A Novel Hybrid Spatial-Temporal Attention-LSTM Model for Heat Load Prediction

TAO LIN\textsuperscript{1}, YU PAN\textsuperscript{1}, GUIXIANG XUE\textsuperscript{1}, JIANCAI SONG\textsuperscript{2},
AND CHENGYING QI\textsuperscript{3}

\textsuperscript{1}School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401, China
\textsuperscript{2}School of Information and Engineering, Tianjin University of Commerce, Tianjin 300134, China
\textsuperscript{3}School of Energy and Environment Engineering, Hebei University of Technology, Tianjin 300401, China

Corresponding authors: Guixiang Xue (xueguixiang_tju@163.com) and Jiancai Song (cn sjc@163.com)

This work was supported in part by the Ministry of Science and Technology of China under Grant 2016YFC070070702, in part by the National Natural Science Foundation of China under Grant 91746205 and Grant 61702157, and in part by the Tianjin Technical Expert Project of China under Grant 19JCTPJC51000.

ABSTRACT Accurate heat load prediction algorithm provides important support for the stable and efficient operation of smart district heating system (SDHS) and helps to realize energy saving and consumption reduction. However, previous researches on heat load prediction are mostly carried out on various regression analyses and modeling prediction, without considering the inherent time delay and spatial dependence between heat exchange stations during regulation. Therefore, a novel heat load prediction model based on the hybrid spatial-temporal attention long short-term memory (STALSTM) is proposed. The STALSTM model introduces the spatial dependence characteristics of heating pipe network into the heat load prediction model, and the influencing factors of heat consumption are considered comprehensively from the time and space dimensions. Then, the LSTM algorithm is used to memory the information of historical data sequence, and the attention mechanism is used to realize the adaptive estimation of the characteristic weight of each influencing factor, which improves the prediction accuracy. And in order to verify the effectiveness of the proposed model, a detailed experimental comparison is made between the STALSTM model and the state-of-the-art algorithms. The results show that the STALSTM model has the best prediction accuracy, and the correctness of introducing the spatial-temporal characteristics and the attention mechanism is also proved.

INDEX TERMS Spatial-temporal, attention mechanism, heat load prediction, district heating system.

I. INTRODUCTION

With the rapid development of urbanization in China, the scope of central heating is expanding. However, with the increase of the heating area, the imbalance between supply and demand gradually emerged. At present, many district heating systems (DHS) in China are still regulated according to human experience, and the problem of unbalanced supply and demand often occurs. Therefore, in the planning of sustainable development of urban energy system, in order to further improve the DHS, it is necessary to take into account the changes of heat load demand in the future [1]. The establishment of data-driven intelligent DHS (SDHS) will be the development trend in the future. SDHS effectively uses the Internet of things (IoT) sensors to complete the real-time collection of multi-level monitoring data, and combines the weather forecast to rationally improve energy utilization efficiency, so as to achieve on-demand heating and balanced heating. And accurate prediction of heat load on demand side is one of the core technologies for SDHS to be realized. The improvement of heat load prediction accuracy can ensure the comfort of users and improve energy utilization.

Heat load prediction has attracted much more attention of scholars, and the literatures about heat load prediction method have been constantly emerging. Most works mainly construct the prediction model from two aspects optimizing the model input and improving the prediction algorithm.

In Ref [2], ensemble weather predictions are introduced in the operation of DHS to create a heat load prediction with dynamic uncertainties. In Ref [3], indoor temperature is innovatively introduced. The historical heat consumption and indoor temperature are taken as the inputs of the model, which improves the accuracy of the model prediction results and provides a feasible method for heat load prediction. In Ref [4],
indoor temperature and building thermal inertia coefficient are introduced into the prediction model of secondary supply temperature. However, without increasing the burden of prediction model, it is also crucial to realize the reasonable selection of internal and external influencing factors. In Ref [5], scholars conduct multi-frequency analysis and correlation analysis on heat load samples to preliminarily extract internal and external influencing factors. Then, principal component analysis is introduced to avoid repeated selection of similar influencing factors. Although these prediction models prove to be effective to a certain degree, the improvement of model prediction performance by optimizing model input is limited. Therefore, many scholars turn to study better predictive algorithms and the improved algorithms to further optimize the prediction model.

In recent years, machine learning methods for energy demand estimation and prediction have attracted much attention due to their advantages over linear and nonlinear programming models. In Ref [6], a data-driven machine learning model is proposed to predict the heat load of buildings in DHS. The comparison of support vector regression (SVR) with partial least square method and random forest shows that the SVR method has the higher efficiency in heat load prediction. In Ref [7], particle swarm optimization (PSO) algorithm is used to improve SVR for heat load prediction. The results show that the model has good robustness and is applicable to the actual heating system. In Ref [8], a short-term multistep heat load prediction model based on extreme learning machine (ELM) is proposed. Through sufficient experimental comparison, it is proved that the ELM method can improve the prediction accuracy and generalization ability. In Ref [9], a multi-step forward heat load prediction framework based on machine learning is proposed. It integrated SVR, Deep neural net (DNN) and XGBoost to build prediction model.

Although the machine learning method and its improved algorithm have made some achievements in the field of heat load prediction, the traditional machine learning algorithm still has some limitations. Therefore, with the rapid development of deep learning technology, heat load prediction based on deep learning algorithm has attracted much attention. In Ref [10], DNN is used for heat load prediction. In this study, compared with other machine learning algorithms, DNN obtained the best prediction results. In the exploitation of deeper information, recurrent neural network (RNN) and long short-term memory (LSTM) have shown a higher capacity for analysis and processing. RNN and LSTM can automatically extract and memorize the deep features of data, which are suitable for time series analysis and prediction and can achieve high prediction accuracy [11]. In Ref [12], RNN is used for heat load prediction of district heating and cooling systems, which can obtain a lower mean square error under non-stationary conditions. As a variant of RNN, LSTM has been widely used in time series prediction [13]–[15], and its prediction performance is better than traditional methods. LSTM has a unique gate structure, which can avoid the gradient disappearance of RNN. Therefore, we conduct a deeper study on LSTM.

However, with the increase of sequence length, the errors in LSTM accumulate gradually, which makes it difficult to process long-term time series data. Attention mechanism (AM) can solve the problem of LSTM to obtain effective information between long-term time series [16]. The core idea of AM is to dynamically adjust the weights between different factors, so as to rationally change the model’s attention to data, and ignore irrelevant information. And AM has gradually become a research hotspot in the field of deep learning. Some existing studies have applied AM to prediction, such as traffic flow prediction [17], [18], subway passenger flow prediction [19], and power load prediction [20], [21], and achieved good prediction results. This provides a theoretical guide for introducing the AM into the research of heat load prediction.

In the whole DHS, the pipe pressure and heat transfer efficiency of heat exchange stations are different, and the heat load between adjacent heat exchange stations is mutually affected. In Ref [22], authors presented the relationship between the thermal inertia of the pipe network and other parameters, such as pipe network structure, water supply temperature, return water temperature, circulating water flow rate, operation mode, and so on. It proves that pipe diameter, water flow rate, and pipe length have a great impact on the lag time. Therefore, in the heating network, there is an influence between the components. And it is not enough to only consider the information of the time dimension. But previous researches on heat load prediction are majority focused on the characteristics of single heat exchange station, and ignored the spatial dependence of adjacent heat exchange stations on pipeline topology. Fully considering the influencing factors of heat exchange station can further improve the accuracy of the prediction model. Therefore, the spatial characteristics of heat exchange station are introduced in this paper. In related fields, some scholars have introduced spatial characteristics into prediction researches and obtained good prediction results, which have great reference value. It mainly involves radar signal [23], weather information [24], wind speed [25] and traffic flow [26], [27] prediction. It provides that the introduction of spatial characteristics is helpful to improve the prediction performance of the model, and provides a theoretical basis for introducing spatial characteristics into the study of heat load prediction.

In this paper, a hybrid spatial-temporal attention-LSTM (STALSTM) model for heat load prediction is proposed. The contributions of this paper mainly include the following three aspects.

1) The spatial characteristics of heat load prediction are introduced. The heat load of heat exchange station is not only affected by its own operating parameters, but also affected by adjacent heat exchange stations. Therefore, in the STALSTM model, the heat exchange station adjacent to the target heat exchange station in the network topology is introduced as a
spatial feature, so as to improve the accuracy of heat load prediction.

(2) This paper establishes a mathematical model of spatial-temporal combination, and analyses the characteristic factors of heat load prediction from two dimensions of time and space, so that different spatial dependence and time delay characteristics can be accurately considered.

(3) The attention mechanism is used to optimize the LSTM model. It can help LSTM solve the problem that errors accumulate gradually with the increase of sequence length and automatically calculate the attention weight of each influencing factor from time and space dimension. Attention-LSTM can pay close attention to the key areas of all information quickly, and improve the accuracy of heat load prediction.

The rest of this paper is organized as follows. Section 2 is a description of the data preprocessing and feature selection. Section 3 represents the methodology which proposed in this paper. Section 4 is the experimental analysis and evaluation of predictive performance. Section 5 is the conclusion.

II. DATA PREPROCESSING AND FEATURE SELECTION

In this section, data preprocessing and feature selection of heat load data are carried out. Firstly, DHS is introduced, and the importance of secondary side heating system in heat load prediction is explained. Secondly, considering the spatial location relationship between heat exchange stations, and taking the general spatial relationship as the research object, the data of three adjacent heat exchange stations are selected for research. Thirdly, the missing values and outliers of the original data are processed. Finally, the correlation analysis of the processed data is carried out, and the factors whose absolute correlation value is greater than 0.5 are selected as the input of the prediction model.

A. BACKGROUND

DHS is an important public infrastructure in northern cities of China, which consists of heat source, heat exchange station, heating pipe network and users. With the rapid development of information technology, SDHS has gradually become a trend. Traditional DHS can be monitored and controlled automatically and intelligently by installing various sensors and automation devices. As the key intermediate part of the whole heating network, heat exchange station mainly realizes energy transfer from primary side to secondary side with high efficiency. It is the key link to realize energy-saving operation of DHS. The classical process of heat exchange station in SDHS is shown in Fig. 1. The heat exchange station is equipped with temperature sensors, pressure gauges and heat meters, etc. The operation parameters of real-time monitoring include primary supply temperature, primary return temperature, secondary supply temperature, secondary return temperature and instantaneous flow rate, etc. Where, instantaneous flow rate represents the amount of fluid flowing through the effective section of the heating pipe network in an hour.

In the heating system, the secondary side heating system is closest to the user side, and the secondary return temperature can directly reflect the indoor temperature of the building, so managers usually take the secondary return temperature as a reference to control the heat energy supply. At the same time, the accurate prediction of the heat load on the demand side is of great help to the accurate adjustment of heat supply.
in the heat exchange station. And there is an interaction between the secondary side and the primary side. Therefore, to avoid repeated input of similar influencing factors, the external factors and influence factors of the secondary side are considered in the experiment.

In the DHS, the heating pipe network connects the heat exchange stations with the heat exchange stations, so there is interaction between adjacent heat exchange stations in space. Therefore, in order to fully mine the time lag and spatial dependence between the heat load data, this paper selects three adjacent heat exchange stations in the heating network of Xingtai City in northern China as the research object, and selects the data of two heating seasons in the time dimension. The real location relationship of the three adjacent heat exchange stations in the heating system is shown in Fig.2. Where, the heat exchange station B is located upstream of heat exchange station A, and heat exchange station C is located downstream of heat exchange station A. We take the heat exchange station A as the target ones, the heat exchange station B and C as the spatial influencing factors of heat exchange station A.

FIGURE 2. The real location relationship of the three adjacent heat exchange stations in the heating system.

B. DATA PREPROCESSING

In the actual data acquisition of heating monitoring system, there are outliers and missing values in the original data collected by SCADA system due to the influence of various uncertain factors such as sensor failure, network transmission failure and electromagnetic interference. And the measurement error of the sensor itself will also lead to the measurement noise in the data. Missing values will affect the continuity of time series, leading to the losses of useful information in the data, and when the noise data reaches a certain volume, it will affect the accuracy of prediction. Therefore, it is necessary to fill the missing data and smooth the noise data. In order to facilitate the model to process indicators of different units or magnitudes, and to be able to compare with other models, it is necessary to normalize the input data.

1) THE PREPROCESS OF ORIGINAL DATA

The data used in this paper is time series data, and there is a time-dependent relationship between the data, so the existence of missing values make the model unable to fully mine the characteristics of the data, resulting in the output of unreliable prediction results. Therefore, it is necessary to supplement the missing data in the data set. Hot deck imputation [28] are used to fill the missing data. For an object containing missing values, the method finds the most similar object in the data, and then fills the missing value with the value of the similar object. For the abnormal values in SCADA system, the 3σ rule is used to exclude the sample points which deviate from the standard deviation twice the sample mean from the data set, which assumes that the heat load and its influencing factors of the heating system obey the normal distribution. For the outliers contained in the original data, this paper uses gauss filtering to smooth the data. Gauss filtering is a weighted averaging of data. The filtering value of each data is obtained by weighted averaging of itself and other data in its neighborhood. By weighted averaging, high-frequency random measurement noise is suppressed to some extent, which reduces its impact on prediction accuracy.

2) THE NORMALIZATION

The normalization processing of data is to scale the input features with different value ranges to the same interval. Normalization is a key step of data preprocessing, and the deep learning optimization algorithm usually adopts this method. After the data is standardized, multiple indexes can be used to predict the heat load. Therefore, in order to facilitate the construction of the prediction model and the comparison of the prediction results, the sample data is normalized. The normalization method adopted in this paper is z-score normalization. Input features in different value ranges are scaled to a standard normal distribution with mean value 0 and variance 1, as shown in equation (1).

\[
\begin{align*}
    y_i &= \frac{x_i - \bar{x}}{s} \\
    \bar{x} &= \frac{1}{n} \sum_{i=1}^{n} x_i \\
    s &= \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}
\end{align*}
\]

where, \( x_i \) and \( y_i \) represent the original data and the normalized data, respectively; \( \bar{x} \) and \( s \) represent the mean value and standard deviation value, respectively; \( n \) represents the size of sample.

C. THE CORRELATION ANALYSIS AND FEATURE SELECTION

The influence of various factors on heat load is also different, so reasonable selection of influencing factors plays an important role in predicting accuracy. In this paper, Pearson correlation analysis method is used to analyze the correlation...
TABLE 1. The Pearson correlation coefficient between the predicted heat load of station A and the other influencing factors.

| heat exchange station | secondary supply temperature | secondary return temperature | historical heat load | instantaneous flow rate |
|-----------------------|------------------------------|------------------------------|----------------------|------------------------|
| Station A             | 0.9422                       | 0.9137                       | 0.9905               | 0.8461                 |
| Station B             | 0.9424                       | 0.8478                       | 0.9240               | 0.7701                 |
| Station C             | 0.8792                       | 0.8772                       | 0.0380               | -0.0226                |

FIGURE 3. The spatial and temporal relationships of the three heat exchange stations.

between various factors and heat load. Pearson correlation coefficient is a kind of linear correlation coefficient, which is used to reflect the degree of linear correlation of two variables. The factors with high correlation value are selected as the input of the model to achieve data dimension reduction. Pearson correlation coefficient between two variables is defined as equation (2).

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$  (2)

In equation (2), we define the Pearson correlation coefficients of two variables (X, Y) equal to the product of their covariance cov(X, Y) divided by their respective standard deviations \(\sigma_X \sigma_Y\), where, the \(\mu_X\) and \(\mu_Y\) are the average and standard deviation of X samples respectively. \(\mu_Y\) and \(\sigma_Y\) are the mean and standard deviation of Y samples respectively.

Table 1 lists the correlation values between the heat load of heat exchange station A and the other influencing factors of heat exchange station A, B and C. Where, the heat load of heat exchange station A is the prediction target. In order to reduce the dimension of input data, only the factors whose absolute value of correlation is greater than 0.5 are selected as the input of the model.

The correlation values of secondary supply temperature, secondary return temperature, historical heat load and instantaneous flow rate of heat exchange station A are higher than 0.8, so these four features are selected as the influencing factors. The correlation values between the four influencing factors of heat exchange station B and the heat load of heat exchange station A is also greater than 0.5, so these four factors are also selected as the influencing factors. However, the absolute values of the correlation of historical heat load and instantaneous flow rate of heat exchange station C are less than 0.5, so they are not selected as a feature. The other three influencing factors of heat exchange station C are taken as the input of the model. At the same time, outdoor temperature as one of the prediction features has a significant negative correlation with the predicted heat load of station A, and correlation values is \(-0.7565\). Because the three stations are distributed in very close geographical areas, the difference of outdoor temperature corresponding to the three stations can be neglected. Therefore, outdoor temperature is also selected as the input of the prediction model.

III. METHODOLOGY

In this section, the concept of spatial-temporal and the introduction of attention mechanism in STALSTM model are described respectively. A hybrid spatial-temporal mathematical model is established, and the input of the STALSTM model is obtained by combining the characteristics of time and space dimensions. Then the
working mechanism of STALSTM with attention mechanism is explained.

A. THE MATHEMATICAL MODEL OF HYBRID SPATIAL-TEMPORAL

The spatial-temporal relationship of the three heat exchange stations in the STALSTM prediction model is shown in Fig. 3.

Fig. 3 shows the spatial-temporal two-dimensional coordinate system, the horizontal axis represents the time stamp, and the vertical axis represents the spatial positional relationship. In the Fig. 3, the red circles represent the predicted heat load of heat exchange station A; the other blue circles represent the internal and external factors of the heat exchange station A; the purple arrows indicate that the influencing factors of other heat exchange stations are all used to predict the heat load of heat exchange station A.

According to the research content of this paper, the mathematical model of hybrid spatial-temporal is introduced in detail. Firstly, the input factors of the prediction model are explained and analyzed. All the influencing factors of heat exchange station A, B and C from time 1 to time $T$ can be expressed as:

$$
\begin{align}
A &= (x_A^1, x_A^2, \cdots, x_A^T) \\
B &= (x_B^1, x_B^2, \cdots, x_B^T) \\
C &= (x_C^1, x_C^2, \cdots, x_C^T)
\end{align}
$$

where, $x_A^1, x_B^1$ and $x_C^1$ represent all input factors of heat exchange station A, B and C from time 1 to time $T$ respectively. $x_A^T, x_B^T$ and $x_C^T$ represent the input factors of the heat exchange station A, B and C at time $t$ respectively. And $x_A^1, x_B^1$ and $x_C^1$ are all two-dimensional arrays, $x_A^T, x_B^T$ and $x_C^T$ are all one-dimensional arrays.

The heat load of heat exchange station A is affected not only by its own features, but also by the features of adjacent heat exchange stations in space. According to the correlation analysis of section 2.3, the influencing factors selected by each heat exchange station are different. The influencing factors of heat load of heat exchange station A include secondary supply temperature ($T_s$), secondary return temperature ($T_r$), historical heat load ($Q_h$), instantaneous flow rate ($F$), and outdoor temperature ($T_o$). The influencing factors of heat load at $t$ time in each heat exchange station are defined as follows.

$$
\begin{align}
A &= (T_S^1, T_R^1, Q_h^1, F^1, T_o^1) \\
B &= (T_S^1, T_R^1, Q_h^1, F^1, T_o^1) \\
C &= (T_S^1, T_R^1, Q_h^1, F^1, T_o^1)
\end{align}
$$

where, $T_S^1, T_R^1, Q_h^1, F^1$ and $T_o^1$ represent secondary supply temperature, secondary return temperature, historical heat load, instantaneous flow rate, and outdoor temperature of heat exchange station A at time $t$ respectively. Similarly, the meanings of five influencing factors in the $x_B^1$ equally correspond to the factors in the $x_A^1$. After the correlation analysis, the influencing factors at time $t$ in the $x_C^1$ include the secondary supply temperature $T_S^C$, the secondary return temperature $T_R^C$, and the outdoor temperature $T_O^C$.

The input of the proposed STALSTM model at time $t$ can be represented as the combination of features at time $t$ of three heat exchange stations, that is

$$x_t = (x_t^A, x_t^B, x_t^C)$$

where, $x_t$ represents the input vector of the STALSTM model at time $t$, and the dimensional of it is 13.

The predicted heat load at time $t$ of the heat exchange station A is used to predict the heat load at time $t$. The expression of it is

$$Q_t = f(x_{t-1}, x_{t-2}, \cdots, x_{t-t'})$$

where, $Q_t$ represents the predicted heat load at time $t$ of the heat exchange station A; $f$ represents the nonlinear prediction function; $x_{t-1}$ and $x_{t-t'}$ represents the spatial-temporal input vector of the STALSTM model from time $t - 1$ to time $t - t'$.

B. THE ATTENTION MECHANISM ENHANCEMENT MODEL

The introduction of time and space features can improve the prediction accuracy of the model, but the increase of input dimensions increases the burden of the model. Moreover, the errors generated by LSTM will accumulate with the increase of sequence length, so in the prediction task of multivariate time series, the standard LSTM cannot fully capture the different influences of time series on the target sequence [29]. Therefore, this paper introduces the attention mechanism to solve this problem of LSTM. The attention mechanism is to imitate the object recognition principle of human brain, which can quickly focus on the local information of interest. The framework of attention mechanism in STALSTM is shown in Fig. 4, which is composed of encoder and decoder. Encoder is used to extract features and analyze the similarity between different features and target predicted values. Decoder is used to decode the information and predict the heat load. The attention mechanism enhancement model can dynamically pay attention to the useful information of influencing factors for time and space dimension, which can improve the prediction accuracy of heat load.

In Fig. 4, the input $x_t$ to $x_T$ of the encoder represents the inputs of the prediction model. The inputs are the combination of time and space features, as shown in equation (5).

$$x_t = (x_t^A, x_t^B, x_t^C)$$

where, $h_t$ represents the input composed of the influencing factors of the three heat exchange stations at time $t$. The output of hidden layer of the encoder at time $t$ can be obtained by equation (7), that is

$$h_t = LSTM(x_t, h_{t-1})$$

where, LSTM is the nonlinear activation function selected in this paper, which uses gating logic to control the memory of information [30]. The LSTM includes three gates, where a forget gate that determines the information to be discarded; an input gate that determines the new information to be
saved; and an output gate that determines what information to output to the next layer [31]. The three gate structures of LSTM promote the long-term memory regulation. And the unique gate structure can avoid the gradient disappearance of RNN [32]. However, as the length of the sequence increases, the performance of the model decreases rapidly. When LSTM processes a long information sequence, the encoder can only calculate a uniform intermediate variable, but cannot independently determine the influence degree of the information at each moment on the future. After introducing the attention mechanism, it can concentrate on the differences of input features to better extract features when different aspects are considered [33]. In this way, the weights of the data at different historical moments are different when the heat load prediction can be improved.

The attention weight obtained in the encoder is calculated by the hidden layer output \( h_t \) of the encoder at the current moment and the hidden layer output \( s_{t-1} \) of the decoder at the previous moment. The attention weight is

\[
\alpha_{t'} = V^T \tanh(W_s s_{t-1} + W_b h_t + b) \tag{8}
\]

where, \( V, W_s, W_b \) and \( b \) are model parameters that can be learned, \( \alpha_{t'} \) represents the attention weight. And the \( \alpha_{t'} \) is the correlation value between the currently input \( x_t \) and the predicted heat load.

Convert the resulting attention weight \( \alpha_{t'} \) through softmax operations to the probability \( \theta_{t'} \). When \( t' \) is given, the value of weight \( \theta_{t'} \) is a probability distribution at time \( t = 1, \ldots, T \). That is

\[
\theta_{t'} = \text{softmax}(\alpha_{t'}) = \frac{\exp(\alpha_{t'})}{\sum_{k=1}^{T} \exp(\alpha_{t'_k})} \tag{9}
\]

The background vector of the decoder at time step \( t' \) is the weighted average of the all outputs of the encoder hidden layers. Equation (10) shows that the decoder will use a variable background vector at each time step.

\[
c_{t'} = \sum_{t=1}^{T} \theta_{t'} h_t \tag{10}
\]

The length of background vector \( c_{t'} \) is equal to the number of hidden units in the encoder.

The output \( s_{t'} \) of the decoder hidden layer is updated by the output \( y_{t'-1} \) of the decoder at time \( t' - 1 \), the hidden layer output \( s_{t'-1} \) at time \( t' - 1 \), and the background vector \( c_{t'} \) at time \( t' \). The nonlinear activation function selected by the decoder and encoder is LSTM.

\[
s_{t'} = \text{LSTM}(y_{t'-1}, c_{t'}, s_{t'-1}) \tag{11}
\]

The final predicted value \( y_{t'} \) can be obtained according to the background vector \( c_{t'} \) and the updated hidden layer output \( s_{t'} \) of the decoder at time \( t' \).

\[
y_{t'} = V_o^T (W_w(s_{t'}, c_{t'}) + b_w) + b_o \tag{12}
\]

where, \( V_o, W_w, b_w \) and \( b_o \) are model parameters that can be learned. The final output \( y_{t'} \) of the model is also the final predicted heat load \( Q_t \).

**IV. EXPERIMENT AND DISCUSSION**

In this section, sufficient experiments are carried out. Firstly, the data used in the experiment are described in detail. Secondly, the hyperparameters of STALSTM model are selected. Finally, the prediction performance of STALSTM is compared with the state-of-the-art algorithms.

**A. CASE STUDY AND DATA DESCRIPTION**

In order to verify the performance of the proposed STALSTM heat load prediction model, this paper uses a real intelligent DHS in Xingtai City, northern China. The DHS has two distributed heat sources and more than 200 heat exchange stations. The author’s team has built a four-stage SCADA system covering heat sources, heat exchange stations, buildings and households. The real-time data acquisition frequency of heat
TABLE 2. The Initial values of hyperparameters.

| Parameter         | Value | Description                  |
|-------------------|-------|------------------------------|
| learning rate     | 0.0001| the value of the learning rate|
| hidden unit       | 15    | the number of hidden cells    |
| batch size        | 10    | the value of the batch size   |
| time step         | 10    | the value of the time step    |
| iterations        | 200   | the number of iterations      |

exchange station is every 20 seconds, and the storage period of historical data persistence is every 10 minutes.

In the effort to achieve the coordination of accurate prediction of heat load and optimal regulation of energy saving, heat load prediction based on hour dimension is more convenient to accurately analyze the non-linear relationship between heat load and its influencing factors.

In this paper, three adjacent heat exchange stations are studied. Among them, heat exchange station A is the predicted target heat exchange station, and adjacent heat exchange station B and C are the spatial influencing factors of heat exchange station A. They are located upstream and downstream of heat exchange station A, respectively. The connecting pipes of the three heat exchange stations are closely adjacent in spatial dimension. In the time dimension, the data of 2017-2018 and 2018-2019 heating seasons of three adjacent heat exchange stations are selected, totaling 4750 pieces, and the first 80% of the data are taken as the training set, and the remaining 20% are taken as the test set.

B. EVALUATION INDEX

In this paper, several index functions are used to evaluate the prediction performance of STALSTM model, such as root mean square error (RMSE), mean absolute error (MAE), determination coefficient ($R^2$) and mean absolute percentage error (MAPE). They can be defined as follows:

$$\begin{align*}
RMSE &= \sqrt{\frac{1}{n} \sum_{i=1}^{n} \frac{P_i - O_i}{P_i}} \\
MAE &= \frac{1}{n} \sum_{i=1}^{n} |P_i - O_i| \\
R^2 &= \left[ \frac{\sum_{i=1}^{n} \frac{(P_i - \bar{P})(O_i - \bar{O})}{\sum_{i=1}^{n}(P_i - \bar{P})^2}}{\sum_{i=1}^{n}(P_i - \bar{P})^2} \right] \\
MAPE &= \frac{100}{n} \sum_{i=1}^{n} \frac{P_i - O_i}{P_i}
\end{align*}$$

(13)

where, $O_i$ and $P_i$ represent the actual and predicted values of heat load, respectively. And $n$ represents the size of sample.

C. THE OPTIMAL SELECTION OF THE HYPERPARAMETERS OF STALSTM

The value of hyperparameters has a great impact on the prediction performance of the proposed model, so it is necessary to select the hyperparameters before building the model. The first step of the experiment in this paper is to carry out the hyperparameters tuning to avoid the improper selection of hyperparameters which may degrade the performance of the model.

The initial values of hyperparameters are listed in Table 2, and on this basis, parameters are selected one by one.

In the experiment, according to the order of parameters in the Table 2, the learning rate, hidden unit, batch size, time step and iterations are selected respectively. In order to avoid mutual interference between multiple parameters, this paper adjusts parameters based on the principle of single variable. Select the hyperparameters in the order of the parameters in Table 2. Set several different values of the hyperparameter without changing other parameter values, and then take the parameter value with the best result as the final hyperparameter value. For example, when selecting the number of hidden unit after completing the selection of learning rate, the value of learning rate uses the previously selected value. When selecting subsequent parameters, the values of other parameters that have been previously selected are also used.

Table 3 lists the performance indexes (include RMSE, MAE, $R^2$, and MAPE) of STALSTM model in training set and test set respectively under different values of hyperparameters. When determining the learning rate parameter value, the other four hyperparameters are fixed. We choose the best learning rate value of MAPE index as the value of this parameter in the subsequent experiments. As can be seen from Table 3, when the learning rate is 0.002, MAPE shows the best performance. Therefore, in the experiment of selecting the other four hyperparameters, the learning rate uses 0.002 to replace the previous initial value of 0.0001. Similarly, the MAPE value of the test set is still used as the selection basis in the subsequent experiments. The parameter value corresponding to the smallest MAPE is the final hyperparameter value.

The final hyperparameters selection is shown in Table 4. In order to ensure the consistency of the experiment, the values are selected as the hyperparameters in the subsequent relevant experiments which are shown in Table 4.

D. COMPARISON ANALYSIS WITH THE STATE-OF-THE-ART ALGORITHMS

To demonstrate the effectiveness of the introduction of attention mechanism and the combination of space and time, STALSTM is compared with the attention-LSTM model based on spatial characteristics (SALSTM), the attention-LSTM model based on temporal characteristics (TALSTM),
TABLE 3. Performance comparison of STALSTM algorithm with different hyperparameters values.

| Parameters Set | Turning Parameter | Value | Train | Test |
|----------------|-------------------|-------|-------|------|
|                |                   |       | RMSE  | MAE  | R²   | MAPE | RMSE  | MAE  | R²   | MAPE  |
| hidden unit = 15 | learning rate     | 0.0001| 0.0633| 0.0428| 0.9849| 0.0162| 0.0968| 0.0806| 0.9853| 0.0343|
| batch size = 10   |                   | 0.0005| 0.0659| 0.0469| 0.9836| 0.0178| 0.0904| 0.0711| 0.9872| 0.0290|
| time step = 10    |                   | 0.001 | 0.0596| 0.0430| 0.9866| 0.0166| 0.0695| 0.0547| 0.9924| 0.0231|
| iterations = 20   |                   | 0.0002| 0.0644| 0.0439| 0.9844| 0.0180| 0.0682| 0.0518| 0.9927| 0.0205|
| learning rate = 0.002 | learning rate   | 0.0005| 0.1000| 0.0718| 0.9622| 0.0283| 0.1363| 0.1161| 0.9709| 0.0499|
| batch size = 10   | hidden unit       | 10    | 0.0753| 0.0456| 0.9786| 0.0166| 0.0742| 0.0510| 0.9914| 0.0206|
| time step = 10    |                   | 12    | 0.0895| 0.0576| 0.9698| 0.0270| 0.0685| 0.0521| 0.9926| 0.0239|
| iterations = 200  |                   | 15    | 0.0644| 0.0439| 0.9844| 0.0180| 0.0682| 0.0518| 0.9927| 0.0205|
| learning rate = 0.002 | hidden unit     | 20    | 0.0521| 0.0357| 0.9898| 0.0141| 0.0525| 0.0362| 0.9957| 0.0172|
| batch size = 10   |                   | 25    | 0.0692| 0.0506| 0.9819| 0.0188| 0.0867| 0.0688| 0.9882| 0.0281|

TABLE 4. The final value of the hyperparameters.

| Parameter     | Value | Description          |
|---------------|-------|----------------------|
| learning rate | 0.002 | the value of the learning rate |
| hidden unit   | 20    | the number of hidden cells |
| batch size    | 10    | the value of the batch size |
| time step     | 7     | the value of the time step |
| iterations    | 200   | the number of iterations |

and the hybrid spatial-temporal LSTM model (STLSTM). The data used in the SALSTM model is the data of the heat exchange stations A, B and C in the heating season of 2018-2019, totaling 2903 pieces; the TALSSTM uses the data of heat exchange station A in the heating season of 2017-2018 and 2018-2019, totaling 4750 pieces; the STLSTM does not introduce an attention mechanism, and the data used is consistent with STALSTM, which is the data of heat exchange stations A, B and C in the heating season of 2017-2018 and 2018-2019, totaling 4750 pieces. In the experiment, the first 80% of the data are taken as the training set, and the remaining 20% are taken as the test set.

The comparison results of the four models are shown in the Table 5. The predicted results of the four models in the training set and the test set are compared in the table. As can be seen from the Table 5, the STALSTM model shows a best predictive performance. The RMSE, MAE, R² and MAPE of it in the test set are all better than SALSTM, TALSTM and STLSTM. It can be analyzed that the introduction of attention mechanism enhances the learning ability of LSTM, which can learn useful information for a long-term time series. And the data of other adjacent heat exchange stations are introduced to modify the prediction results. The results of SALSTM are the worst, indicating that only increasing the spatial dimension would increase the burden of attention-LSTM. The results of STLSTM are also poor, indicating that only adding the dimensions of the data reduces the predictive power of the model. This proves the necessity of introducing attention mechanism. The prediction ability of TALSTM is inferior to that of STALSTM, which also proves that the introduction of spatial features can improve the prediction performance of the model. The results show that the attention mechanism and the combination of space and time can effectively improve the accuracy of heat load prediction.

As shown in Fig. 5, the predicted data of the next 48 hours are selected to compare the predicted performance of STLSTM, SALSTM, TALSTM and STLSTM. It can be seen from the Fig 5 (a) that the prediction results of STALSTM are superior to the other three models, and it is closer to the actual value. The most of predicted values of TALSTM are lower...
than the actual value. The predicted values of SALSTM fluctuate greatly, and much predicted values of it are lower than other algorithms. The predicted values of STLSTM fluctuate greatly in some time periods, which do not conform to the change law of actual value.

Fig. 5(b) shows the relative error between the predicted value and the actual value in the four models. The relative error of SALSTM and STLSTM fluctuates greatly, and the maximum relative error of SALSTM reaches 9%. The relative error of TALSTM is smaller, but the overall predicted values are poor. The relative error of STALSTM fluctuates less, most of which are lower than other algorithms.

From the two subfigures, it is obvious that the STALSTM has a better performance. And the combination of attention mechanism and spatial-temporal concept significantly improves the prediction performance for the heat load in SDHS.

To further validate the validity of the model, STALSTM is compared with the state-of-the-art algorithms, such as LSTM, back propagation neural network (BPNN), support vector regression (SVR), random forest regression (RFR), gradient boost decision tree (GBDT) and extreme tree regression (ETR).

- LSTM [32]: A variant of a recurrent neural network. Through the unique “gate” structure, it is possible to learn and remember information that is dependent on for a long time. The hyperparameters used in this algorithm are same as the STALSTM.
- BPNN [34]: A multilayer feed forward neural network is trained according to the error inverse propagation algorithm. BPNN has strong nonlinear mapping capabilities and a flexible network structure. In the experiment, the hidden layer of gradient descent cycle is 3, the number of iterations is 20000, and the learning rate is 0.002.
- SVR [35]: The SVR algorithm specially developed for the regression problem. It is based on statistical learning theory and structural risk minimization principles. And it not only consider the error approximation of the data in the sample, but also generalizes the model and is applicable to various regression problems. In this paper, the kernel function of SVR is radial basis function. Because radial basis function can map samples to higher space dimensionality, it is suitable for training and testing of medium data samples.
- RFR [6]: RFR consists of multiple decision trees, and there is no correlation between each tree in the forest. The final output of the model is determined by each decision tree in the forest. When dealing with regression problems, the final result is the average value of output by each decision tree.
- GBDT [36]: GBDT mainly combines the ideas of regression decision tree and boosting decision tree, and utilizes residual gradient to optimize the integration process of regression tree. GBDT can find distinguishing features and feature combinations in the data.
- ETR [37]: When constructing the split node of each tree, ETR will not arbitrarily select features, but first collect some features randomly, and then use information entropy or index to select the best node features.

It can be seen from Table 6, the STALSTM algorithm has best prediction performance, and followed by BPNN. RMSE, MAE, R², and MAPE of STALSTM are 0.0460, 0.0356, 0.9967, and 0.0147 respectively. After introducing temporal and spatial characteristics, STALSTM can still achieve best prediction performance compared with the state-of-the-art algorithms such as LSTM, SVR, RFR, GBDT and ETR. And the prediction performance of STALSTM algorithm is much better than other algorithms.

In order to further verify the performance of STALSTM algorithm, it will be compared with other state-of-the-art algorithms. Combined with hourly weather forecast data to predict heat load of the next 48 hours, and the comparison of STALSTM with other algorithms for predict results and relative errors are shown in Fig. 6.

It can be seen from the Fig.6(a) that the prediction results of LSTM and BPNN fluctuated greatly, and there are big gap between the predicted value and the actual value in some periods of time. The prediction results of LSTM and BPNN are poor. The prediction results of SVR, RFR and GBDT have the same variation trend, but are inconsistent with the change trend of the actual value. In the later period of prediction, the prediction results of ETR are poor and the error with the actual value is large. The predicted values of STALSTM fluctuate slightly around the actual value, but they are closest to the actual value.

It can be seen from Fig.6(b) that the relative error of STALSTM is less than 5%. And compared to other algorithms the error of STALSTM in the whole prediction process is very small. The relative error of BPNN is large. In the later period of prediction, the relative errors of LSTM and ETR are large. The relative errors of other algorithms in the process of prediction are small, but the fluctuation is large.

### TABLE 5. The Comparison of the predictive performance.

| Algorithm | Train | Test |
|-----------|-------|------|
|           | RMSE  | MAE  | R²  | MAPE | RMSE  | MAE  | R²  | MAPE |
| SALSTM    | 0.0938| 0.0665| 0.9731| 0.0291| 0.1547| 0.1240| 0.9624| 0.0550|
| TALSTM    | 0.0622| 0.0395| 0.9854| 0.0153| 0.0545| 0.0425| 0.9953| 0.0176|
| STLSTM    | 0.0490| 0.0375| 0.9909| 0.0136| 0.1525| 0.0942| 0.9632| 0.0348|
| STALSTM   | 0.0623| 0.0412| 0.9853| 0.0173| 0.0460| 0.0356| 0.9967| 0.0147|

T. Lin et al.: Novel Hybrid STALSTM Model for Heat Load Prediction
the prediction performance is not stable. It can be clearly seen that STALSTM has a better predictive performance compared with other algorithms.

E. COORDINATION OF HEAT LOAD PREDICTION AND ENERGY SAVING STRATEGY

Accurate heat load prediction algorithm provides theoretical guidance for optimal regulation and operation of heat exchange station, which is the basic guarantee for energy saving and consumption reduction. Heat load prediction algorithm is the core technology of SDHS. Based on the actual operation data of SDHS in Xingtai, this paper makes a thorough theoretical study and practical verification of the proposed STALSTM algorithm for heat load prediction. The detailed results show that STALSTM algorithm could pay attention to the most useful features from two dimensions of time and space, and effectively solves the intrinsic complex non-linear heat load prediction problem of DHS. The
FIGURE 6. The compare of STALSTM with other algorithms for (a) prediction results and (b) relative error.

application of STALSTM prediction algorithm will be real-
ized in the construction of SDHS in the future.

V. CONCLUSION

DHS is a non-linear system with wide spatial distribution of pipeline topology, and there is a time delay caused by building thermal inertia. In order to improve the accuracy of heat load prediction, a novel hybrid spatial-temporal attention-LSTM model is proposed in this paper. On the one hand, the time attention mechanism is used to accurately calculate the influencing of each factor on the prediction results at different times; on the other hand, the spatial attention mechanism is used to obtain the influencing of the characteristics of adjacent heat exchange stations on the heat load. By considering the combination of time and space, attention mechanism and the combination of internal and external features, the STALSTM model can dynamically focus on and learn the most critical features of all information, which can significantly improve the accuracy of heat load prediction. In order to prove the superiority of STALSTM algorithm, this paper compares STALSTM with various state-of-the-art algorithms in detail, such as SALSTM, TALSTM, STLSTM, LSTM, BPNN, BPNN, SVR, GDBT, etc. Detailed comparison results show that the prediction performance of STALSTM model is significantly better than other models, and MAPE reaches 1.47%, which also proves the superiority of three principles in feature processing: spatial-temporal integration, the combination of internal and external features and the introduction of attention mechanism.

In the future work, we will continue to explore and improve the heat load prediction algorithm on the basis of considering heating comfort, and further optimize the energy-saving control strategy optimization based on artificial intelligence.

REFERENCES

[1] D. Koschwitz, E. Spinnräker, J. Frisch, and C. V. Treeck, “Long-term urban heating load predictions based on optimized retrofit orders: A cross-scenario analysis,” Energy Buildings, vol. 208, Feb. 2020, Art. no. 109637.
[2] M. Dahl, A. Brun, and G. B. Andresen, “Using ensemble weather predictions in district heating operation and load forecasting,” Appl. Energy, vol. 193, pp. 455–465, May 2017.

[3] J. Gu, J. Wang, C. Qi, C. Min, and B. Sundén, “Medium-term heat load prediction for an existing residential building based on a wireless on-off control system,” Energy, vol. 152, pp. 709–718, Jun. 2018.

[4] J. Yuan, C. Wang, and Z. Zhou, “Study on refined control and prediction model of district heating station based on support vector machine,” Energy, vol. 189, Dec. 2019, Art. no. 116193.

[5] Y. Ding, Q. Zhang, T. Yuan, and K. Yang, “Model input selection for building heating load prediction: A case study for an office building in Tianjin,” Energy Buildings, vol. 159, pp. 254–270, Jan. 2018.

[6] F. Dalipi, S. Yıldırım Yavılın, and A. Gebremedhin, “Data-driven machine-learning model in district heating system for heat load prediction: A comparison study,” Appl. Comput. Intell. Soft Comput., vol. 2016, pp. 1–11, Jun. 2016.

[7] M. Wang and Q. Tian, “Dynamic heat supply prediction using support vector regression optimized by particle swarm optimization algorithm,” Math. Problems Eng., vol. 2016, pp. 1–10, May 2016.

[8] S. Sajjadi, S. Shamsirband, M. Alizamir, P. L. Yee, Z. Mansor, A. A. Manaf, T. A. Altameem, and A. Mostafaeipour, “Extreme learning machine for prediction of heat load in district heating systems,” Energy Buildings, vol. 122, pp. 222–227, Jun. 2016.

[9] P. Xue, Y. Jiang, Z. Zhou, X. Chen, X. Fang, and J. Liu, “Multi-step ahead forecasting of heat load in district heating systems using machine learning algorithms,” Energy, vol. 188, Dec. 2019, Art. no. 116085.

[10] G. Suryanarayana, J. Lago, D. Gysen, P. Aleksiejuk, and C. Johansson, “Thermal load forecasting in district heating networks using deep learning and advanced feature selection methods,” Energy, vol. 157, pp. 141–149, Aug. 2018.

[11] G. Xue, Y. Pan, T. Lin, J. Song, C. Qi, and Z. Wang, “District heating load prediction algorithm based on feature fusion LSTM model,” Energies, vol. 12, no. 11, p. 2122, Jun. 2019.

[12] K. Kato, M. Sakawa, K. Ishimaru, S. Ushiro, and T. Shibano, “Heat load prediction through recurrent neural network in district heating and cooling systems,” in Proc. IEEE Int. Conf. Syst., Man Cybern., Oct. 2008, p. 1400.

[13] G. Lai, W.-C. Chang, Y. Yang, and H. Liu, “Modeling long- and short-term temporal patterns with deep neural networks,” in Proc. 41st Int’l ACM SIGIR Conf. Res. Develop. Inf. Retr., Jul. 2018, pp. 95–104.

[14] Y. Song, C. Ning, L. Fa, X. Liu, X. Shu, Z. Li, and J. Tang, “Multi-part boosting LSTMs for skeleton based human activity analysis,” in Proc. IEEE Int. Conf. Multimedia Expo Workshops (ICMEW), Jul. 2017, pp. 605–608.

[15] X. Shu, J. Tang, G.-J. Q., Y. Song, Z. Li, and L. Zhang, “Concurrence-aware long short-term sub-memories for person-person action recognition,” in Proc. IEEE Comput. Vis. Pattern Recognit. Workshops (CVPRW), Jul. 2017, pp. 1–8.

[16] B. Yang, S. Sun, J. Li, X. Lin, and Y. Tian, “Traffic flow prediction using LSTM with feature enhancement,” Neurocomputing, vol. 332, pp. 320–327, Mar. 2019.

[17] B. Liao, S. Tang, S. Yang, W. Zhu, and F. Wu, “Multi-modal sequence to sequence learning with content attention for hotspot traffic speed prediction,” in Proc. Pacific Rim Conf. Multimedia, 2018, pp. 212–222.

[18] Z. He, C.-Y. Chow, and J.-D. Zhang, “STANN: A spatio-temporal attentive neural network for traffic prediction,” Inter. J. Electr. Power Energy Syst., vol. 7, pp. 4795–4806, 2019.

[19] Y. Danfeng and W. Jing, “Subway passenger flow forecasting with multi-station and external factors,” IEEE Access, vol. 7, pp. 57415–57423, 2019.

[20] Y. G. Cinar, H. Mirisaee, P. Goswami, E. Gaussier, and A. Aït-Bachir, “Period-aware content attention RNNs for time series forecasting with missing values,” Neurocomputing, vol. 312, pp. 177–186, Oct. 2018.

[21] S. Wang, X. Wang, S. Wang, and D. Wang, “Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting,” Int. J. Electr. Power Energy Syst., vol. 109, pp. 470–479, Jul. 2019.

[22] P. He, Z. Tian, S. Yuan, and N. Zhu, “Modeling the dynamic characteristics of a district heating network,” Energy, vol. 39, no. 1, pp. 126–134, Mar. 2012.

[23] Y. Xia, H. Leung, and H. Chan, “A prediction fusion method for reconstructing spatial temporal dynamics using support vector machines,” IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 53, no. 1, pp. 62–66, Jan. 2006.

[24] M. Xu, Y. Yang, M. Han, T. Qiu, and H. Lin, “Spatio-temporal interpolated echo state network for meteorological series prediction,” IEEE Trans. Neural Netw. Learn. Syst., vol. 30, no. 6, pp. 1621–1634, Jun. 2019.

[25] Z. Zhu, J. Chen, L. Zhu, X. Duan, and Y. Liu, “Wind speed prediction with spatio- temporal correlation: A deep learning approach,” Energies, vol. 11, no. 4, p. 705, Mar. 2018.

[26] Q. Liu, Y. Cai, H. Jiang, X. Chen, and J. Lu, “Traffic state spatial-temporal characteristic analysis and short-term forecasting based on manifold similarity,” IEEE Access, vol. 6, pp. 9690–9702, 2018.

[27] F. Lin, Y. Xu, Y. Yang, and H. Ma, “A spatio-temporal hybrid model for short-term traffic prediction,” Math. Problems Eng., vol. 2019, pp. 1–12, Jan. 2019.

[28] W. Zhao-hong, “Numeric missing value’s hot deck imputation based on cloud model and association rules,” in Proc. 2nd Int. Workshop Educ. Technol. Comput. Sci., Mar. 2010, pp. 238–241.

[29] J. Hu and W. Zheng, “Multistage attention network for multivariate time series prediction,” Neurocomputing, vol. 383, pp. 122–137, Mar. 2020.

[30] Y. Yang, J. Dong, X. Sun, E. Lima, Q. Mu, and X. Wang, “A CFCC-LSTM model for sea surface temperature prediction,” IEEE Geosci. Remote Sens. Lett., vol. 15, no. 2, pp. 207–212, Feb. 2018.

[31] Q. Zhang, H. Wang, J. Dong, G. Zhong, and X. Sun, “Prediction of sea surface temperature using long short-term memory,” IEEE Geosci. Remote Sens. Lett., vol. 14, no. 10, pp. 1745–1749, Oct. 2017.

[32] S. Hochreiter and J. J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[33] X. Ran, Z. Shan, Y. Fang, and C. Lin, “An LSTM-based model with attention mechanism for travel time prediction,” Sensors, vol. 19, no. 4, p. 861, Feb. 2019.

[34] Z. Wen, R. S. Srinivasan, and J. Shi, “Artificial intelligent models for improved prediction of residential space heating,” J. Energy Eng., vol. 142, no. 4, Dec. 2016, Art. no. 04016006.

[35] M. Protic, S. Shamsirband, M. H. Anisi, D. Petkovic, D. Mitic, M. Raos, M. Arif, and K. A. Alam, “Appraisal of soft computing methods for short term consumers’ heat load prediction in district heating systems,” Energy, vol. 82, pp. 697–704, Mar. 2015.

[36] J. Wang, P. Li, R. Ran, Y. Che, and Y. Zhou, “A short-term photovoltaic power prediction model based on the gradient boost decision tree,” Appl. Sci., vol. 8, no. 5, p. 689, Apr. 2018.

[37] B. Thiagarajan, L. Srinivasan, A. V. Sharma, D. Sreekanthan, and V. Vijayaraghavan, “A machine learning approach for prediction of on-time performance of flights,” in Proc. IEEE/IAIA 36th Digit. Avionics Syst. Conf. (DASC), Sep. 2017, pp. 1–6.
JIANCAI SONG received the B.S. degree in computer science and technology from the Hebei University of Technology, Tianjin, China, in 2003, and the Ph.D. degree in communication engineering from Tianjin University, in 2016. He is currently a Senior Engineer with the Tianjin University of Commerce. His research interests include artificial intelligence, deep learning, and their applications in design and optimization on the smart city.

CHENGYING QI received the B.S. and master’s degrees in engineering thermo-physics and the Ph.D. degree in thermal engineering from Tianjin University, in 1985, 1988, and 1999, respectively. He is currently a Professor with the Hebei University of Technology. His research interests include the theory and technology of renewable energy utilization, enhanced heat transfer technology, smart heating systems, and heat metering systems.