Journal of Intelligent & Fuzzy Systems 43 (2022) 3479–3491
DOI:10.3233/JIFS-213176
IOS Press

RETRACTED: Multiscale convolutional recurrent neural network for residential building electricity consumption prediction

Hongxia Wang[a], Wubin Ma[b,*], Zhiru Wang[a,*] and Chenyang Lu[b]
[a]*School of Economics and Management, University of Science and Technology Beijing, Beijing, China*
[b]*Information System Engineering Laboratory, National University of Defense Technology, Changsha, China*

This article has been retracted. A retraction notice can be found at https://doi.org/10.3233/JIFS-219433.

1. Introduction

Reducing energy consumption and related carbon emissions has become one of the most important issues in the world. A building's end-use energy consumption accounts for a large proportion of the total energy consumption. For instance, the residential and tertiary sectors consumed 40% of the European Union's total energy [1, 2]. In 2020, residential and commercial buildings accounted for approximately 22% and 18%, respectively, of the total U.S. end-use energy according to statistics from the U.S. Energy Information Administration (EIA). Therefore, the reduction in end-use energy consumption by buildings is crucial to meet the goal of energy conservation [3]. Accurately predicting the energy use in buildings is important for energy planning and energy savings.

The prediction of building or residential energy consumption has attracted much attention. For example, previous research showed that unnecessarily leaving computers on or on standby contribute to 20–30% of energy consumption in the UK. In China, especially in public service buildings and university research rooms, the inappropriate use of electrical appliances leads to a large amount of energy waste [4]. The prediction of energy consumption can provide an early warning for the abnormal use of energy

*Corresponding author. Wubin Ma, Information System Engineering Laboratory, National University of Defense Technology, Deya Road, 410073 Changsha, China. E-mail: wb_ma@nudt.edu.cn. (W.M.); Zhiru Wang, School of Economics and Management, University of Science and Technology Beijing, No. 30 Xueyuan Road, 100083 Beijing, China. E-mail: wangzhiru@ustb.edu.cn. (Z.W.); Tel.: (++8613548696947)

and provide decision support for energy supply strategies and the energy supply scheduling department [5].

Among the sources of energy consumption in residential buildings, electricity is the most consumed energy source. According to the 2015 Residential Energy Consumption Survey (RECS) by the U.S. EIA, electricity consumption accounted for 47% of the total energy consumption in all U.S. households. Therefore, this study focuses on electricity consumption in residential buildings.

Currently, the prediction accuracy of electricity consumption is insufficient. Traditional machine learning (ML) methods can predict electricity consumption. However, there are many factors affecting electricity consumption, and the relationships between these factors are very complicated. Therefore, traditional machine learning methods have difficulty in obtaining the long-term dependency and the time-series information of the various factors. Recently, some researchers have used deep learning methods to predict energy consumption, including the recurrent neural network (RNN), the long short-term memory network model (LSTM), the gated recurrent unit network model (GRU), and bidirectional long short-term memory (Bi-LSTM). To some extent, these methods can extract some information that traditional ML methods may miss. However, neither method can concurrently obtain the different scales of correlation information and long-time dependency.

A hybrid model of a multiscale convolutional recurrent neural network (MCRNN) is proposed in this paper. The parameters of MCRNN include historical indoor temperature and humidity and outdoor atmospheric pressure, temperature, humidity, wind, and visible light data. The main characteristics of MCRNN are as follows:

First, in the MCRNN model, a bidirectional recurrent neural network (BiRNN) structure is used to identify data collected by indoor and outdoor sensors and collect long-term dependence information.

Second, a multiscale convolutional recurrent neural network unit is proposed to collect information with different scales. Two units are used to obtain the impact of temperature, humidity, and other weather information on electricity consumption for different time periods.

Third, an integrated model named MCRNN that uses multiscale recurrent neural network units and BiRNNs is proposed to ensure that the long-term dependence on information and multiscale influence information can be collected.

We summarize our contributions as follows:

- We propose a new neural network model, the multiscale convolutional recurrent neural network (MCRNN), which can collect both the long-term dependence on information and multiscale influence information.
- We apply MCRNN to the prediction of residential building electricity consumption. An experiment using data from a residential building in Belgium proves the prediction accuracy of this model.
- The MCRNN model is compared with eight frequently used ML models, including SVM (support vector machine), RF (random forest), LSTM, GRU, Bi-LSTM, Bi-GRU (bidirectional gated recurrent unit network model), Bi-Conv-LSTM (the combination of a convolutional neural network and bidirectional long short-term memory), and Bi-Conv-GRU (the combination of a convolutional neural network and bidirectional gated recurrent unit network model). The advantages of MCRNN are verified from multiple aspects, such as validation loss, training loss, prediction accuracy and efficiency.

The rest of this paper is arranged as follows. Section 2 introduces related work, and the progression of the method of MCRNN is described in Section 3. Section 4 describes the experiments. Finally, the research content of this article is summarized and prospective work is proposed in Section 5.

2. Related work

Researchers have used different methods for building energy consumption predictions. These studies can be divided into two types: nonneural network-based methods and neural network-based methods.

2.1. Nonneural network-based methods

Nonneural network-based methods include linear regression (LR), ensemble learning (EL), and support vector machine/regression (SVM/SVR).

A linear regression can describe the relationship between multiple factors. Regression models have been widely used in predicting energy consumption in office buildings [6, 7] and higher education buildings [8]. A multivariate linear regression was used in the prediction of a rental house's energy consumption [9] and the estimation of the energy consumption by utilizing data such as schedules, operating behaviours, and sensor devices of rental housing employees. A prediction method combining multivariate linear regression with a back-propagation neural network (BPNN) was proposed [10]. This method focused on selecting and optimizing training samples with linear regression. However, the sample selection method based on LR is obviously more suitable for prediction with BPNN. LR was also used to forecast building energy consumption [11, 12]. However, as it is difficult to describe complex systems, there are certain limitations in linear regression models. The change in building energy consumption is influenced by many factors, including light, temperature and humidity. There are many linear and nonlinear relationships between these factors, so it is difficult for the regression model to make effective predictions.

Ensemble learning uses multiple learning models to randomly collect building energy information and optimally selects from multiple model combinations to obtain the final model [13]. Typical ensemble learning models include random forest (RF) and gradient boosting methods. An RF method was applied for the prediction of energy consumption of mobile educational institutions in North Central Florida [14]. The different characteristic contribution degrees affecting the energy consumption of the building are analysed as well. A combination model that includes RF and nonlinear autoregressive was proposed to predict energy consumption [15]. In addition, the calculation accuracy of the regression models was compared with that of the RF and nonlinear autoregressive models. A gradient boosting machine (GBM) method was used to forecast energy consumption in commercial buildings [16]. The experiment proved that the accuracy (i.e., RMSE) of a gradient-based calculation method exceeds that of the LR and RF methods by 80%. EL has certain advantages. However, as a traditional machine learning method, EL cannot sufficiently obtain the sequence and contextual feature information of building energy, resulting in lower prediction accuracy.

SVM/SVR is a generalized linear classifier that tries to find a hyperplane to segment samples into different categories. The principle of segmentation is to find the maximum interval between different categories and finally transform the original problem. Support vector regression (SVR) was proposed to forecast the energy consumption value of public buildings [17]. The traditional SVR was improved, and the MSE loss function was substituted with the information theory cost function to solve the insensitivity problem of SVR in building energy prediction [18]. Therefore, a vector field-based SVR model was applied, especially for extremely high-dimensional samples [19]. The high-dimensional multiple-distortion samples were mapped to a vector field. SVM and SVR have good performance on classification problems, but for energy consumption, their prediction accuracy is still very low.

2.2. *Neural network-based methods*

With the widespread use of neural networks in the field of artificial intelligence, an increasing number of researchers have begun to use neural network models in the energy consumption prediction of a building [8, 20–26]. A back-propagation artificial neural network (ANN) was used to forecast the electricity consumption of residential buildings [27]. The ANN-based method can solve nonlinear problems effectively and quickly while minimizing training errors. Based on the back-propagation ANN, an ANN model based on a stack-type noise reduction autoencoder was proposed [28]. To improve the prediction performance of building energy consumption, a hybrid model based on an improved deep belief network (DBN) was used [29]. The contrastive divergence (CD) algorithm was used to improve the model's hidden parameters, while the least squares method was used to improve the output weighting vectors.

Due to the time-series characteristics of building energy consumption data, the most suitable neural network model is the recurrent neural network (RNN) and its variants [30]. RNN has shown very good performance in modelling 24-hour ahead prediction [30]. Among these methods, the most representative recurrent neural networks are the gated recurrent unit (GRU) network model and the long short-term memory (LSTM) network model. For instance, convolutional neural networks (CNNs) and GRUs (GRUs) were combined (Conv-GRU) to forecast short-term residential loads [31]. LSTM has different kinds of gates that can decide which information should be retained or discarded, so it is well suited for learning from historical series to extract long-distance sequence dependency. The LSTM network was also combined with the stationary wavelet transform (SWT) to predict building energy consumption [32]. SWT can capture the characteristics of stationary sequence information and reduce fluctuations. To improve computational efficiency, the k-means clustering algorithm and

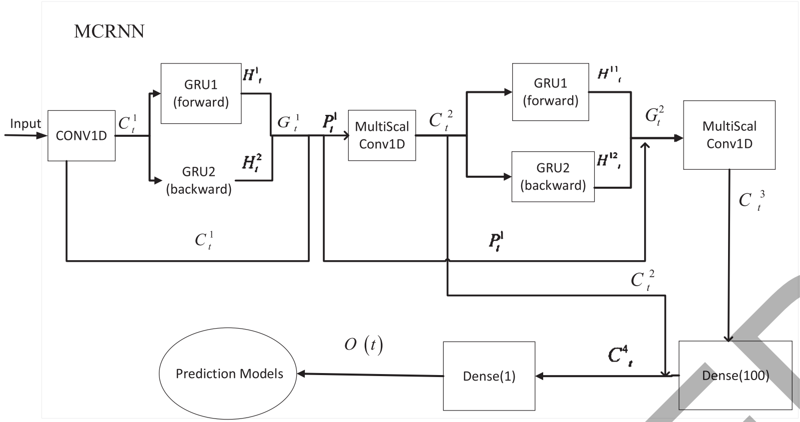

Fig. 1. The structure of the MCRNN model.

transfer learning were combined with LSTM models [33]. Good results were achieved in predicting the energy consumption of smart buildings in South Korea. In addition, the method combining convolutional networks with LSTM (Conv-LSTM) could collect sequence information through convolutional networks and obtain long-term dependencies with LSTM [34]. The combination of a convolutional neural network (CNN) and bidirectional long short-term memory (Bi-Conv-LSTM) was applied to predict electric energy consumption and achieved good results [35]. However, currently, Conv-LSTM simply connects these two types of networks. It does not extract the sequence information at different scales, and the improvement effect is not obvious. Moreover, Conv-LSTM does not capture the bidirectional sequence characteristics, and the connection method needs to be further optimized.

3. Materials and methods

A novel method (MCRNN) that can obtain bidirectional long-term sequence information at different scales is proposed. The structure of the MCRNN is shown in Fig. 1.

Definition 1. Energy consumption prediction model. It is assumed that there is a set of time-series data $(x_0, ..., x_N)$ that collects information about temperature, wind speed and direction, and pressure. A set of observational time-series data for building energy consumption $(y_0, ..., y_U)\,, U < N$ and a set of data series $(y_{\mathrm{U}+1}, ..., y_N)$ must be predicted. The estimated value of the data is expressed as $(\tilde{y}_{U+1}, ..., \tilde{y}_N)$. Equation (1) gives the building energy consumption time-series prediction model.

$$(\tilde{y}_{U+1}, ..., \tilde{y}_N) = f\,(x_0, ..., x_U, y_0, ..., y_U) \qquad (1)$$

It is required that the time-series estimated value $\tilde{y}_{U+1}$ depend only on the previous U time-series values. The goal of this model is to obtain the best $f\,(X)$, making the predicted value $(\tilde{y}_{U+1}, ..., \tilde{y}_N)$ closest to the true value $(y_{\mathrm{U}+1}, ..., y_N)$.

A one-dimensional convolutional network unit is used to recognize the input data in MCRNN. Two one-dimensional convolutional networks of different scales connect two bidirectional GRU convolutions, which can simultaneously identify sequence features of different scales and long-term dependency.

The core of the MCRNN structure is two multiscale convolutional (MC) operations and two bidirectional gated recurrent units (BiGRUs). This structure forms a tandem connection. Assuming there is a time series $X = \{x_1, x_2, ...x_t\}$, the process is as follows:

The first layer of convolution $\eta^1\,(x_t)$ accepts the input of sequence data, $\eta^1\,(u) = \sum_{i=0}^{k-1} \beta\,(i)\, X_{u-di}$, where k indicates the filter size, d represents the convolution dilation factors in this convolutional layer, and $\beta\,(i)$ is a convolution kernel function. Convolution scales are adjusted by these two parameters. The output C_t^1 is shown in Equation (2).

$$C_t^1 = \eta^1(u) = \sum_{i=0}^{k-1} \beta\,(i)\, X_{u-di} \qquad (2)$$

where C_t^1 is connected to the update gate of the first and the second BI-GRU as input, $\sigma\,(x) = 1/\left(1 + e^{-x}\right)$. The output of the forget gate is f_t^1 and

f_t^2, as shown in Equations (3) and (4), respectively.

$$f_t^1 = \sigma\left(W_f^1[H_{t-1}^1, C_t^1] + B_f^1\right) \quad (3)$$

$$f_t^2 = \sigma\left(W_f^2[H_{t-1}^2, C_t^1] + B_f^2\right) \quad (4)$$

The output of the update gate is z_t^1 and z_t^2:

$$z_t^1 = \sigma\left(W_z^1[H_{t-1}^1, C_t^1] + B_z^1\right) \quad (5)$$

$$z_t^2 = \sigma\left(W_z^2[H_{t-1}^2, C_t^1] + B_z^2\right) \quad (6)$$

$$\tilde{h^1}_t = \tanh\left(W_h^1[f_t^1 \cdot H_{t-1}^1, C_t^1] + B_h^1\right) \quad (7)$$

$$\tilde{h^2}_t = \tanh\left(W_h^2[f_t^2 \cdot H_{t-1}^2, C_t^1] + B_h^2\right) \quad (8)$$

The output of the first-layer bidirectional GRU is H_t^1 and H_t^2, as shown in Equations (9) and (10).

$$H_t^1 = \left(1 - z_t^1\right) \cdot h_{t-1}^1 + \tilde{h^1}_t \cdot z_t^1 \quad (9)$$

$$H_t^2 = \left(1 - z_t^2\right) \cdot h_{t-1}^2 + \tilde{h^2}_t \cdot z_t^2 \quad (10)$$

The output of the first Bi-GRU layer is $\left[H_t^1, H_t^2\right]$, which is the concatenation of the forward GRU output H_t^1 and backwards GRU output H_t^2. As shown in Equation (11), G_t^1, which is the output of the first fusional layer, is the result of multiplying the Bi-GRU output by the weight vector W_{g1}^1 and adding the offset vector B_{g1}^1:

$$G_t^1 = W_{g1}^1 \cdot \left[H_t^1, H_t^2\right] + B_{g1}^1 \quad (11)$$

Concatenating G_t^1 with C_t^1, which is the output of $\eta^1(x_t)$, we can obtain P_t^1:

$$P_t^1 = \left[G_t^1, C_t^1\right] \quad (12)$$

where P_t^1 is the input of the second convolutional layer. Then, C_t^2 is the output of the second convolutional layer with the scale of *Scale*1:

$$C_t^2 = MutiScalConv\left(P_t^1, Scale1\right) \quad (13)$$

where C_t^2 is connected to the second Bi-GRU layer and used as the input. This step is repeated on the first Bi-GRU layer in Equations (3)–(11) to finally obtain C_t^3:

$$C_t^3 = MutiScalConv\left(P_t^1, Scale2\right) \quad (14)$$

A convolution operation with the scale of *Scale*2 is used on $[C_t^2, C_t^3]$. The sequence information that is more important to the target can be retained in this way. The output C_t^4 is obtained through a fully connected operation O_t, as shown in Equations (15) and (16).

$$C_t^4 = \eta^2\left([C_t^2, C_t^3]\right) \quad (15)$$

$$O_t = W_O^1 \cdot \left[C_t^4\right] + B_O^1 \quad (16)$$

The above calculations show the process of the MCRNN model. The experimental result of this algorithm is evaluated in Section 4.

4. Experiments and discussions

4.1. Dataset and experiment background

Dataset: As the prediction model needs to be validated using high-frequency energy consumption data, the residential building studied should be equipped with devices to record energy consumption data every hour or every 10 minutes. We use the electricity consumption dataset of a residential building in Belgium. The amount of electricity consumption in this building is recorded every 10 minutes. The items included in the dataset are shown in Table 1. There are eight areas in the building. The distribution of the building energy consumption dataset is shown in Fig. 2.

Experimental hyperparameters: Table 2 shows the hyperparameters of the MRCNN model.

Experimental processing unit: The computer is configured with a Pentium(R) Dual-core 3.06 CPU and 8 G RAM memory.

Evaluation functions: The functions used in the performance evaluation of different models are the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and coefficient of determination (R^2):

The RMSE is calculated using Equation (17).

$$RMSE = \sqrt{\sum_{i=1}^{N} (\tilde{y}_i - y_i)^2 \Big/ N} \quad (17)$$

Table 1
Items in the dataset

Item name	Unit	Meaning
T1	Celsius	Kitchen area's temperature
RoomH_1	%	Kitchen area's humidity
T2	Celsius	Living room's temperature
RoomH_2	%	Living room's humidity
T3	Celsius	Laundry room's temperature
RoomH_3	%	Laundry room's humidity
T4	Celsius	Study room's temperature
RoomH_4	%	Study room's humidity
T5	Celsius	Bathroom's temperature
RoomH_5	%	Bathroom's humidity
T6	Celsius	Temperature outside the building
RoomH_6	%	Humidity outside the building
T7	Celsius	Ironing room's temperature
RoomH_7	%	Ironing room's humidity
T8	Celsius	Teenager room's temperature
RoomH_8	%	Teenager room's humidity
T9	Celsius	Parents' room's temperature
RoomH_9	%	Parents' room's humidity
T_out	Celsius	Temperature outside
Pressure	MmHg	Pressure outside
RH_out	%	Humidity outside
Wind speed	m/s	Wind speed outside
Visibility	km	Visibility outside
Tdewpoint	Â°C	Dew point outside

The MAE is calculated as shown in Equation (18).

$$MAE = \sum_{i=1}^{N} |\tilde{y}_i - y_i| \Big/ N \tag{18}$$

The MAPE is calculated using Equation (19).

$$MAPE = 100\% \cdot \sum_{i=1}^{N} \left| \frac{\tilde{y}_i - y_i}{y_i} \right| \Big/ N \tag{19}$$

The R^2 is calculated as shown in Equation (20).

$$R^2 = 1 - \frac{\mathrm{E}\left[\tilde{y} - \mathrm{E}\tilde{y}\right]^2}{\mathrm{E}\left[Y - \mathrm{E}Y\right]^2} \tag{20}$$

It should be noted that RMSE, MAE, and MAPE are all measures of prediction error, and R^2 represents the relationship between two sequences of data. The larger the R^2 value is, the greater the correlation between two sequences of data, and the better the prediction result.

4.2. Dataset analysis

The data sample distribution is shown in Fig. 2. Figure 2(a) shows the relationship between temperature and electricity consumption in different areas, T1-T9 are the temperature data for different rooms, among which T4 (study room) and T5 (bathroom) have the highest average temperature, and T6 (outside), T2 (living room) and T9 (parents' room) have the lowest average temperature. These results show that the temperature change in the house has the same trend (T1-T5 and T7-T9), and the temperature change outside the house also has the same trend (T6, T_out, Tdewpoint). From Fig. 2(a), it can be seen that there is no strong correlation between the temperature change and electricity consumption (Appli). Figure 2(b) shows the temperature difference between indoors and outdoors. The temperature inside the house is always higher than outside the house, and when the difference in temperature is too large, the power consumption will increase significantly.

Figure 2(c) shows the relationship between humidity and power consumption in different areas of the house, where RH_1-RH_9 is the change in air humidity in different rooms. The difference in air humidity between the nine rooms in the house is not obvious. The average air humidity between RH_6 (outside) and RH_8 (teenager room 2) is slightly larger than in other rooms in the house. Similar to the temperature, the humidity change inside the house has the same trend except for RH_5 (bathroom). Figure 2(d) shows the humidity difference between indoors and outdoors. There is no obvious correlation between the humidity differences and electricity consumption.

Figure 2(e) shows the curves of pressure, wind speed, visibility, and electricity consumption. There is no periodic characteristic or correlation between these variables.

The dataset we used includes four months of statistical data. To display the changes in various parameters more dynamically and explore the monthly periodicity of parameters, we use the monthly average of each parameter. Figure 3 shows the monthly changes in each parameter. The results show that the periodicity of each sequence of data changes is not very strong, which means that electricity consumption does not show periodic changes. From a monthly point of view, during the 4–5 months of data collection, as the weather gradually became hotter (the data from T1-T9 show an upwards trend), the air humidity gradually decreased (the data from RH_1-RH_9 show a downwards trend), and the overall electricity consumption tended to decrease.

The box-plot diagram of each item is shown in Fig. 4. The "appliance" item in the left-most of Fig. 4 is the electricity consumption data, which is also the

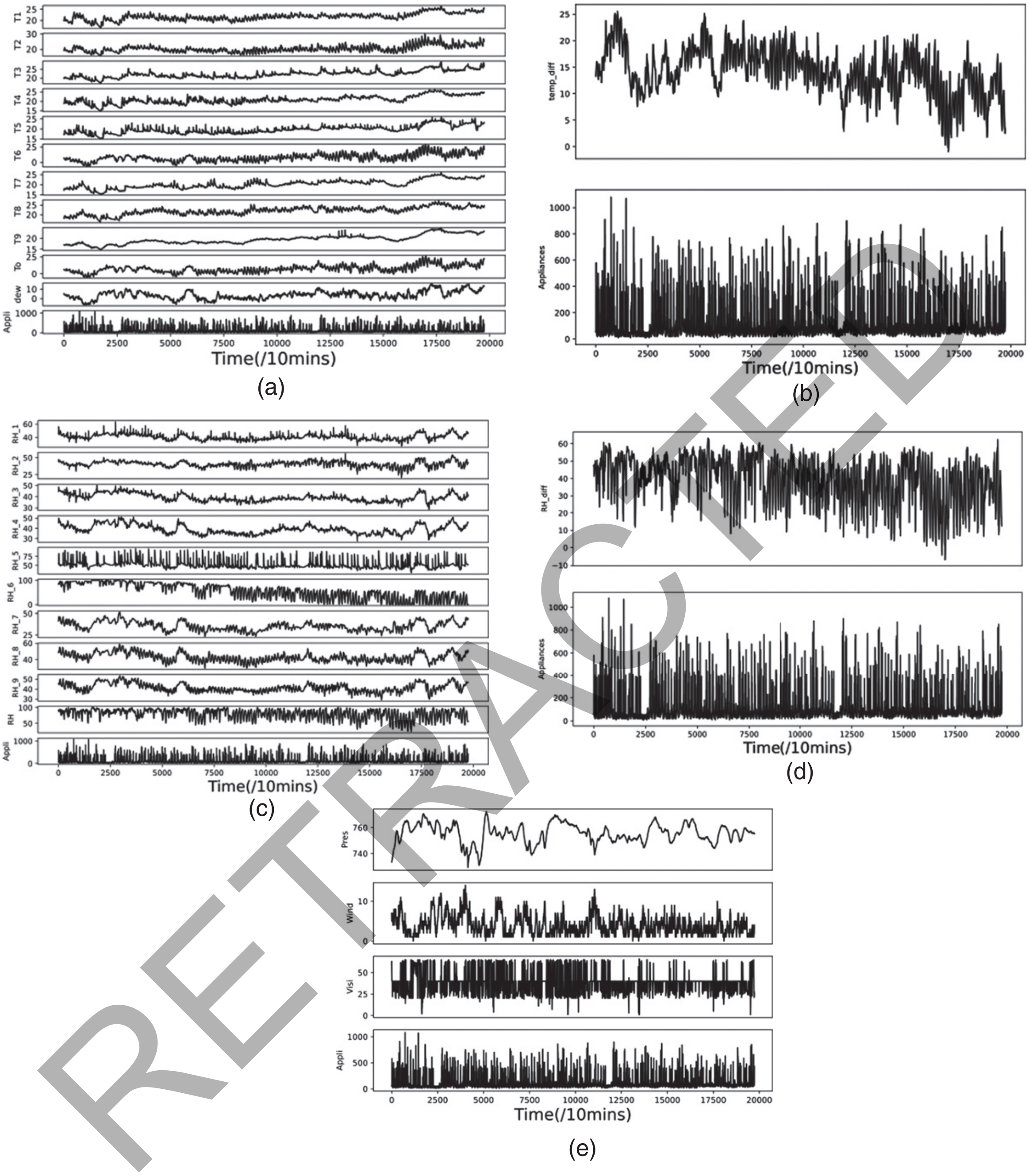

Fig. 2. Distribution of building electricity consumption dataset.

label that needs to be trained and predicted. The maximum value of label data is 1080, the minimum is 10, the median number is 97, and 75% of the data lies between 71 and 119. The data has been normalized for model training.

As shown in Fig. 5, the electricity consumption of the entire house does not have a strong correlation with any certain factor. All correlation coefficients are less than 0.2. This shows that the building's electricity consumption is the result of the joint action of

Table 2
The values of parameters

Name	Value
LEARNING_RATE	0.0004
WINDOW_SIZE	20
BATCH_SIZE	100
TRAIN_RATE	0.8
VALIDATE_RATE	0.1
TEST_RATE	0.1
SCALE1	10
SCALE2	20

multiple rooms. The future electricity consumption in the building needs to be analysed from the overall correlation among many factors.

4.3. Performance comparations and discussions

4.3.1. Prediction accuracy

Nine machine learning models are used to predict electricity consumption in the building. The prediction accuracy is shown in Tables 3 and 4. The best of each kind of model is in bold.

As shown in Tables 3 and 4, the prediction accuracy of MCRNN is generally higher than that of the other models. Compared to the SVM, RF, LSTM, GRU, Bi-LSTM, Bi-GRU, Bi-Conv-LSTM, Bi-Conv-GRU models, the RMSE is reduced by 47.83%, 38.72%, 16.62%, 15.67%, 13.29%, 13.58%, 7.55%, and 3.09%, respectively, and the MAE

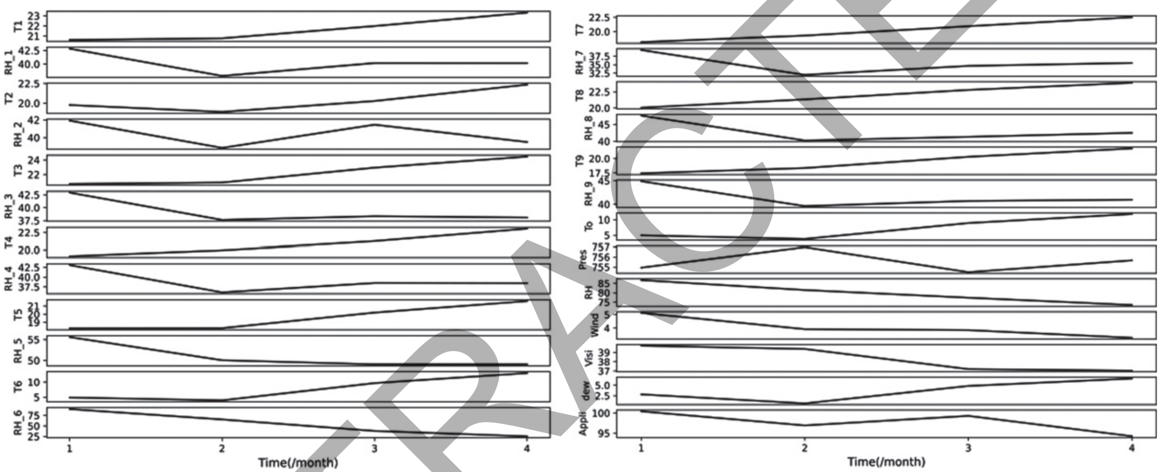

Fig. 3. Monthly average of parameters.

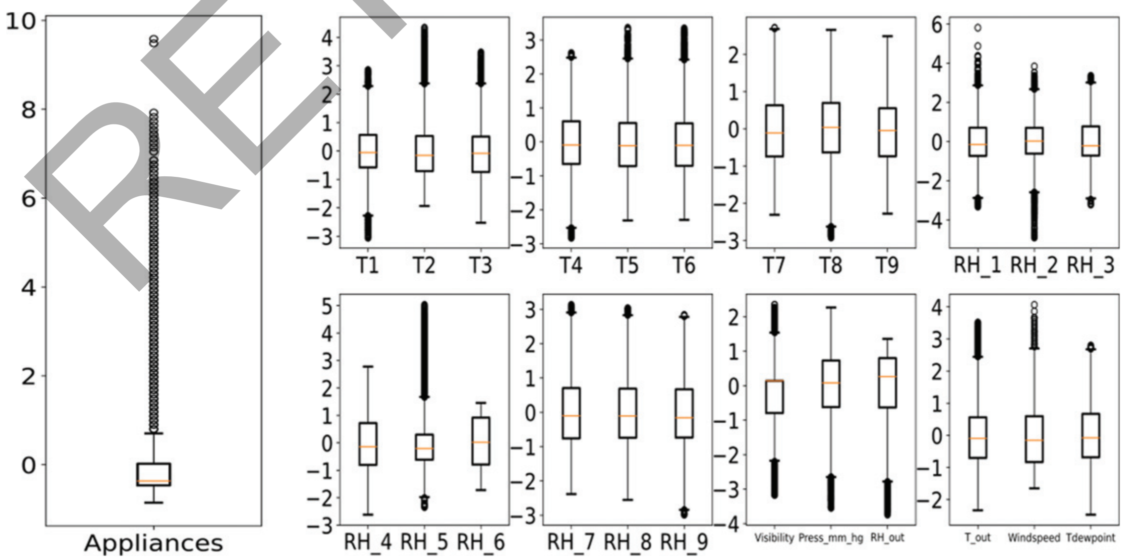

Fig. 4. Box-plot distribution of the building electricity consumption dataset.

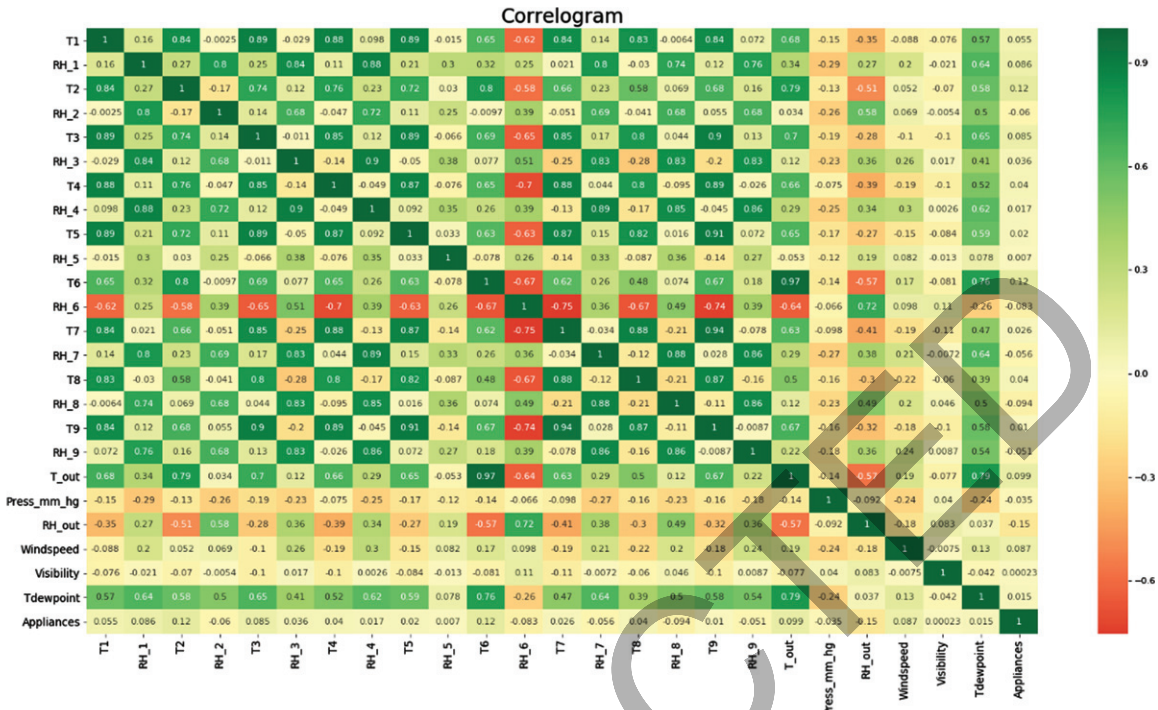

Fig. 5. Correlation diagram of the building energy consumption dataset.

Table 3
Predicted values of nonneural network-based algorithms

Model Name	RMSE	MAE	MAPE	R^2
SVM	56.6225	**24.8826**	**0.2429**	0.1315
RF	**53.1312**	30.7618	0.3951	**0.2381**

Table 4
Predicted values of neural network-based algorithms

Model Name	RMSE	MAE	MAPE	R^2
LSTM	44.6686	23.9224	0.2970	0.7144
GRU	44.3031	20.9606	0.2515	0.6883
Bi-LSTM	43.3905	22.6755	0.2853	0.7091
Bi-GRU	43.5030	23.0072	0.3005	0.7096
Bi-Conv-LSTM	41.1922	18.9461	0.2157	0.7369
Bi-Conv-GRU	39.4847	18.8488	0.2180	0.7613
MCRNN	**38.3016**	**18.0553**	**0.2042**	**0.7775**

is reduced by 37.81%, 70.38%, 32.50%, 16.09%, 25.59%, 27.43%, 4.93%, and 4.39%, respectively, in the MCRNN model. In addition, the prediction correlation is increased by 83.09%, 69.38%, 8.12%, 11.47%, 8.80%, 8.73%, 5.22%, and 2.08%, respectively, in the MCRNN model. Finally, MCRNN has better performance than other models on MAPE.

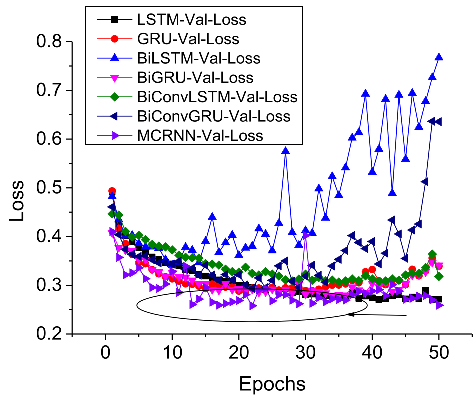

Fig. 6. Validation loss of different models with the number of model iterations.

4.3.2. Convergence and validation loss

From the perspective of the solving process, the convergence of MCRNN is stronger than other models in terms of the validation loss of the model once it is trained. Since the validation dataset is not included in the back-propagation calculation of the model, validation loss is often one of the most important criteria for evaluating the convergence of a model. As shown in Fig. 6, the validation loss of MCRNN is the lowest

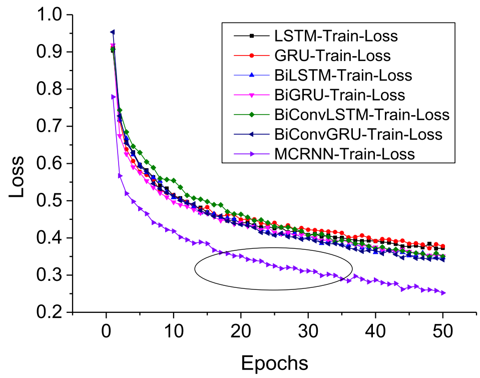

Fig. 7. Training loss of different models with the number of model iterations.

among all models, reaching the minimum value after the 13th epoch and remains the model with the lowest validation loss until the 37th epoch.

As shown in Fig. 6, the validation loss varies greatly in different models. The smoothness of the validation loss with the Bi-LSTM model is the worst as it fluctuates greatly after 15 iterations. It is obvious that overfitting occurs in Bi-LSTM. The Bi-Conv-GRU model fluctuates greatly in later iterations. This shows that the simple connection of the convolution and recurrent neural networks cannot improve the prediction accuracy and may sometimes introduce negative effects. The validation loss of the MCRNN model proposed in this paper has been maintained at a low level, which shows that the convolution

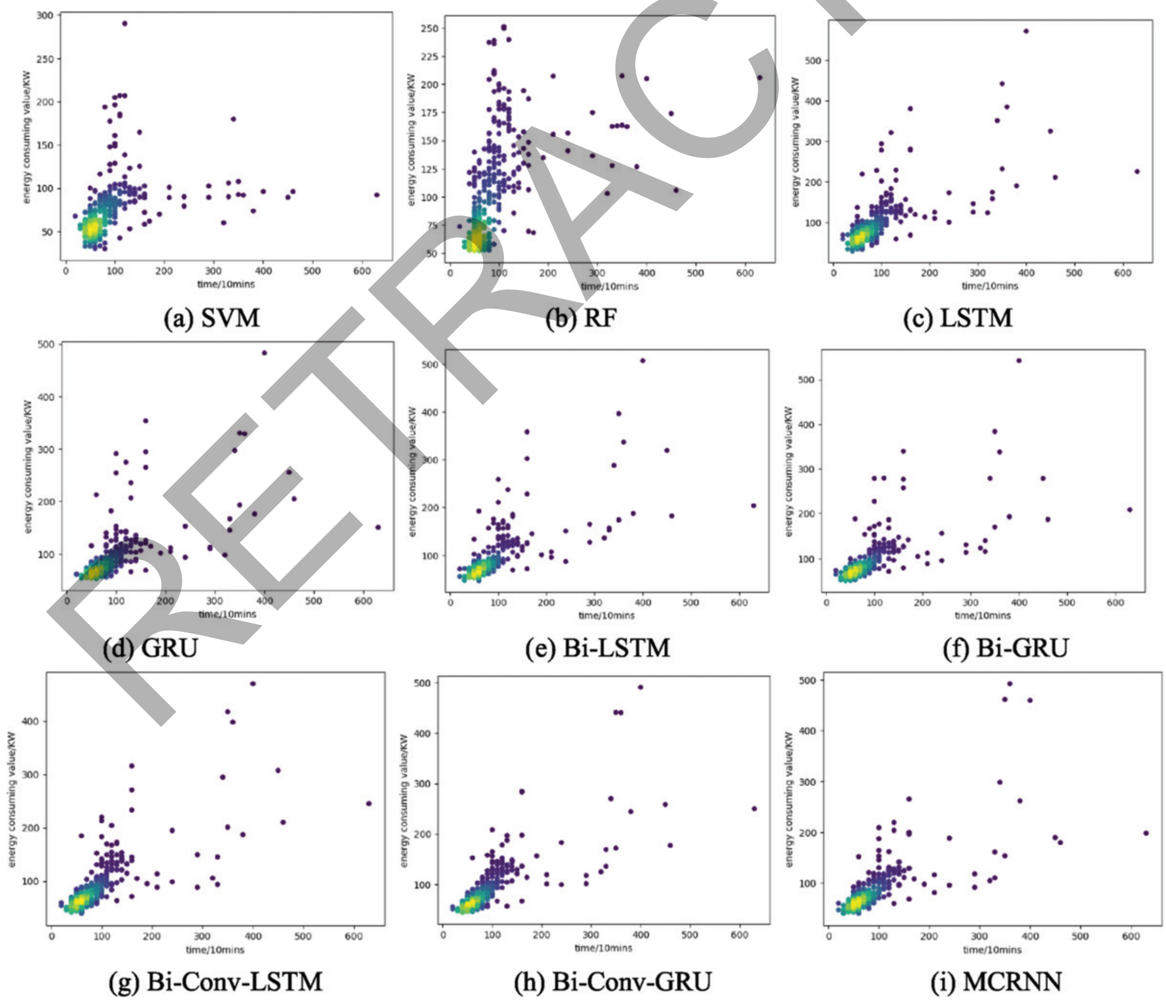

Fig. 8. Prediction accuracy of the different models studied.

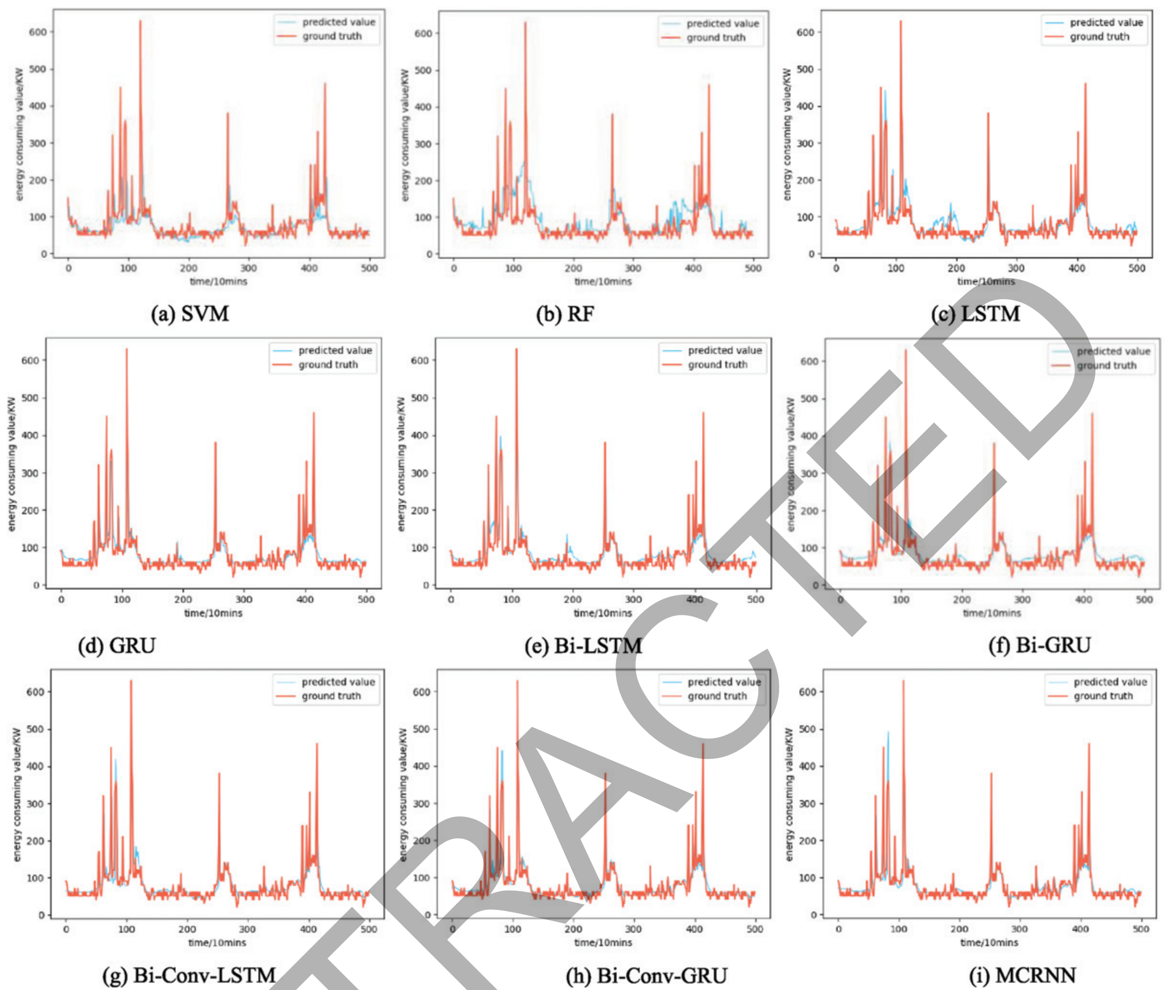

Fig. 9. Prediction results of the different models studied.

and recurrent neural network connection methods using multiple different scales can achieve good results.

4.3.3. *Training accuracy*

In terms of training accuracy, the performance of MCRNN, as shown in Fig. 7, indicates that this model is significantly better than other models. The minimum training loss reaches 0.25 in MCRNN, while other models are generally higher than 0.3. This also shows that the MCRNN model can better capture the most important information of the factors that affect building energy consumption, and can achieve a good fitting effect. Therefore, regardless of the perspective of training or verification loss, the MCRNN model has higher accuracy and stronger convergence than the other models.

4.3.4. *Predictive effect*

The 500 results in the test set are chosen for predictive effect evaluation in different models. We use the real value as the abscissa and the predicted value as the ordinate. The prediction effect results are shown in Fig. 8. The closer the points are to a straight line with a line of 45 degrees, the better the predictive effect is. It can be seen from Figs. 8(a) and (b) that SVM is slightly better than RF in prediction effect. The prediction models based on RNN are better than traditional ML methods (SVM and RF). The MCRNN model is better than the other models. When we compare the real value with the predicted value by different models, similar conclusions can be obtained. As shown in Fig. 9, the MCRNN and RNN models can better capture the sequence features, and the prediction effect is better than that of the SVM and RF models.

5. Conclusions

The accurate prediction of building electricity consumption can help decision-making departments plan the construction of energy facilities and provide an early warning of abnormal energy consumption for energy supply departments. A building electricity consumption prediction model named MRCNN is proposed in this paper. Multiple heterogeneous convolutional neural units are used to collect and obtain historical data at different scales. At the same time, the long-term dependence is obtained through a bidirectional recurrent neural network. Through experimental comparison with real data, the accuracy is further improved.

The building electricity consumption prediction problem can be further investigated in the future. First, the electricity consumption correlation in multiple areas needs to be considered. The electricity consumption in different functional areas may be different. It is necessary to use mathematical models to describe the electricity consumption relationship between different areas more accurately. Second, it is necessary to make more effective predictions for long-term energy consumption. Finally, as the neural network model has poor interpretability, more interpretive models need to be researched to predict energy consumption while maintaining the prediction accuracy.

Acknowledgments

This research was funded by the Hunan Natural Science Foundation (2018JJ3619), the National Natural Science Foundation of China (61871388, 71903013, 71704010, and 72101028), the Humanities and Social Science Fund of the Ministry of Education of China (18YJC630193), the Beijing Natural Science Foundation (L181009 and 9194028), and the Fundamental Research Funds for the Central Universities (FRF-BR-19-004B, FRF-BR-20-03A, 06106263). The authors would also like to thank the anonymous reviewers, whose comments and suggestions helped us to improve the paper. We would like to acknowledge the experts who provided suggestions.

References

[1] D. Chwieduk, Towards sustainable-energy buildings [J], *Applied Energy* **76**(1-3) (2003), 211–217.

[2] V. Bianco, A. Marchitto, F. Scarpa, et al., Forecasting energy consumption in the EU residential sector [J], *International Journal of Environmental Research and Public Health* **17**(7) (2020), 2259.

[3] S. Maki, S. Ashina, M. Fujii, et al., Employing electricity-consumption monitoring systems and integrative time-series analysis models: A case study in Bogor, Indonesia [J], *Frontiers in Energy* **12**(3) (2018), 426–439.

[4] D. Lee and C. Cheng, Energy savings by energy management systems: A review [J], *Renewable and Sustainable Energy Reviews* **56** (2016), 760–777.

[5] K. Amasyali and N.M. El-Gohary, A review of data-driven building energy consumption prediction studies [J], *Renewable and Sustainable Energy Reviews* **81** (2018), 1192–1205.

[6] S. Zhou and N. Zhu, Multiple regression models for energy consumption of office buildings in different climates in China [J], *Frontiers in Energy* **7**(1) (2013), 103–110.

[7] I. Korolija, Y. Zhang, L. Marjanovic-Halburd, et al., Regression models for predicting UK office building energy consumption from heating and cooling demands [J], *Energy and Buildings* **59** (2013), 214–227.

[8] D. Hawkins, S.M. Hong, R. Raslan, et al., Determinants of energy use in UK higher education buildings using statistical and artificial neural network methods [J], *International Journal of Sustainable Built Environment* **1**(1) (2012), 50–63.

[9] Y.R. Yoon and H.J. Moon, Energy consumption model with energy use factors of tenants in commercial buildings using Gaussian process regression [J], *Energy and Buildings* **168** (2018), 215–224.

[10] T. Yuan, N. Zhu, Y. Shi, et al., Sample data selection method for improving the prediction accuracy of the heating energy consumption [J], *Energy and Buildings* **158** (2018), 234–243.

[11] T. Zhang, L. Liao, H. Lai, et al., Electrical energy prediction with regression-oriented models [M], *Proceedings of the Fifth Euro-China Conference on Intelligent Data Analysis and Applications*. Cham: Springer International Publishing, 2018, 146–154.

[12] R. Wang, S. Lu and Q. Li, Multi-criteria comprehensive study on predictive algorithm of hourly heating energy consumption for residential buildings [J], *Sustainable Cities and Society* **49** (2019), 101623.

[13] Z. Dong, J. Liu, B. Liu, et al., Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification [J], *Energy and Buildings* **241** (2021), 110929.

[14] Z. Wang, Y. Wang, R. Zeng, et al., Random Forest based hourly building energy prediction [J], *Energy and Buildings* **171** (2018), 11–25.

[15] T. Ahmad and H. Chen, Nonlinear autoregressive and random forest approaches to forecasting electricity load for utility energy management systems [J], *Sustainable Cities and Society* **45** (2019), 460–473.

[16] S. Touzani, J. Granderson and S. Fernandes, Gradient boosting machine for modeling the energy consumption of commercial buildings [J], *Energy and Buildings* **158** (2018), 1533–1543.

[17] Y. Liu, H. Chen, L. Zhang, et al., Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: A case study in China [J], *Journal of Cleaner Production* **272** (2020), 122542.

[18] J. Duan, X. Tian, W. Ma, et al., Electricity consumption forecasting using support vector regression with the mixture

maximum correntropy criterion [J], *Entropy* **21** (2019), 707.
[19] H. Zhong, J. Wang, H. Jia, et al., Vector field-based support vector regression for building energy consumption prediction [J], *Applied Energy* **242** (2019), 403–414.
[20] J. Run Ge and R. Zmeureanu, Forecasting energy use in buildings using artificial neural networks: A review [J], *Energies* **12**(17) (2019), 3254.
[21] H. Wang, Z. Lei, X. Zhang, et al., A review of deep learning for renewable energy forecasting [J], *Energy Conversion and Management* **198** (2019), 111799.
[22] A.S. Ahmad, M.Y. Hassan, M.P. Abdullah, et al., A review on applications of ANN and SVM for building electrical energy consumption forecasting [J], *Renewable and Sustainable Energy Reviews* **33** (2014), 102–109.
[23] G. Habib and S. Qureshi, Optimization and acceleration of convolutional neural networks: A survey [J], *Journal of King Saud University – Computer and Information Sciences*, 2020.
[24] L. Li, C.J. Meinrenken, V. Modi, et al., Short-term apartment-level load forecasting using a modified neural network with selected auto-regressive features [J], *Applied Energy* **287** (2021), 116509.
[25] L.G.B. Ruiz, M.I. Capel and M.C. Pegalajar, Parallel memetic algorithm for training recurrent neural networks for the energy efficiency problem [J], *Applied Soft Computing* **76** (2019), 356–368.
[26] C. Tian, J. Ma, C. Zhang, et al., A deep neural network model for short-term load forecast based on long short-term memory network and convolutional neural network [J], *Energies* **11**(12) (2018), 3493.
[27] M.A.R. Biswas, M.D. Robinson and N. Fumo, Prediction of residential building energy consumption: A neural network approach [J], *Energy* **117** (2016), 84–92.
[28] P. Liu, P. Zheng and Z. Chen, Deep learning with stacked denoising auto-encoder for short-term electric load forecasting [J], *Energies* **12**(12) (2019), 2445.
[29] C. Li, Z. Ding, J. Yi, et al., Deep belief network based hybrid model for building energy consumption prediction [J], *Energies* **11** (2018), 242.
[30] C. Fan, J. Wang, W. Gang, et al., Assessment of deep recurrent neural network-based strategies for short-term building energy predictions [J], *Applied Energy* **236** (2019), 700–710.
[31] M. Sajjad, A.Z. Khan, A. Ullah, et al., A novel CNN-GRU-based hybrid approach for short-term residential load forecasting [J], *IEEE Access* **8** (2020), 143759–143768.
[32] K. Yan, W. Li, Z. Ji, et al., A hybrid LSTM neural network for energy consumption forecasting of individual households [J], *IEEE Access* **7** (2019), 157633–157642.
[33] T. Le, M.T. Vo, T. Kieu, et al., Multiple electric energy consumption forecasting using a cluster-based strategy for transfer learning in smart building [J], *Sensors* **20**(9) (2020), 2668.
[34] T. Kim and S. Cho, Predicting residential energy consumption using CNN-LSTM neural networks [J], *Energy* **182** (2019), 72–81.
[35] T. Le, M.T. Vo, B. Vo, et al., Improving electric energy consumption prediction using CNN and Bi-LSTM [J], *Applied Sciences* **9**(20) (2019), 4237.