and power system driven by geothermal energy considering the building multi-load requirements. Energy Convers Manage 2022;270:116189.

[5] Dalipi F, Yildirim Yayilgan S, Gebremedhin A. Data-driven machine-learning model in district heating system for heat load prediction: A comparison study. Appl Comput Intell Soft Comput 2016;2016:3403150. http://dx.doi.org/10.1155/2016/3403150.

[6] Gong M, Wang J, Bai Y, Li B, Zhang L. Heat load prediction of residential buildings based on discrete wavelet transform and tree-based ensemble learning. J Build Eng 2020;32:101455. http://dx.doi.org/10.1016/j.jobe.2020.101455.

[7] Shamshirband S, Petković D, Enayatifar R, Abdullah AH, Marković D, Lee M, et al. Heat load prediction in district heating systems with adaptive neuro-fuzzy method. Renew Sustain Energy Rev 2015;48:760–7. http://dx.doi.org/10.1016/j.rser.2015.04.020.

[8] Kuan L, Zhenfu B, Xin W, Xiangrong M, Honghai L, Wenxue S, et al. Short-term CHP heat load forecast method based on concatenated LSTMs. In: Proceedings of the 2017 Chinese automation congress. Jinan, China: IEEE; 2017, p. 99–103. http://dx.doi.org/10.1109/CAC.2017.8242744.

[9] Adamski M, Ruszczyk J. New weather controlled central heating system. Ciepłownictwo Ogrzewnictwo Wentylacja 2012;43:278–83.

[10] Kapalo P, Adamski M. The analysis of heat consumption in the selected city. In: Proceedings of the 1st international scientific conference ecocomfort and current issues of civil engineering. Lviv, Ukraine: Springer; 2020, p. 158–65. http://dx.doi.org/10.1007/978-3-030-57340-9_20.

[11] Guelpa E, Marincioni L. Demand side management in district heating systems by innovative control. Energy 2019;188:116037. http://dx.doi.org/10.1016/j.energy.2019.116037.

[12] Wang Y, Li Z, Liu J, Zhao Y, Sun S. A novel combined model for heat load prediction in district heating systems. Appl Therm Eng 2023;227:120372. http://dx.doi.org/10.1016/j.applthermaleng.2023.120372.

[13] Sakkas NP, Abang R. Thermal load prediction of communal district heating systems by applying data-driven machine learning methods. Energy Rep 2022;8:1883–95. http://dx.doi.org/10.1016/j.egyr.2021.12.082.

[14] Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The graph neural network model. IEEE Trans Neural Netw 2008;20(1):61–80. http://dx.doi.org/10.1109/TNN.2008.2005605.

[15] Zhang Q, Tian Z, Ma Z, Li G, Lu Y, Niu J. Development of the heating load prediction model for the residential building of district heating based on model calibration. Energy 2020;205:117949. http://dx.doi.org/10.1016/j.energy.2020.117949.

[16] Stoffel P, Maier L, Kümpel A, Schreiber T, Müller D. Evaluation of advanced control strategies for building energy systems. Energy Build 2023;280:112709. http://dx.doi.org/10.1016/j.enbuild.2022.112709.

[17] Zhao A, Mi L, Xue X, Xi J, Jiao Y. Heating load prediction of residential district using hybrid model based on CNN. Energy Build 2022;266:112122. http://dx.doi.org/10.1016/j.enbuild.2022.112122.

[18] Thilker CA, Bacher P, Bergsteinsson HG, Junker RG, Cali D, Madsen H. Non-linear grey-box modelling for heat dynamics of buildings. Energy Build 2021;252:111457. http://dx.doi.org/10.1016/j.enbuild.2021.111457.

[19] Yang X, Liu S, Zou Y, Ji W, Zhang Q, Ahmed A, et al. Energy-saving potential prediction models for large-scale building: A state-of-the-art review. Renew Sustain Energy Rev 2022;156:111992. http://dx.doi.org/10.1016/j.rser.2021.111992.

[20] Zhu J, Dong H, Zheng W, Li S, Huang Y, Xi L. Review and prospect of data-driven techniques for load forecasting in integrated energy systems. Appl Energy 2022;321:119269. http://dx.doi.org/10.1016/j.apenergy.2022.119269.

[21] Wang Y, Li Z, Liu J, Zhao Y, Sun S. A novel combined model for heat load prediction in district heating systems. Appl Therm Eng 2023;227:120372. http://dx.doi.org/10.1016/j.applthermaleng.2023.120372.

[22] Song J, Xue G, Pan X, Ma Y, Li H. Hourly heat load prediction model based on temporal convolutional neural network. IEEE Access 2020;8:16726–41. http://dx.doi.org/10.1109/ACCESS.2020.2968536.

[23] Yuan J, Huang K, Lu S, Zhang J, Han Z, Zhou Z. Analysis of influencing factors on heat consumption of large residential buildings with different occupancy rates-Tianjin case study. Energy 2022;238:121834. http://dx.doi.org/10.1016/j.energy.2021.121834.

[24] Lu C, Li S, Lu Z. Building energy prediction using artificial neural networks: A literature survey. Energy Build 2022;262:111718. http://dx.doi.org/10.1016/j.enbuild.2021.111718.

[25] Werner S. The heat load in district heating systems [Ph.D. thesis], Chalmers tekniska högskola; 1984.

[26] Yuan J, Zhou Z, Tang H, Wang C, Lu S, Han Z, et al. Identification heat user behavior for improving the accuracy of heating load prediction model based on wireless on-off control system. Energy 2020;199:117454. http://dx.doi.org/10.1016/j.energy.2020.117454.

[27] Khalil M, McGough AS, Pourmirza Z, Pazhoohesh M, Walker S. Machine learning, deep learning and statistical analysis for forecasting building energy consumption — A systematic review. Eng Appl Artif Intell 2022;115:105287. http://dx.doi.org/10.1016/j.engappai.2022.105287.

[28] Torres JF, Hadjout D, Sebaa A, Martínez-Álvarez F, Troncoso A. Deep learning for time series forecasting: A survey. Big Data 2021;9(1):3–21. http://dx.doi.org/10.1089/big.2020.0159.

[29] Bujalski M, Madejski P, Fuzowski K. Day-ahead heat load forecasting during the off-season in the district heating system using generalized additive model. Energy Build 2023;278:112630. http://dx.doi.org/10.1016/j.enbuild.2022.112630.

[30] Jagait RK, Fekri MN, Grolinger K, Mir S. Load forecasting under concept drift: Online ensemble learning with recurrent neural network and ARIMA. IEEE Access 2021;9:98992–9008. http://dx.doi.org/10.1109/ACCESS.2021.3095420.

[31] Cui M. District heating load prediction algorithm based on bidirectional long short-term memory network model. Energy 2022;254:124283. http://dx.doi.org/10.1016/j.energy.2022.124283.

[32] Ding Y, Timoudas TO, Wang Q, Chen S, Bratteb H, Nord N. A study on data-driven hybrid heating load prediction methods in low-temperature district heating: An example for nursing homes in nordic countries. Energy Convers Manage 2022;269:116163. http://dx.doi.org/10.1016/j.enconman.2022.116163.

[33] Song J, Zhang L, Xue G, Ma Y, Gao S, Jiang Q. Predicting hourly heating load in a district heating system based on a hybrid CNN-LSTM model. Energy Build 2021;243:110998. http://dx.doi.org/10.1016/j.enbuild.2021.110998.

[34] Gong M, Zhao Y, Sun J, Han C, Sun G, Yan B. Load forecasting of district heating system based on informer. Energy 2022;253:124179. http://dx.doi.org/10.1016/j.energy.2022.124179.

[35] Zhou H, Zhang S, Peng J, Zhang S, Li J, Xiong H, et al. Informer: Beyond efficient transformer for long sequence time-series forecasting. In: Proceedings of the 35th AAAI conference on artificial intelligence. Virtual Event: AAAI Press; 2021, p. 11106–15. http://dx.doi.org/10.1609/aaai.v35i12.17325.

[36] Wang C, Wang Y, Ding Z, Zheng T, Hu J, Zhang K. A transformer-based method of multienergy load forecasting in integrated energy system. IEEE Trans Smart Grid 2022;13(4):2703–14. http://dx.doi.org/10.1109/TSG.2022.3166600.

[37] Hu Y, Cheng X, Wang S, Chen J, Zhao T, Dai E. Times series forecasting for urban building energy consumption based on graph convolutional network. Appl Energy 2022;307:118231. http://dx.doi.org/10.1016/j.apenergy.2021.118231.

[38] Wu Q, Zheng H, Guo X, Liu G. Promoting wind energy for sustainable development by precise wind speed prediction based on graph neural networks. Renew Energy 2022;199:977–92. http://dx.doi.org/10.1016/j.renene.2022.09.036.

[39] Shahzad MW, Burhan M, Ng KC. A standard primary energy approach for comparing desalination processes. NPJ Clean Water 2019;2(1):1. http://dx.doi.org/10.1038/s41545-018-0028-4.

[40] Ng KC, Burhan M, Chen Q, Ybyraiymkul D, Akhtar FH, Kumja M, et al. A thermodynamic platform for evaluating the energy efficiency of combined power generation and desalination plants. Npj Clean Water 2021;4(1):25. http://dx.doi.org/10.1038/s41545-021-00114-5.

[41] Schaffer M, Tvedebrink T, Marszal-Pomianowska A. Three years of hourly data from 3021 smart heat meters installed in danish residential buildings. Sci Data 2022;9(1):1–13. http://dx.doi.org/10.1038/s41597-022-01502-3.

[42] Ding Y, Zhang Q, Yuan T, Yang K. Model input selection for building heating load prediction: A case study for an office building in Tianjin. Energy Build 2018;159:254–70. http://dx.doi.org/10.1016/j.enbuild.2017.11.002.

[43] Ling J, Dai N, Xing J, Tong H. An improved input variable selection method of the data-driven model for building heating load prediction. J Build Eng 2021;44:103255. http://dx.doi.org/10.1016/j.jobe.2021.103255.

[44] Wang Y, Liu K, Liu Y, Wang D, Liu J. The impact of temperature and relative humidity dependent thermal conductivity of insulation materials on heat transfer through the building envelope. J Build Eng 2022;46:103700. http://dx.doi.org/10.1016/j.jobe.2021.103700.

[45] Zhao H, Wang Y, Duan J, Huang C, Cao D, Tong Y, et al. Multivariate time-series anomaly detection via graph attention network. In: Proceedings of the 20th IEEE international conference on data mining. Sorrento, Italy: IEEE; 2020, p. 841–50. http://dx.doi.org/10.1109/ICDM50108.2020.00093.

[46] Petitjean F, Ketterlin A, Gançarski P. A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognit 2011;44(3):678–93. http://dx.doi.org/10.1016/j.patcog.2010.09.013.

[47] Cai X, Xu T, Yi J, Huang J, Rajasekaran S. DTWNet: A dynamic time warping network. In: Proceedings of the 33rd international conference on advances in neural information processing systems. Vancouver, BC, Canada: Curran Associates, Inc.; 2019, p. 11636–46.

[48] Wang Z, Cai B. COVID-19 cases prediction in multiple areas via shapelet learning. Appl Intell 2022;52(1):595–606. http://dx.doi.org/10.1007/s10489-021-02391-6.

[49] Józefowicz R, Zaremba W, Sutskever I. An empirical exploration of recurrent network architectures. In: Proceedings of the 32nd international conference on machine learning. Lille, France: PMLR; 2015, p. 2342–50.

[50] Cho K, van Merrienboer B, Bahdanau D, Bengio Y. On the properties of neural machine translation: Encoder-decoder approaches. In: Proceedings of the 8th workshop on syntax, semantics and structure in statistical translation. Doha, Qatar: Association for Computational Linguistics; 2014, p. 103–11. http://dx.doi.org/10.3115/v1/W14-4012.

[51] Lim B, Zohren S. Time-series forecasting with deep learning: A survey. Proc R Soc Lond Ser A Math Phys Eng Sci 2021;379(2194):20200209. http://dx.doi.org/10.1098/rsta.2020.0209.

[52] Wu Y, Yang Y, Nishiura H, Saitoh M. Deep learning for epidemiological predictions. In: Proceedings of the 41st international conference on research and development in information retrieval. Ann Arbor, MI, USA: ACM; 2018, p. 1085–8. http://dx.doi.org/10.1145/3209978.3210077.

[53] Shih S, Sun F, Lee H. Temporal pattern attention for multivariate time series forecasting. Mach Learn 2019;108(8–9):1421–41. http://dx.doi.org/10.1007/s10994-019-05815-0.

[54] Lai G, Chang W, Yang Y, Liu H. Modeling long- and short-term temporal patterns with deep neural networks. In: Proceedings of the 41st international conference on research and development in information retrieval. Ann Arbor, MI, USA: ACM; 2018, p. 95–104. http://dx.doi.org/10.1145/3209978.3210006.

[55] Chang Y, Sun F, Wu Y, Lin S. A memory-network based solution for multivariate time-series forecasting. 2018, CoRR abs/1809.02105.

[56] Kingma DP, Ba J. Adam: A method for stochastic optimization. In: Proceedings of the 3rd international conference on learning representations. San Diego, CA, USA; 2015.

[57] Dai Z, Liu H, Le QV, Tan M. CoAtNet: Marrying convolution and attention for all data sizes. In: Proceedings of the 35th international conference on advances in neural information processing systems. Online Event: Curran Associates, Inc.; 2021, p. 3965–77.

[58] Han S, Dong H, Teng X, Li X, Wang X. Correlational graph attention-based long short-term memory network for multivariate time series prediction. Appl Soft Comput 2021;106:107377. http://dx.doi.org/10.1016/j.asoc.2021.107377.

[59] Zhang T, Guo G. Graph attention LSTM: A spatiotemporal approach for traffic flow forecasting. IEEE Intell. Transp. Syst. Mag. 2022;14(2):190–6. http://dx.doi.org/10.1109/MITS.2020.2990165.

[60] Lumbreras M, Garay-Martinez R, Arregi B, Martin-Escudero K, Diarce G, Raud M, Hagu I. Data driven model for heat load prediction in buildings connected to district heating by using smart heat meters. Energy 2022;239:122318. http://dx.doi.org/10.1016/j.energy.2021.122318.

[61] Narvaez G, Giraldo LF, Bressan M, Pantoja A. Machine learning for site-adaptation and solar radiation forecasting. Renew Energy 2021;167:333–42. http://dx.doi.org/10.1016/j.renene.2020.11.089.

[62] Zhu H, Ren C, Cao S. Fast prediction for multi-parameters (concentration, temperature and humidity) of indoor environment towards the online control of HVAC system. Build Simul 2021;14(3):649–65. http://dx.doi.org/10.1007/s12273-020-0709-z.

[63] Priesmann J, Nolting L, Kockel C, Praktiknjo A. Time series of useful energy consumption patterns for energy system modeling. Sci Data 2021;8(1):148. http://dx.doi.org/10.1038/s41597-021-00907-w.