Journal of Network Intelligence ©2025 ISSN 2414-8105 (Online)
Taiwan Ubiquitous Information Volume **10**, Number **3**, August **2025**

Prediction Model Based on Deep Learning and Building Energy Consumption Data for the Electrical Energy Consumption of Intelligent Buildings

Jie Fang*, Jie Ling

Ningbo International Trade and Investment Development Co, Ltd, NingBo 315100, China
21760611@zju.edu.cn, lg1314tt@163.com

Ziyin Xu

Ningbo Yinzhou District Construction Engineering Quality Management Service Station
NingBo 315100, China
zyx_sc@163.com

Jingnan Zhao

Department of Civil and Environmental Engineering Rutgers
The State University of New Jersey, Piscataway, USA
jingnanzhao.cee@gmail.com

*Corresponding author: Jie Fang

Received May 13, 2024, revised August 21, 2024, accepted December 2, 2024.

Abstract. *Energy is continuously consumed, and thus, the issue of energy consumption in buildings has become a growing concern. Traditional methods for predicting building energy consumption typically rely on linear models or simple statistics, which cannot effectively capture nonlinear features and long-term dependencies from building energy consumption data. These methods suffer from low accuracy and poor robustness when dealing with complex time series data. To meet the energy management needs of modern intelligent buildings, this study uses the long short-term memory (LSTM) algorithm in deep learning to construct a prediction model for the electrical energy consumption of an intelligent building. This model aims to improve the accuracy and reliability of energy consumption prediction. In the specific implementation, this study performed preprocessing steps on collected building energy consumption data, including data cleaning, missing value filling, and normalization processing. Then, it constructed an energy consumption prediction model architecture based on the LSTM algorithm and added an attention mechanism. Mean squared error (MSE) was used as the loss function during model training, and the parameters were updated using the Adam optimization algorithm. To evaluate model performance, several metrics were used in this study and the constructed model was compared with other algorithmic models. After model testing, MSE was 3.31, mean absolute error was 1.45, the coefficient of determination was 0.9, and mean absolute percentage error was 2.98. Compared with different algorithms, all the algorithm models used in this study have the best indicator levels. The energy consumption prediction model introduced in this work, which combines the LSTM algorithm with an attention mechanism, significantly improves the accuracy and robustness of energy consumption prediction, providing important support for the energy management of intelligent buildings.*

Keywords: Intelligent Building, Energy Consumption, Deep Learning, Attention Mechanism, Long Short-term Memory

1. **Introduction.** The energy problem has always been an important issue that plagues the development of various countries. In such an environment, the prediction and control of building energy consumption have become a major area of research [?, ?]. Traditional building energy consumption prediction methods mostly rely on linear models and simple statistical methods, which frequently fail to effectively capture nonlinear features and long-term dependencies in complex time series data, resulting in the low accuracy of prediction results. Deep learning (DL) methods, particularly long short-term memory (LSTM) algorithms, exhibit advantages in dealing with time series data and are effective tools for solving this problem. LSTM avoids the gradient vanishing problem, which is common in traditional methods, by capturing long-term dependencies through its special gating mechanism [?, ?]. Meanwhile, the powerful nonlinear feature extraction capability of a relevant model can identify key features in complex data patterns, significantly improving the robustness of prediction [?].

This work provides predictive results for controlling building energy consumption by constructing an intelligent prediction model to save energy effectively. The energy consumption prediction model is constructed by combining an attention mechanism with the LSTM algorithm to capture nonlinear features and long-term dependencies in data. In the model implementation process, operations, such as data cleaning, missing value filling, and normalization, are performed one by one on raw data. Mean squared error (MSE) is adopted as the loss function, and the parameters are updated using the Adam optimization algorithm. Finally, the model proposed in this study is evaluated through simulation experiments by using several indicators. The evaluation result that indicates that the model exhibits good prediction performance is attained by analyzing the obtained indicators. After comparing with other algorithmic models, the LSTM + attention algorithmic model is found to have the best level of metrics. The results show that the LSTM algorithm combined with the attention mechanism significantly improves the accuracy and robustness of energy consumption prediction, providing an important support for the energy management of intelligent buildings.

The introduction part of this article outlines the background of the building energy consumption problem and points out the shortcomings of traditional prediction methods. The second part reviews related research and discusses a variety of existing methods and their advantages and disadvantages. The third part describes the design and architecture of the model in detail, including the specific steps of data collection, data preprocessing, and model construction. The fourth part describes the process of model training, including dataset partitioning, loss function selection, optimization algorithms, hyperparameter tuning, and methods for preventing overfitting. The fifth part presents the evaluation results of the model by using multiple metrics to comprehensively measure the predictive performance of the model and compare it with other models. Finally, the article summarizes the major results of this study, points out challenges encountered by the model in practical applications, and proposes future research directions.

2. **Realted work.** In recent years, research on building energy consumption prediction has gradually increased and achieved progress. Shao et al. proposed a home energy consumption prediction model based on a Markov chain and improved prediction accuracy through state transfer analysis [?]. Liu et al. analyzed the application of machine learning methods in energy consumption prediction; then, they compared the advantages and disadvantages of neural networks and support vector machines [?]. Wang et al. used various optimization algorithms to enhance the correlation vector machine model; they constructed an energy model for public buildings with excellent performance and a score

of 0.9523 [?]. Irfan et al. improved the accuracy and robustness of building energy prediction by using a partial least squares (PLS) regression model [?]. Hakawati et al. used PLS structural equation modeling for smart energy prediction and management to improve overall building performance [?]. Although these studies have improved prediction accuracy, most of these methods still exhibit limitations in dealing with complex time series and anomalous data.

DL algorithms have contributed significantly in further improving the accuracy of prediction models, and researchers have started adopting these methods for energy consumption prediction. Zhao et al. explored the application, advantages, and disadvantages of data mining techniques in the field of building energy and proposed future research directions [?]. Zhang et al. optimized residential energy management through multi-intelligence deep reinforcement learning [?]. Zheng et al. proposed a building air-conditioning load prediction method based on principal component analysis (PCA) and the LSTM network; this method combined PCA and the LSTM network and effectively guaranteed the stability of the system [?]. Le et al. proposed an electrical energy consumption prediction model that combined a convolutional neural network (CNN) and bidirectional LSTM for predicting electrical energy consumption; their experimental results showed that the model performed well over multiple time spans [?]. Lin et al. significantly improved the accuracy of energy consumption prediction by enhancing the ability of the LSTM model to capture key features through an attention mechanism [?]. These studies have demonstrated that DL methods exhibit significant advantages in dealing with the building energy consumption prediction problem. However, some challenges remain in model optimization and practical applications.

3. LSTM Energy Consumption Prediction Model Combined with an Attention Mechanism.

3.1. **Data Collection.** Data collection is a fundamental step in building energy consumption prediction models, and the quality of the collected data can affect the accuracy of model training to a certain extent [?]. To ensure the integrity and accuracy of data, various methods are adopted in the data collection process, and strict processing is performed on data. Energy consumption and environmental parameters inside and outside a building are mostly recorded based on power monitoring instruments, temperature and humidity sensors, and equipment operating status. These parameters include the overall power consumption of a building and its various zones, temperature and humidity changes in different areas inside the building, and operating time and power of major electrical equipment. By using various sensors and instruments in a building energy consumption monitoring system to automatically collect data and transmit them in real time to the data center through Internet of things technology, data are stored in a cloud database, ensuring their security while making their access easy for relevant personnel.

This article selects two buildings in a certain residential area as prototypes for simulation experiments, records building energy consumption related data for a period of time, and creates a small dataset for subsequent training. The process of data collection includes data sampling, data integration, feature extraction, and data partitioning. Specifically, it is necessary to extract energy consumption data from the cloud database during this period, while ensuring that the data covers the entire time period without significant missing data [?]. Electricity consumption data, environmental data, equipment data, and time data can be integrated to form a comprehensive dataset, and key features such as total electricity consumption, partition electricity consumption, environmental temperature and humidity, and equipment power can be extracted from the integrated dataset. Among them, the time stamp collects data every 20 min. The recorded types of electricity include total electricity consumption and the electricity consumption of each partition. The environmental parameters are primarily temperature and humidity, and relevant indoor and outdoor parameters are recorded. Table 1 provides the raw data collected during the data collection phase.

TABLE 1. Building energy consumption data

Time Stamp	Total (kWh)	Zone 1 (kWh)	Zone 2 (kWh)	Indoor Temper ature (°C)	Indoor Humidity (%)	outdoor Tempe rature (°C)	outdoor Humidity (%)	HVAC Status (On/Off)
0:00:00	50.2	20.1	30.1	24.5	55	22	60	On
0:20:00	49.8	19.9	29.9	24.6	56	21.8	61	On
0:40:00	49.5	20	29.5	24.4	55	21.7	60	Off
1:00:00	49.3	20.2	29.1	24.2	54	21.5	59	Off
1:20:00	49	19.8	29.2	24.3	55	21.6	58	On
1:40:00	48.8	19.7	29.1	24.5	54	21.4	57	On
2:00:00	48.5	19.5	29	24.3	53	21.3	56	Off
2:20:00	48.2	19.4	28.8	24.1	52	21.2	55	Off
2:40:00	48	19.3	28.7	24	51	21.1	54	On
3:00:00	47.8	19.2	28.6	23.9	50	21	53	On
3:20:00	47.5	19.1	28.4	23.8	49	20.9	52	Off

An example of building energy data collection that covers electricity consumption, environmental conditions, and the operational status of the heating, ventilation, and air-conditioning (HVAC) system at a given point in time in a building is presented in Table 1. Each row of data corresponds to a time stamp, which is a data collection point every 20 min. At 0:00:00, total electricity consumption was 50.2 kilowatt-hour (kWh), with the first area consuming 20.1 kWh and the second area consuming 30.1 kWh, reflecting the distribution of energy consumption in different areas within the building. Meanwhile, indoor temperature was 24.5 °C, humidity was 55%, and outdoor temperature and humidity were 22 °C and 60%, respectively. These environmental parameters are crucial for understanding energy consumption patterns, because they indirectly affect the load demand of an HVAC system. As time progressed, along with the operation of the HVAC system, total power consumption dropped to 49.3 kWh at 1:00:00, and room temperature correspondingly decreased to 24.2 °C. Notably, although the HVAC system was still "on" at 1:40:00, total power consumption and room temperature stabilized at 48.8 kWh and 24.5 °C, respectively, indicating that the system reached a state of equilibrium, i.e., power consumption no longer increased and room temperature remained at a relatively stable level.

These data provide insight into the dynamics of building energy consumption and the interactions between environmental factors and HVAC system operations. Such insight is essential for building accurate energy consumption prediction models. When models capture such nonlinear relationships and long-term dependencies, they can effectively predict future energy consumption trends.

3.2. **Data Preprocessing.** Raw data directly used for model training can most likely introduce noise and increase inaccuracy, affecting the predictive performance of the model; hence, data preprocessing adopts a strict method to ensure the integrity, consistency, and availability of data. The whole process includes data cleaning, data normalization, data transformation, data labeling, and feature selection [?].

In the data cleaning process, outliers in the raw data are first processed. Outliers are usually caused by equipment failures, data acquisition errors, etc., and have very large data differences compared to normal values. The outliers are identified and removed by analyzing the data using Box Plot method (Box Plot) [?]. For power consumption data, if the data points were outside the upper and lower quartile ranges by a factor of 1.5, they were considered as outliers and were eliminated. For missing values, they were filled in by Linear Interpolation to ensure the continuity of the time series data [?]. The interpolation method utilizes the linear relationship between the known data points before and after a missing point to calculate the values of missing data points, ensuring the integrity of the data. The mathematical formula is expressed as

$$x(t) = x(t_0) + \frac{(x(t_1) - x(t_0)) \times (t - t_0)}{t_1 - t_0} \tag{1}$$

where $x(t)$ is the point to be interpolated; $x(t\_0)$ and $x(t\_1)$ are the nearest known data points before and after a missing point, respectively; and t is the time of the point to be interpolated.

By avoiding the effect of data range on model training, all feature data are scaled to the [0,1] interval by using the Min–Max normalization method [?]. The specific formula for normalization is

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

Data transformation performs operations, including time stamp conversion and feature engineering processing, on the original data to extract more representative features from the original data, such as lag and rolling features, which are commonly used in feature engineering [?,?]. Data annotation classifies and annotates data in accordance with actual application scenarios to improve the prediction accuracy and generalization ability of the model [?]. Feature selection directly affects the performance and training effect of the model, and the correlation analysis method is used to evaluate the relationship between each feature and energy consumption, helping select features that are closely related to energy consumption [?]. The Pearson correlation coefficient is used to calculate the correlation between each feature and total energy consumption, and features that exhibit a strong correlation are selected as input to the model with the following mathematical formula:

$$r = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2 \sum(y_i - \overline{y})^2}} \tag{3}$$

Removing features with low correlation or redundancy can reduce data dimensionality, avoid excessive model complexity, and thus, improve training speed and prediction performance. Table 2 provides the preprocessed data.

As indicated in Table 2, the building energy consumption data obtained after a series of preprocessing steps are cleaned, normalized, and converted into a format suitable for model training. The relevant cleaning process is not reflected in Table 2 due to the absence of outliers and missing values in the presented raw data. Each row of data represents the observed value at a time point, listing key characteristics, such as total power consumption, power consumption in each region, and indoor and outdoor temperature and humidity. All data are between 0 and 1 after normalization. When the represented data are the maximum value under this feature, the value is 1. When the represented data is the minimum value under this feature, the value is 0.

TABLE 2. Preprocessed data

serial Number	Total power Consumption	Zone 1 Power Consumption	Zone 2 Power Consumption	Indoor Temperature	Indoor Humidity	Outdoor Temperature	Outdoor Humidity
1	1	0.909	1	0.875	0.857	1	0.889
2	0.852	0.727	0.944	1	1	0.8	1
3	0.741	0.818	0.833	0.75	0.857	0.7	0.889
4	0.667	1	0.722	0.625	0.714	0.6	0.778
5	0.556	0.727	0.778	0.75	0.857	0.5	0.667
6	0.481	0.636	0.722	0.875	0.714	0.4	0.556
7	0.37	0.454	0.667	0.75	0.571	0.3	0.444
8	0.259	0.364	0.556	0.5	0.429	0.2	0.333
9	0.185	0.273	0.5	0.375	0.286	0.1	0.222
10	0.111	0.182	0.444	0.25	0.143	0	0.111
11	0	0.091	0.333	0.125	0	0	0

Through the detailed data preprocessing steps mentioned above, the quality of data is significantly improved, providing a solid data foundation for subsequent model training and prediction. The strict execution of each step ensures the integrity, consistency, and accuracy of data, making the constructed prediction model reliable and precise. Data preprocessing is not only an important part of data analysis, but also a prerequisite and guarantee for building high-performance prediction models.

3.3. **Model Design and Architecture** The major components of the model include two parts: the LSTM layer and the attention layer [**?**, **?**]. The reason for choosing a combination of LSTM and attention mechanism is that LSTM performs well in handling time series data and is able to capture long time dependencies. However, LSTM models alone may miss some important feature information when dealing with long time dependencies. By introducing the attention mechanism, the model is able to better focus on key features, thus improving the accuracy and robustness of the prediction. The attention mechanism assigns weights to different time steps by calculating the degree of match between query vectors, key vectors, and value vectors, enabling the model to identify and utilize important features more accurately. This combination allows the model to not only accurately predict energy consumption, but also to provide the contribution of each time step to the final prediction, thus increasing the interpretability and trust in the model. By visualizing the attention weights, users can intuitively understand the key features that the model focuses on when making predictions. In practical applications, real-time data collection and processing can be realized by deploying on edge computing devices. At the same time, the model can periodically update its parameters to adapt to changes in real-time data streams, thus improving the accuracy and timeliness of predictions. In addition, online learning and incremental learning techniques can be used to enable the model to continuously learn and optimize to improve its adaptability in dynamic environments. To improve the understanding of the role played by each component, a conceptual diagram of LSTM and the attention mechanism is presented in Figure 1.

A conceptual diagram of the structure of the LSTM unit block is shown in Figure 1 (A), while a conceptual diagram of the operation of the attention mechanism is presented in Figure 1 (B). The combination of the two constitutes the model of the LSTM + attention algorithm required for this study. In Figure 1 (A), h, C, and X represent the hidden state,

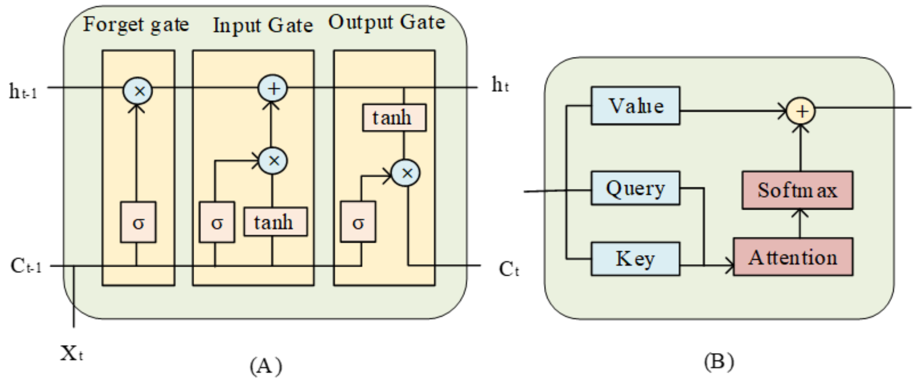

FIGURE 1. (A) LSTM (B) Attention mechanism
Conceptual diagram of the algorithm model structure

cell state, and input data, respectively. Meanwhile, t, σ , and tanh refer to the current moment, sigmoid function, and hyperbolic tangent function, respectively. In Figure 1 (B), Query is the query vector, Key is the key vector, Value is the value vector, and Attention represents the weighted sum process of attention. The LSTM unit block controls the storage and output of information through gating mechanisms, i.e., input, forget, and output gates. Meanwhile, the attention mechanism focuses on key features by calculating the matching degree among query, key, and value vectors. By utilizing the long-term dependency processing capability of LSTM and the key feature focusing ability of the attention mechanism, the accuracy and robustness of prediction are improved, making the model effective in identifying energy consumption patterns.

The input layer of the model receives normalized multidimensional time series data, including features, such as total energy consumption, zonal energy consumption, ambient temperature and humidity, lighting power, and elevator power. To enable the model to process these data, the input data are organized into a 3D (number of samples, time step, and number of features) tensor.

Systematic hyperparameter tuning is performed to ensure optimal model performance. The optimal combination of the number of LSTM cells and layers is selected through grid search and cross-validation [?]. In particular, the weight dimensions that best reflect the key time step features are determined experimentally to ensure the effectiveness of the attention mechanism in identifying important features. In addition, a learning rate scheduling strategy is adopted to start training from higher values and gradually reduce learning rate to ensure the stability and efficiency of model convergence. An appropriate batch size can be determined through experiments to achieve balance between training efficiency and model performance.

Finally, the designed model structure parameters are as follows. The LSTM layer consists of three stacked layers, with each layer containing 128 units. The weight dimension of the attention layer is 64, the initial learning rate is 0.001, and batch size is 32. Through such model design and optimization, an efficient and accurate electrical energy consumption prediction model for intelligent buildings is constructed. This model can provide strong support for energy management in intelligent buildings. Figure 2 shows the structural diagram of the model.

Figure 2 shows the training structure of the electrical energy consumption prediction model for intelligent buildings. The input layer receives preprocessed multidimensional time series data and organizes it into a 3D tensor. The input data are passed to a multilayered stacked LSTM structure, where each layer of LSTM consists of 128 units

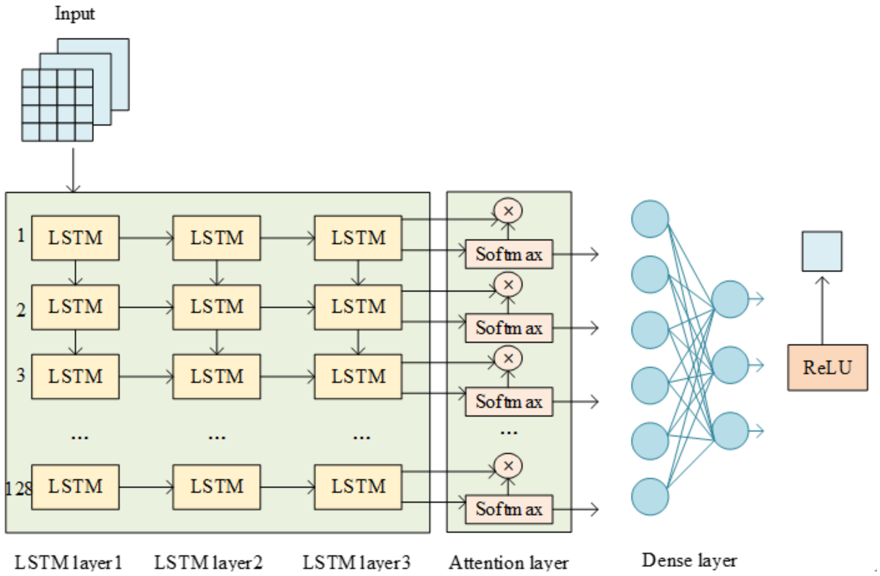

FIGURE 2. Structure diagram of model training

and time dependencies are processed through gating mechanisms. After the output of the LSTM layer enters the attention layer, the importance weights of each time step are calculated and these weights are normalized via the softmax function. The normalized weights are summed with the LSTM output to obtain the output of the attention layer, which can go through the fully connected layer. The output can be mapped to the predicted values via linear transformation for energy consumption prediction. The two core components of the model work together to ensure the performance of the entire model training.

3.4. **Model Training.** This section provides a detailed description of the methods used in model training and the specific process of solving the problem, including measures, such as data preparation, training configuration, and model training. Table 3 provides the distribution of the sample data in the dataset, with the training, testing, and validation sets accounting for 70%, 15%, and 15% of the total data, respectively.

TABLE 3. Sample classification of the dataset

Serial number	Type	Training set	Test set	Validation set
1	Room	1500	320	320
2	Corridor and stairs	850	182	182
3	Underground parking	650	140	140
4	Others	1200	258	258

As indicated in Table 3, 6000 samples are used for the experiment, including 4200 training samples, 900 testing samples, and 900 validation samples. The sample data collection mostly originates from indoor rooms, outdoor corridors and stairs, and underground parking lots. In addition, some difficult-to-classify data sources are uniformly classified into other categories. Among them, corridor stairs and underground parking belong to outdoor, room belongs to indoor, and others contain half of the indoor data and half of the outdoor data. The ratio of the indoor and outdoor data is 1:1, which can effectively measure the electrical consumption level inside and outside the building, better assisting the model in predicting electrical energy consumption.

MSE can be selected as the loss function. It can effectively measure the deviation between the predicted and true values. The formula is

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2 \tag{4}$$

where N is the number of samples, y_i is the true value, and $\widehat{y_i}$ is the predicted value. By minimizing MSE, the model can fit training data accurately. This study uses the Adam optimization algorithm for parameter updating. This algorithm combines the advantages of the momentum method and the root mean square propagation algorithm, and it can adaptively adjust learning rate to improve the convergence speed and stability of the model [?]. The update rule for the Adam optimization algorithm is

$$\theta_{t+1} = \theta_t - \eta\frac{\hat{m}_t}{\sqrt{\hat{v}_t} - \epsilon} \tag{5}$$

where η is the learning rate; $\widehat{m}_t$ and $\widehat{v}_t$ are the deviation correction estimates of momentum and gradient squared, respectively; and ϵ is a constant that prevents zero division errors. Batch size is set to 32, *i. e.* , 32 samples are used for gradient calculation each time the parameters are updated. A batch size of 32 can achieve balance between memory consumption and training efficiency, ensuring efficient model training and avoiding memory shortage issues.

During model training, randomly shuffling training data is necessary to ensure that data distribution is different for each epoch to avoid overfitting the model into specific training samples. Then, batch-by-batch training can be used to input the training data into the model for training in batches of different sizes. During the model training process, the training time of each epoch was recorded to evaluate the training efficiency of the model. The experimental results show that it takes about 2 minutes on average to train one epoch, and the total training time is about 100 minutes. By using an efficient optimization algorithm and reasonable hyperparameter settings, the model achieves good convergence performance in a relatively short time. The input data are processed through the LSTM and attention layers to capture long-term dependencies, calculate the importance weight of each time step in the input sequence, and focus on key features. The accuracy of model prediction can be measured by calculating the loss function value. The gradient of loss relative to the model parameters can be calculated based on the loss function value. Through the backpropagation algorithm, gradient information can be transmitted to each layer of the model to guide parameter updates. The Adam optimization algorithm can be used to update model parameters based on the calculated gradient. The optimal number of LSTM units and layers were selected through grid search and cross-validation, and the weight dimensions of the attention layer were experimentally determined. Meanwhile, a learning rate scheduling strategy was adopted to start training from larger values and gradually reduce the learning rate to ensure the stability and efficiency of model convergence. The above process is iteratively repeated on the training set until the loss function value converges or reaches the preset number of training rounds. At the end of each epoch, the performance of the model is evaluated using validation set data, and the generalization ability of the model is determined through the validation loss.

In this paper, the computational complexity of the model is analyzed.The computational complexity of the LSTM layer is mainly reflected in its gating mechanism and loop structure, while the computational complexity of the attention mechanism is squarely related to the length of the input sequence. In order to reduce the computational cost, the following optimization strategies can be adopted: 1) Reduce the computational complexity by reducing the dimensionality and time step of the input features; 2) Adopt more efficient optimization algorithms, such as distributed training and parallel computation; and 3) Optimize the model by using quantization and pruning techniques, which reduces the number of parameters and the computational volume.

To prevent overfitting of the model during training, various regularization methods are employed, including dropout and early stopping [?]. The introduction of dropout into the LSTM and fully connected layers can prevent the model from overly relying on specific features and improve its generalization ability by randomly discarding some neurons. During the training process, the dropout probability is set to 0.5, *i. e.* , each neuron has a 50% chance of being discarded. By monitoring loss on the validation set, when validation loss no longer decreases for several consecutive epochs, training is terminated in advance to avoid overfitting of the model into the training set [?, ?]. The specific setting is to stop training and save the best model parameters when validation loss fails to decrease within 10 consecutive epochs. Figure 3 shows the parameter changes related to the training process of the model.

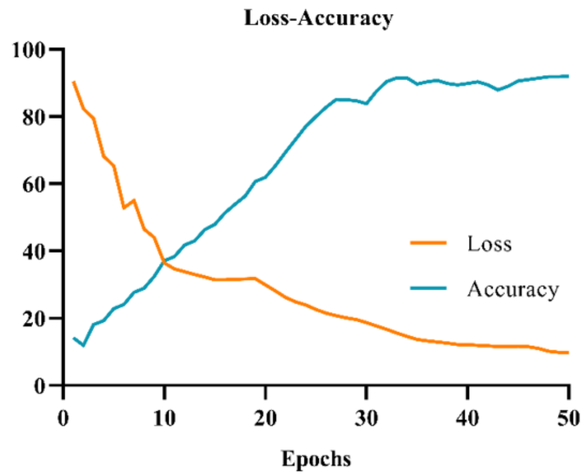

FIGURE 3. Change in loss accuracy during model training

As shown in Figure 3, two curves are used to illustrate the trend of model accuracy and loss over 50 iterations. As the number of iterations increases, the loss value of the model continuously decreases, while accuracy gradually improves, both conditions reflect the learning progress and performance improvement of the model. After approximately 30 iterations, the loss value tends to stabilize and fluctuates within a small range, while accuracy remains stable at around 90%, indicating that the model has reached a good convergence state. This stable accuracy and low loss value signify that the model effectively learns the patterns of the input data during training and can make accurate predictions on new data. Overall, Figure 3 clearly depicts the evolution trajectory of the model during the training process, illustrating that the model reaches the ideal state after multiple iterations and laying a solid foundation for subsequent practical applications.

4. **Model Evaluation.**

4.1. **Model Evaluation Indicators.** During the experimental evaluation process, a comprehensive measure of the model's predictive performance and accuracy was used with a variety of evaluation metrics, each of which played a different role and significance in assessing the performance of the model.

The performance of the energy consumption prediction model proposed in this paper may vary under different geographical locations and climatic conditions. In order to assess the adaptability of the model, experiments are conducted under different geographical locations and climatic conditions to analyze the predictive performance of the model. Preliminary experiments show that the model's prediction accuracy is higher in temperate and subtropical climates, while the model's prediction performance may decrease in extreme climates. This suggests that future research can combine multimodal data to further optimize the model's adaptive ability and improve its prediction accuracy in different environments.

MSE is sensitive to large errors and can effectively capture large prediction deviations.

The mean absolute error (MAE) is insensitive to large errors and provides a smooth picture of the overall model error. In general, the lower the MSE and MAE values, the closer the model's predictions are to the true values. The formula for MAE is

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{6}$$

The coefficient of determination (R^2) is a measure of the model's ability to explain the variation in the true values and takes values within the range [0, 1]. The formula is

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}. \tag{7}$$

R^2 provides the overall performance of the model in fitting data, and it is an important reference for evaluating model performance.

The mean absolute percentage error (MAPE) is an assessment of the percentage error of the predicted value relative to the true value. It can visualize the relative magnitude of the prediction error. The calculation formula is

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} |\frac{y_i - \hat{y}_i}{y_i}| \times 100\% \tag{8}$$

The lower the MAPE value, the higher the prediction accuracy of the model, and the predicted results are closer to the true values. Figure 4 shows the scatterplot under 50 sample predictions, while Figure 5 shows the final metric data of the model on the test set.

The results of the model on 50 samples are presented in Figure 4, visualizing the gap between the predicted and actual energy consumption values. The dotted line in Figure 4 is the baseline, which represents the consistency between the predicted and actual values, and the predicted points are closely distributed around the baseline. This finding indicates that the predicted results of the model are highly consistent with the actual energy consumption values, proving that the model exhibits a high degree of accuracy in predicting the electrical energy consumption of buildings. Among the predicted results, an extremely small number deviated a considerable distance, indicating that some results had relatively large prediction errors in the experiment. This finding is an acceptable error in the experiment. These errors can be gradually reduced in the further optimization of the experimental model in the future to obtain the prediction model with the best predictive performance.

As shown in Figure 5, the final performance indicators of the model on the test set are presented in the form of histograms, including MSE, MAE, R^2, and specific values of MAPE. Among them, the values of MSE, MAE, and MAPE are relatively low, with two

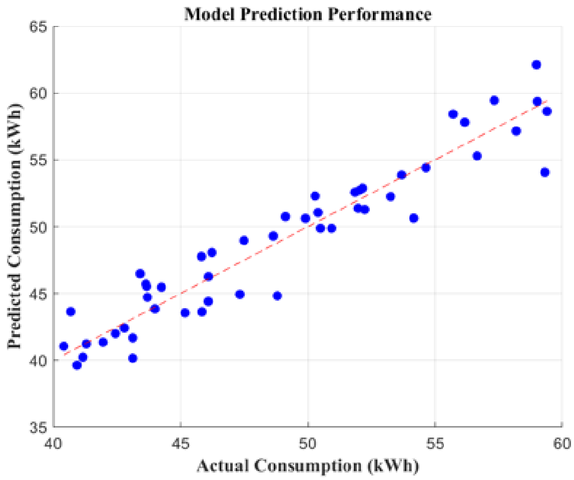

Figure 4. Scatterplot of the predicted results

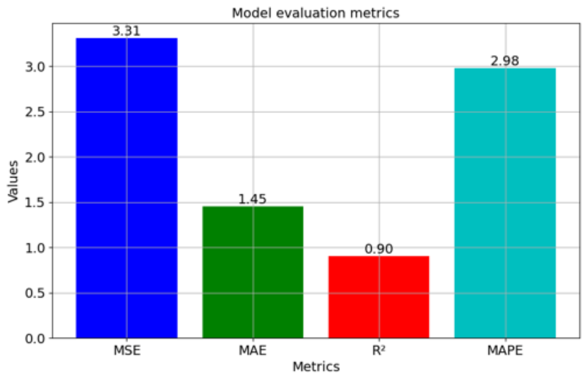

Figure 5. Histogram of the final performance

decimal places of 3.31, 1.45, and 2.98, respectively. The R^2 value is 0.9, which is close to 1. This result not only reflects the high accuracy and stability of the model in the prediction process, but also indicates that the model can efficiently explain the variation of energy consumption data. Hence, the model exhibits excellent predictive ability and generalization performance.

The following conclusion can be drawn from Figures 4 and 5: the energy consumption prediction model that combined the LSTM algorithm with an attention mechanism significantly improves the accuracy and robustness of energy consumption prediction. This study provides strong technical support for energy management in intelligent buildings, verifying the effectiveness and practicality of the model.

4.2. **Comparison of Model Prediction Performance.** To fully evaluate the performance of the prediction models in this study, three common DL models were chosen for comparison. These models are multilayer perceptron (MLP), CNN, and the traditional LSTM model. The evaluation indicators for comparison include MSE, MAE, R^2, and MAPE. To better demonstrate the advantages of the model designed in this study, it was compared with three other algorithms. The parameter conditions of these algorithms are provided in Table 4.

The design scenarios of the four algorithms for comparison are presented in Table 4, including the LSTM model with an attention mechanism, the traditional LSTM model without an attention mechanism, the CNN model, and the MLP model. The traditional

Table 4. Comparison parameters of different algorithms

Serial number	Algorithm	Activation function	Introduction of channel attention mechanism	Fully connected layer
1	LSTM	ReLU	Yes	Dense layer
2	LSTM	ReLU	No	Dense layer
3	CNN	ReLU	No	Dense layer
4	MLP	ReLU	No	Dense layer

LSTM model is used to verify the importance of the attention mechanism in the model presented in this study and compare the changes in model testing metrics with and without an attention mechanism. In addition, all the four models use the rectified linear unit (ReLU) as the linear activation function and the dense layer to fuse features, reducing design differences among models and making the performance comparison among algorithms more evident.

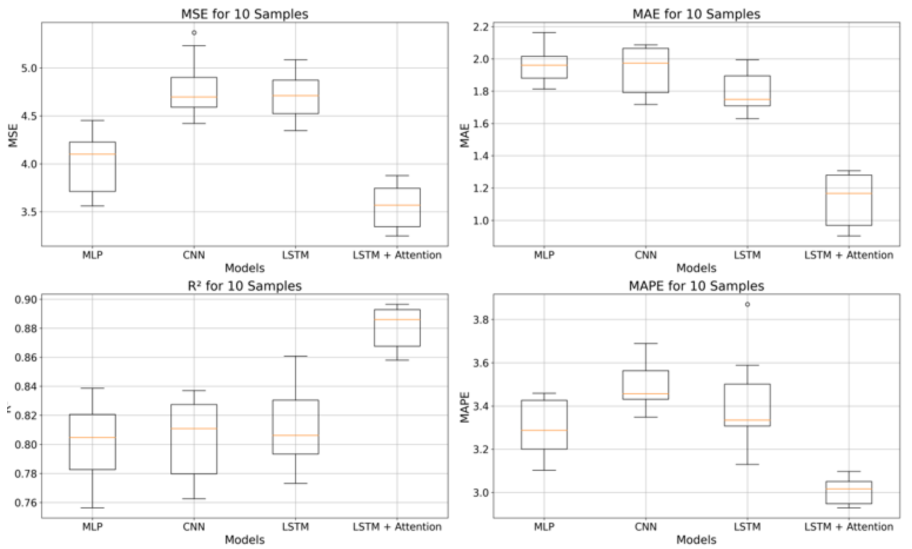

Figure 6. Performance comparison of different algorithms

Figure 6 presents a comparison of the predictive performance of four different DL models (MLP, CNN, traditional LSTM model, and LSTM model with an attention mechanism) on ten random samples. The LSTM model with an attention mechanism proposed in this study performs the best in all four evaluation metrics. The LSTM model with an attention mechanism has the smallest MSE, MAE, and closest R^2 to 1, indicating that it has the smallest prediction error, the highest prediction accuracy, and is able to better explain data changes compared with the three other models. The MAPE of the proposed model is also the smallest among the four models. The smaller the MAPE, the higher the prediction accuracy of the model. The above comparison shows that the LSTM model with an attention mechanism is considerably stronger than the four other classes of algorithmic models in terms of prediction performance, and it can be helpful in this field.

5. **Conclusions.** The electrical energy consumption prediction model based on the LSTM network and an attention mechanism for intelligent buildings successfully combines DL algorithms with the characteristics of building energy consumption data, effectively improving the accuracy and reliability of energy consumption prediction. Through systematic data preprocessing and model design, the model proposed in this study performs well in evaluation metrics, such as MSE, MAE, R^2, and MAPE, demonstrating significant advantages over traditional models. The energy consumption prediction model based on LSTM and attention mechanism proposed in this paper not only excels in prediction accuracy and robustness, but also has a wide range of application potential in energy optimization and energy saving strategy formulation. By accurately predicting building energy consumption, it can provide data support for energy management of intelligent buildings and help formulate more scientific energy-saving strategies. For example, the operating time and intensity of air conditioning and lighting equipment can be dynamically adjusted according to the prediction results, thus achieving the purpose of energy saving and consumption reduction. In addition, the model can be used to monitor abnormal energy consumption, identify and solve energy consumption problems in a timely manner, and further improve energy utilization efficiency. Although the model presented in this study has achieved good results in energy consumption prediction, it still experiences challenges in practical applications. First, the performance of the model depends on a large amount of high-quality historical data, but in practical environments, the integrity and accuracy of data may be difficult to guarantee. In addition, the robustness of the model in handling extreme weather or sudden events still requires further validation. Future research directions include the following: (1) further optimizing data collection and preprocessing methods to enhance the predictive ability of models in the absence of data, (2) exploring multimodal data fusion to improve the adaptability of models in complex environments, and (3) combining real-time monitoring and feedback mechanisms to achieve dynamic optimization and management of energy consumption in intelligent buildings. Through continuous improvement, the proposed model is expected to provide comprehensive and effective technical support for the energy-saving management of intelligent buildings.

REFERENCES

[1] S. Birkha Mohd Ali, M. Hasanuzzaman, N. Rahim, M. Mamun, and U. Obaidellah, "Analysis of energy consumption and potential energy savings of an institutional building in Malaysia," *Alexandria Engineering Journal*, vol. 60, no. 1, pp. 805–820, 2021.

[2] E. Masanet, A. Shehabi, N. Lei, S. Smith, and J. Koomey, "Recalibrating global data center energy-use estimates," *Science*, vol. 367, no. 6481, pp. 984–986, 2020.

[3] S. M. Al-Selwi, M. Fadzil, S. J. Abdulkadir, and A. Muneer, "Lstm inefficiency in long-term dependencies regression problems," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:259819976

[4] S. Poornima and M. Pushpalatha, "Prediction of Rainfall Using Intensified LSTM Based Recurrent Neural Network with Weighted Linear Units," *Atmosphere*, vol. 10, no. 11, p. 668, 2019.

[5] T. Guo, L. Wu, C. Wang, and Z. Xu, "Damage detection in a novel deep-learning framework: a robust method for feature extraction," *Structural Health Monitoring*, vol. 19, no. 2, pp. 424–442, 2020.

[6] C. Shao, D. Zhaobin, E. Chauveau, and C. Lidan, "Household load forecasting model based on markov state transition," *Power Demand Side Management*, vol. 23, no. 01, pp. 55–60, 2021.

[7] Z. Liu, D. Wu, Y. Liu, Z. Han, L. Lun, J. Gao, G. Jin, and G. Cao, "Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction," *Energy Exploration & Exploitation*, vol. 37, no. 4, pp. 1426–1451, 2019.

[8] H. Wang, W. Wen, Z. Zhang, and N. Gao, "Construction of Building Energy Consumption Prediction Model Based on Multi-Optimization Model," *Buildings*, vol. 13, no. 7, p. 1677, 2023.

[9] Muhammad Irfan and Faizir Ramlie, "Analysis of Parameters which Affects Prediction of Energy Consumption in Buildings using Partial Least Square (PLS) Approach," *Journal of Advanced Research in Applied Sciences and Engineering Technology*, vol. 25, no. 1, pp. 61–68, 2021.

[10] B. Hakawati, A. Mousa, and F. Draidi, "Smart energy management in residential buildings: the impact of knowledge and behavior," *Scientific Reports*, vol. 14, no. 1, p. 1702, 2024.

[11] Y. Zhao, C. Zhang, Y. Zhang, Z. Wang, and J. Li, "A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis," *Energy and Built Environment*, vol. 1, no. 2, pp. 149–164, 2020.

[12] H. Zhang, X. Shen, H. Mu *et al.*, "Online optimal scheduling of residential energy consumption based on multi-agent asynchronous deep reinforcement learning," *Chinese Journal of Electrical Engineering*, vol. 40, no. 1, pp. 117–127, 2020.

[13] M. Zheng, K. Yang, C. Shang, and Y. Luo, "A PCA-LSTM-Based Method for Fault Diagnosis and Data Recovery of Dry-Type Transformer Temperature Monitoring Sensor," *Applied Sciences*, vol. 12, no. 11, p. 5624, 2022.

[14] T. Le, M. Vo, B. Vo, E. Hwang, S. Rho, and S. Baik, "Improving Electric Energy Consumption Prediction Using CNN and Bi-LSTM," *Applied Sciences*, vol. 9, no. 20, p. 4237, 2019.

[15] Z. Lin, L. Cheng, and G. Huang, "Electricity consumption prediction based on LSTM with attention mechanism," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 15, no. 4, pp. 556–562, 2020.

[16] M. K. Alam, "A systematic qualitative case study: questions, data collection, NVivo analysis and saturation," *Qualitative Research in Organizations and Management: An International Journal*, vol. 16, no. 1, pp. 1–31, 2020.

[17] G. Li, H. Dong, and C. Zhang, "Cloud databases: new techniques, challenges, and opportunities," *Proceedings of the VLDB Endowment*, vol. 15, no. 12, pp. 3758–3761, 2022.

[18] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, 2022.

[19] X. Sun, "Application of boxplot in examination results based on psychological evaluation," *Global Transitions Proceedings*, vol. 33, no. 8, pp. 459–463, 2021.

[20] Y. Liu, Y. Xue, and M. Taniguchi, "Robust Linear Interpolation and Extrapolation of Stationary Time Series in L^{p}," *Journal of Time Series Analysis*, vol. 41, no. 2, pp. 229–248, 2020.

[21] M. Razaviyayn, T. Huang, S. Lu, M. Nouiehed, M. Sanjabi, and M. Hong, "Nonconvex Min-Max Optimization: Applications, Challenges, and Recent Theoretical Advances," *IEEE Signal Processing Magazine*, no. 5, pp. 55–66, 2020.

[22] Z. Chen, P. Zhao, F. Li, T. T. Marquez-Lago, A. Leier, J. Revote, Y. Zhu, D. R. Powell, T. Akutsu, G. I. Webb, K.-C. Chou, A. I. Smith, R. J. Daly, J. Li, and J. Song, "iLearn: an integrated platform and meta-learner for feature engineering, machine-learning analysis and modeling of DNA, RNA and protein sequence data," *Briefings in Bioinformatics*, vol. 21, no. 3, pp. 1047–1057, 2019.

[23] B. Venkatesh and J. Anuradha, "A Review of Feature Selection and Its Methods," *Cybernetics and Information Technologies*, vol. 19, no. 1, pp. 3–26, 2019.

[24] N. F. Greenwald, G. Miller, E. Moen, A. Kong, A. Kagel, T. Dougherty, C. C. Fullaway, B. J. McIntosh, K. X. Leow, M. S. Schwartz, C. Pavelchek, S. Cui, I. Camplisson, O. Bar-Tal, J. Singh, M. Fong, G. Chaudhry, Z. Abraham, J. Moseley, S. Warshawsky, E. Soon, S. Greenbaum, T. Risom, T. Hollmann, S. C. Bendall, L. Keren, W. Graf, M. Angelo, and D. Van Valen, "Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning," *Nature Biotechnology*, vol. 40, no. 4, pp. 555–565, 2022.

[25] X. Zhuang, Z. Yang, and D. Cordes, "A technical review of canonical correlation analysis for neuroscience applications," *Human Brain Mapping*, vol. 41, no. 13, pp. 3807–3833, 2020.

[26] K. Smagulova and A. P. James, "A survey on LSTM memristive neural network architectures and applications," *The European Physical Journal Special Topics*, vol. 228, no. 10, pp. 2313–2324, 2019.

[27] M.-H. Guo, T.-X. Xu, J.-J. Liu, Z.-N. Liu, P.-T. Jiang, T.-J. Mu, S.-H. Zhang, R. R. Martin, M.-M. Cheng, and S.-M. Hu, "Attention mechanisms in computer vision: A survey," *Computational Visual Media*, vol. 8, no. 3, pp. 331–368, 2022.

[28] T. Yan, S.-L. Shen, A. Zhou, and X. Chen, "Prediction of geological characteristics from shield operational parameters by integrating grid search and k-fold cross validation into stacking classification algorithm," *Journal of Rock Mechanics and Geotechnical Engineering*, vol. 14, no. 4, pp. 1292–1303, 2022.

[29] D. Yi, J. Ahn, and S. Ji, "An Effective Optimization Method for Machine Learning Based on ADAM," *Applied Sciences*, vol. 10, no. 3, p. 1073, 2020.

[30] S. Gazzola, P. C. Hansen, and J. G. Nagy, "IR Tools: a MATLAB package of iterative regularization methods and large-scale test problems," *Numerical Algorithms*, vol. 81, no. 3, pp. 773–811, 2019.

[31] X. Fan, S. Zhao, X. Zhang, and L. Meng, "The Impact of Improving Employee Psychological Empowerment and Job Performance Based on Deep Learning and Artificial Intelligence:," *Journal of Organizational and End User Computing*, vol. 35, no. 3, pp. 1–14, 2023.

[32] G. Wang, X. Zhang, Y. Gao, A. L. Yee, and X. Wang, "The Use of an Internet of Things Data Management System Using Data Mining Association Algorithm in an E-Commerce Platform:," *Journal of Organizational and End User Computing*, vol. 35, no. 3, pp. 1–19, 2023.