Article

High-Dimensional Energy Consumption Anomaly Detection: A Deep Learning-Based Method for Detecting Anomalies

Haipeng Pan, Zhongqian Yin and Xianzhi Jiang *

School of Mechanical and Automatic, Zhejiang Sci-Tech University, Hangzhou 310018, China
* Correspondence: xianzhi@zstu.edu.cn

Abstract: With the increase of energy demand, energy wasteful behavior is inevitable. To reduce energy waste, it is crucial to understand users' electricity consumption habits and detect abnormal usage behavior in a timely manner. This study proposes a high-dimensional energy consumption anomaly detection method based on deep learning. The method uses high-dimensional energy consumption related data to predict users' electricity consumption in real time and for anomaly detection. The test results of the method on a publicly available dataset show that it can effectively detect abnormal electricity usage behavior of users. The results show that the method is useful in establishing a real-time anomaly detection system in buildings, helping building managers to identify abnormal electricity usage by users. In addition, users can also use the system to understand their electricity usage and reduce energy waste.

Keywords: deep learning; anomaly detection; time series analysis; high-dimensional energy consumption

1. Introduction

Energy is essential for people's social life and scientific and technological development. With social development and technological innovation, the demands of society for energy are increasing. Of these, commercial and residential buildings account for a significant proportion of energy consumption. With buildings accounting for approximately 40% of energy consumption, 36% of energy-related greenhouse gas emissions, and 80% of the energy consumed by citizens for heating, cooling, and domestic hot water in Europe, buildings are the largest single consumer of energy in Europe [1]. In addition, commercial and residential buildings are major contributors to global carbon emissions. Energy for buildings is mainly derived from the burning of coal, oil, and natural gas. The use of fossil fuels increases carbon dioxide emissions, contributes to climate warming, and leads to accelerated environmental degradation. In the current form of energy, reducing energy consumption has become a goal that we have to achieve.

For commercial and residential buildings, energy consumption is concentrated in air conditioning, lighting systems, and various modern appliances. As the quality of life has improved, people's electricity needs have become more diverse. The increasing variety of modern appliances used in life and the consequent increase in the use of electrical equipment has increased the share of energy consumption in buildings year on year. Final energy use in buildings increased from 118 EJ in 2010 to almost 130 EJ in 2019 at an average annual rate of 1%. The fastest-increasing end uses of energy in buildings—for space cooling, appliances, and electric plug-loads—drive buildings sector electricity demand growth. While electricity made up one-third of building energy use in 2020, fossil fuel use has also increased at a marginal annual average growth rate of 0.7% since 2010 [2]. It is therefore particularly important to understand the electricity consumption of buildings to reduce CO_2 emissions and energy consumption. To analyze the electricity consumption of a building, detailed building electricity and environmental data need to be obtained. Usually, data are obtained by installing appropriate sensors at various locations in the building. However, in

practice, this traditional detection method is easily rejected by users because it violates the privacy of others. In addition, distributed installations are more expensive, cumbersome to count, and do not allow building managers easy access to analytical data. The advent of smart meters has adequately addressed the problems of the traditional approach. Smart meters use fewer sensors, can record electricity consumption at hourly or even shorter intervals, and have more than just a billing function. Instead of collecting data on-site, managers can obtain the corresponding energy consumption data in real-time through wireless signal transmission, facilitating energy consumption statistics and analysis [3].

Since abnormal power consumption behavior results in higher electricity consumption and wasted energy, identifying abnormal power consumption behavior of users from the obtained energy consumption data allows for more efficient use of electricity [4,5]. There are many reasons for abnormal electricity consumption by end-users, for example, damage to equipment, wasteful behavior by end-users (forgetting to switch off the equipment after use or using incorrectly configured equipment), electricity theft attacks [6,7]. As a result, end-users can check their electrical equipment in terms of abnormal electrical behavior and develop good electrical habits. The method of finding patterns in the data that do not conform to expected or normal behavior is generally referred to as anomaly detection. Identifying abnormal energy consumption increases by specific end users should be seen as a way of early warning to reduce energy waste in buildings [8].

The methods for anomaly detection can be classified into distance-based methods, density-based methods, dimensionality reduction-based methods, and deep learning-based methods.

Among the distance-based methods, the K-Nearest Neighbor (KNN) algorithm is one of the more popular methods. This algorithm calculates the average distance between each sample point and its nearest K samples in turn and then uses the calculated distance for anomaly detection [9,10]. The distance-based approach, although effective in some cases, performs better with a priori knowledge of the anomaly duration and the number of anomalies.

The density-based approach is to investigate the density of each power consumption pattern and its neighbors. Among them, the local density cluster-based outlier factor (LDCOF) applies the concept of local density in assigning anomaly scores. Refs. [11,12] used the density-based spatial clustering of applications with noise (DBSCAN) method to detect anomalous power consumption in the wind farm environment. However, the density-based approach cannot take into account time correlation and is therefore not applicable to multivariate time series data.

The method based on dimensionality reduction can be used as a classification method that removes irrelevant power patterns and redundancies, possessing a low computational cost [13]. Principal component analysis (PCA) is a multivariate data analysis method that preserves as much as possible the relationships between data extracted from process measurements and reduces the dimensionality of a large number of raw data [14]. However, methods based on dimensionality reduction are only valid for highly correlated data and require that the data follow a multivariate Gaussian distribution [15].

In recent years, deep learning-based methods have been widely used, and work on anomaly detection of time series data has increased significantly [16].

First, convolutional neural networks (CNNs) have proven their effectiveness in different research applications and have superior performance in detecting time series data anomalies compared to artificial neural network (ANN) algorithms. In [17], the authors propose a new anomaly detection technique, FuseAD, which utilizes a statistical ARIMA (Autoregressive Integrated Moving Average model) and convolutional neural network (CNN) based approach to fusing them in a residual manner. The results obtained show that this fusion-based technique can achieve the best of both by combining their strengths and complementing their weaknesses. In addition, deep CNNs can accurately identify the non-periodicity of electricity theft and the periodicity of normal electricity consumption based on two-dimensional (2D) electricity consumption data, solving the problem of low

accuracy when detecting electricity theft [18]. In [19], the authors use convolutional neural networks for feature extraction and then use random forest algorithms to detect electricity theft to help utilities solve the problem of inefficient electricity detection and irregular energy consumption.

On the other side, Recurrent Neural Networks (RNN) also have excellent performance in time series data prediction, especially LSTM (Long Short-Term Memory) networks. As in [20], the authors use deep learning algorithms to remove seasonality and trends from data for better anomaly detection, helping electric utilities to minimize the impact of uncaptured errors in their daily work. Meanwhile, in [21], the authors propose a power consumption prediction and anomaly detection algorithm based on LSTM neural network, which focuses on seasonal and monthly trends, resulting in a significant improvement in power theft identification. Ref. [22] predicted the system energy consumption using pattern decomposition based on the LSTM algorithm and detected abnormal system energy consumption by Grubbs test using the difference between the predicted and actual values, which effectively reduced the energy waste during the system operation. In [23], the authors combined OC-SVM (one class-support vector machine) and SVDD (support vector data description), based on the generic structure of LSTM, with modified formulas to achieve efficient anomaly detection, especially for time series data, capable of handling variable length data sequences.

For the problems of traditional machine learning methods that do not apply to multiple variables, require prior knowledge, and require data to follow a multivariate Gaussian distribution, we follow the successful prospects of deep learning-based anomaly detection methods by combining CNN, Bi-LSTM (Bidirectional Long ShortTerm Memory), and attention mechanisms with a 3σ criterion to propose a new energy consumption anomaly detection method. CNN can extract higher-order features from the input data. Bi-LSTM network has the advantage of acquiring contextual information of time series data compared to the LSTM network, which combines information in both forward and backward directions. In addition, attention mechanisms have been very successful in the fields of machine translation and image description generation. We use it to assign different weights to different hidden units of the neural network to make the hidden layer focus more on the key information in the sequence data.

The method applies CNN, Bi-LSTM, and attention mechanism to the prediction model mines the contextual information in historical high-dimensional energy consumption data and the contribution of different feature dimensions to the prediction results, and then uses the 3σ criterion to make energy consumption outlier judgments.

Therefore, this study develops a method to identify abnormal power consumption behavior of customers in real-time using high-dimensional energy consumption data through experiments. The experimental data includes electricity consumption as well as several environmental parameters that affect electricity consumption. The aim is to detect abnormal user energy consumption behavior in real-time based on high-dimensional energy consumption data. Therefore, the results of this research have potential application in an IoT-based energy management system. In addition, the results of this research are not only applicable to the detection of anomalies in electricity consumption but also have potential applications in the detection of anomalies based on other sensor data. The main contributions of this work can be summarized as follows:

1.　A deep learning based HDEC-AD (High Dimensional Energy Consumption) method for identifying abnormal energy consumption behavior of users;
2.　The method is divided into two stages. The first stage is the prediction stage, where the power at the next moment is predicted from the high-dimensional power-related data collected in real time. The second stage is the anomalous pattern detection stage, where the predicted values are compared with the actual values and the prediction error is calculated and defined as an anomalous activity when it exceeds 3SDs (standard deviations);

3. Anomaly detection helps to build managers understanding users' daily electricity consumption patterns so they can plan reasonable electricity demand, while users can analyze electricity costs from the anomaly results and thus reduce energy waste.

The rest of the paper is structured as follows: In Section 2, we describe the materials and methods. In Section 3, we perform relevant experimental tests and analysis of the results. In Section 4, we provide a discussion. In Section 5, we summarize the full paper and provide an outlook for the future.

2. Materials and Methods

In this section, we explain the various components of the model. Among them, the main body of the prediction model consists of CNN, Bi-LSTM, an attention mechanism, which is used to predict the value of the next time point in a given time series. The predicted values are further passed to the anomaly detection module, which will determine whether the data point is anomalous or not.

This experiment is implemented using the Keras deep learning library, running on the Google TensorFlow framework [24,25]. The hardware uses RTX 2060super for GPU acceleration and AMD Ryzen 7 3700X for the CPU.

As shown in Figure 1, the HDEC-AD proposed includes a prediction model. The model uses high-dimensional data related to the energy consumption of 144 sets of users with a time interval of 10 min to predict the electricity consumption at the next moment. The prediction results are evaluated using Mean Absolute Percentage Error (MAPE) and Mean Absolute Error (MAE) to assess accuracy.
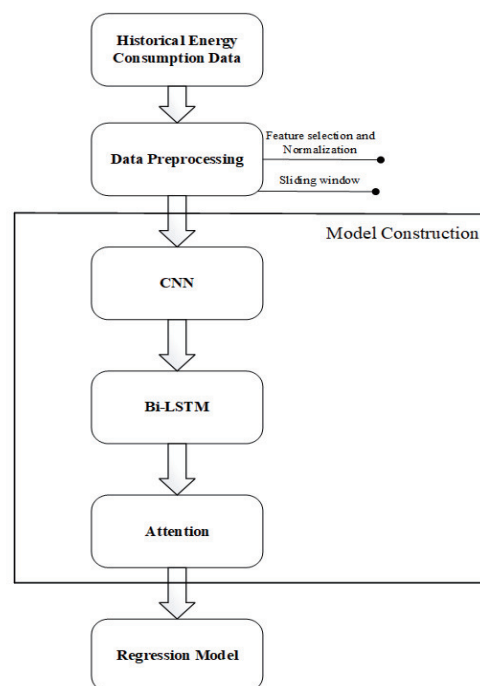
Figure 1. Process of model training for energy consumption forecasting.

As shown in Figure 2, abnormal user energy consumption behavior is then identified by monitoring the difference between predicted consumption and actual consumption. The difference between the predicted consumption and the actual consumption is calculated, and, if it is greater than three times the standard deviation of the actual consumption for the previous 144 moments, then the electricity consumption is identified as abnormal and the next 24 h containing that moment are considered as abnormal electricity consumption periods.
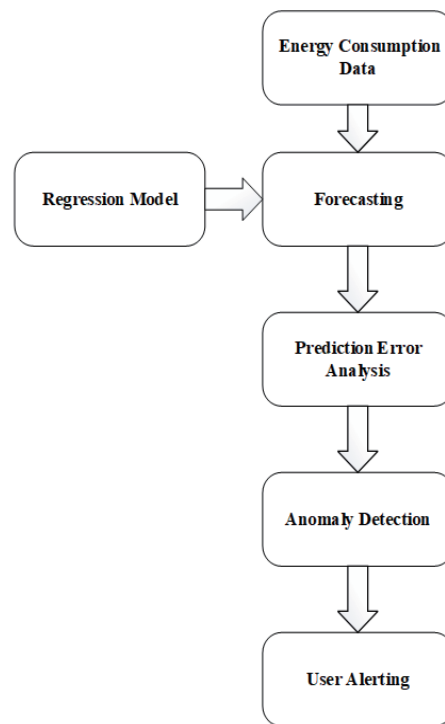
Figure 2. Process of anomaly detection.

2.1. Convolutional Layer

Convolutional neural networks have representational learning capabilities and are capable of extracting higher-order features from the input information. The convolution layer consists of several feature filters, which are used to compute different feature mappings. Each neuron in the convolution layer connects a local region in the previous layer, and the convolutional result is obtained by summing the input features by doing matrix element multiplication and superimposing the amount of deviation. The ReLU activation function is then applied to the convolution result. As shown in Equation (1), the output of the convolutional layer can be expressed as:

$$
y_{conv}(X_l^{f_1}) = \delta \left(\sum_{f_l=1}^{F_l} W_l^{f_1} * X_l^{f_1} + b_l^{f_1} \right), \tag{1}
$$

where δ is the activation function, $*$ is the convolution operation, $X_l^{f_1}$ is the input of the f-th feature filter, and both $W_l^{f_1}$ and $b_l^{f_1}$ are learnable parameters in the f-th feature filter.

2.2. Dropout Layer

In deep learning, models are prone to overfitting when there are too many parameters. Overfitting is a common problem with much deep learning and even machine learning algorithms, as evidenced by high prediction accuracy on the training set and a significant drop in accuracy on the test set. The basic idea of Dropout is shown in the figure. During training, each neuron is retained with probability p (stop working with probability $1 - p$), and each forward propagation retains a different neuron, which allows the model to be less dependent on certain local features and has better generalization performance. During testing, each parameter is also multiplied by p to ensure the same output expectation.

2.3. Bi-LSTM

Traditional (feed-forward) neural networks assume that the data are independent in time. However, this assumption does not apply to continuous time-series data. Therefore, for time series data, recurrent neural networks (RNNs) are commonly used. Recurrent

Neural Network (RNN) is a class of recursive neural network that takes sequence data as input, recursion in the direction of sequence evolution, and all nodes (recurrent units) are connected in a chain. However, in the case of Long Term Dependencies, RNNs modeling sequential data will face the problem of gradient disappearance. Therefore, long and short-term memory networks are used to solve this problem. The LSTM model is composed of an input x_t at moment t, a cell state C_t, a temporary cell state $\tilde{C}_t$, a hidden state h_t, an oblivion gate f_t, a memory gate it, and an output gate o_t. The computational process of the LSTM can be summarized by forgetting information in the cell state and remembering new information so that information useful for subsequent moments of computation is passed on, while useless information is discarded, and the hidden state h_t is output at each time step, where forgetting, memory, and output are controlled by the forgetting gate f_t, the memory gate it, and the output gate o_t, calculated from the implicit state h_{t-1} at the previous moment and the current input of x_t. The overall framework and the memory updates for each time step t are calculated as follows [26]:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{2}$$

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{3}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{4}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{5}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{6}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{7}$$

$$h_t = o_t * \tanh(C_t) \tag{8}$$

In the previous equation, i_t, f_t, o_t, C_t, and h_t denote input gates, oblivion gates, output gates, storage cells, and hidden states, respectively, and $*$ denotes the product of elements. The other parameters are the weight matrices to be learned, shared between all time steps.

LSTMs can better capture dependencies over longer distances, but cannot integrate temporal information about the future, so bi-directional long and short-term memory neural networks (Bi-LSTMs) were chosen to solve this problem. The Bi-LSTM consists of a forward LSTM and a backward LSTM. The input sequence is fed into the two LSTM neural networks in forward and reverse order respectively for feature extraction, and the two output vectors (i.e., the extracted feature vectors) are stitched together to form the final feature representation. The model design concept of the Bi-LSTM is to make the feature data obtained at the moment t have information between the past and the future at the same time.

2.4. Attention Mechanisms

The introduction of an attention mechanism allows for better capturing of information about the entire sequence, selectively focusing on the state of the relevant vectors. The attention model takes the output of the Bi-LSTM as input, places a weight α_{ts} on each moment, α_{ts} is determined by the similarity between the current vector state ht and all vector states $\bar{h} = (h_1, h_2, \ldots, h_s)$, and outputs a series of contextual vectors c_t with the same length [27]. The calculation of the weights and context vectors can be calculated as shown in Equations (9) and (10):

$$\alpha_{ts} = \frac{exp(score(h_t, \bar{h}_s))}{\sum_{s'=1}^{S} exp(score(h_t, \bar{h}_{s'}))} \tag{9}$$

$$c_t = \sum_s \alpha_{ts} \bar{h}_s \tag{10}$$

where α_{ts} is a weight on each moment, c_t is a series of contextual vectors, h_t is the current vector state, and $\bar{h}$ are all vector states.

2.5. Flatten and Dense Layers

In the last part of the model, a *Flatten* layer is used to 'flatten' the input, i.e., to make the multidimensional input one-dimensional for the transition to the fully connected layer, and *Flatten* does not affect the size of the batch. Finally, the fully connected layer uses a *sigmoid* activation function with a one-dimensional output result.

2.6. Anomaly Detection

We assess the trend in electricity consumption over time by using the standard deviation, setting a threshold value of 3σ above the predicted value, where σ is the standard deviation of electricity consumption on the day before the actual moment [28]. A value higher than the threshold for predicted electricity consumption at the actual moment indicates an abnormal state. σ and $Y_{threshold}$ are calculated as shown in Equations (11) and (12):

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \tag{11}$$

$$Y_{threshold} = \widehat{Y} + 3\sigma \tag{12}$$

where σ is the standard deviation, $Y_{threshold}$ is the threshold, $\widehat{Y}$ is the predicted value, x_i is the electricity consumption, $\bar{x}$ is the average electricity consumption, and n is the number of samples.

3. Experimental Results

3.1. Data Set and Pre-Processing

This experiment uses a UCI (University of California, Irvine, CA, USA) appliances energy prediction data set. In total, the UCI (University of California, Irvine) appliances energy prediction dataset recorded data for houses from 11 January 2016, 5:00 p.m. to 27 May 2016, 6:00 p.m. The read interval of the data was 10 min, and the total number of samples was 19,735. The house temperature and humidity conditions were monitored with a ZigBee wireless sensor network. Each wireless node transmitted the temperature and humidity conditions around 3.3 min. Then, the read interval of the data was 10 min. The energy data were logged every 10 min with m-bus energy meters. Weather from the nearest airport weather station (Chievres Airport, Chievres, Belgium) was downloaded from a public data set from Reliable Prognosis (rp5.ru) and merged with the experimental data sets using the date and time column. Two random variables have been included in the data set for testing the regression models and to filter out non-predictive attributes (parameters).

In the data pre-processing step, we select eight features in the dataset as raw data, which include energy use, energy use of light fixtures, temperature, pressure, Humidity, Wind speed, Visibility, and dewpoint. After comparison tests, it can be seen from Tables 1–6 that the method of using the first 80% of the original data as the training set and the last 20% as the test set makes the MAE and MAPE smaller. Therefore, we choose that data split ratio. The time range for the training set is 11 January 2016, 5:00 p.m. to 1 May 2016, 7:50 a.m., and the time range for the test set is 1 May 2016, 8:00 a.m. to 27 May 2016, 6:00 p.m. Then, we need to convert different types of data to the same specification and "nondimensionalize" the data to speed up the model solving. Therefore, we normalize all the training set features by first centering the data on the minimum and then scaling them by the range (difference between the maximum and the minimum), with convergence in the interval [0, 1]. The window length K = 144 means 144 time steps to generate a sample, and the step length (step) S = 1 means the window slides one time step to generate a sample.

Selection of hyperparameters based on hyperparameter studies, common settings, or through experimentation, batch size 64 was used. For the high-dimensional energy consumption data in this experiment, the Adam optimizer converges faster, and the convergence process is more stable. Optimization was performed using the Adam optimizer. The learning rate was 0.01. CNN has 128 units. Bi-LSTM has 128 units, 64 in each direction.

In practical applications, missing data and outliers may occur in sensor acquisition data. For missing data cases, consider the mean value fill approach. If the null value is a numeric attribute, use the average of the values taken by the attribute in all other objects to fill the missing attribute value; if the null value is a non-numeric attribute, use the value with the highest frequency of the attribute in all other objects to fill the missing attribute value. For outliers, you can treat the outlier as a missing value and use the average value fill method to handle it.

The performance of our model improves as the size of the dataset increases. A larger dataset can better capture the electricity consumption habits of users, and accordingly, the model training time will be longer. If the dataset is too small, it will not be able to capture the user's electricity consumption habits well and the model will be less applicable.

Table 1. Evaluation of model performance with different data division ratio with *lstmunits* set to 64 and *epochs* set to 20.

Division Ratio	MAE	MAPE	Training Time
9:1	25.40	18.83%	2250 s
8:2	23.10	18.55%	2008 s
7:3	24.59	21.77%	1697 s
6:4	28.22	22.08%	1474 s

Table 2. Evaluation of model performance with different data division ratio with *lstmunits* set to 64 and *epochs* set to 10.

Division Ratio	MAE	MAPE	Training Time
9:1	25.08	19.40%	1234 s
8:2	24.55	19.93%	1048 s
7:3	25.25	20.55%	991 s
6:4	25.89	20.56%	815 s

Table 3. Evaluation of model performance with different data division ratio with *lstmunits* set to 128 and *epochs* set to 20.

Division Ratio	MAE	MAPE	Training Time
9:1	27.42	19.83%	4680 s
8:2	24.64	19.66%	3915 s
7:3	27.34	21.63%	3930 s
6:4	28.75	21.35%	3391 s

Table 4. Evaluation of model performance with different data division ratio with *lstmunits* set to 128 and *epochs* set to 10.

Division Ratio	MAE	MAPE	Training Time
9:1	28.31	23.63%	2610 s
8:2	26.23	24.02%	2310 s
7:3	25.88	22.61%	2009 s
6:4	30.37	23.25%	1711 s

Table 5. Evaluation of model performance with different data division ratio with *lstmunits* set to 32 and *epochs* set to 20.

Division Ratio	MAE	MAPE	Training Time
9:1	26.62	21.01%	1713 s
8:2	23.99	18.77%	1413 s
7:3	23.91	19.63%	1298 s
6:4	26.67	20.90%	1098 s

Table 6. Evaluation of model performance with different data division ratio with *lstmunits* set to 32 and *epochs* set to 10.

Division Ratio	MAE	MAPE	Training Time
9:1	28.22	20.56%	748 s
8:2	24.19	20.35%	749 s
7:3	24.71	20.65%	690 s
6:4	26.52	22.77%	579 s

3.2. Results

Table 7 compares the results achieved on the UCI dataset for the GRU, LSTM, and Bi-LSTM neural network structures. As can be seen from Table 1, the MAE decreased from 27.28, 26.45 to 23.10 and the MAPE decreased from 21.96%, 21.27% to 18.55% using Bi-LSTM compared to LSTM and GRU, which is satisfactory. These facts show that training with Bi-LSTM is superior to GRU and LSTM.

Table 7. Evaluation of the performance of different RNN models.

| Model	MAE	MAPE
CNN, LSTM, and Attention	27.28	21.96%
CNN, GRU, and Attention	26.45	21.27%
CNN, Bi-LSTM, and Attention	23.10	18.55%

Table 8 shows the results of the ablation experiments, where this experimental method was compared to the network without the attention mechanism, the MAE decreased from 28.31 to 23.10, and the MAPE decreased from 22.23% to 18.55%. For the network without CNN, MAE decreased from 26.22 to 23.10, and MAPE decreased from 20.49 to 18.55%. In addition, the MAE and MAPE of the network without the attention mechanism were smaller than those of the network without the CNN, indicating that the attention mechanism appears to be more important than the CNN in this prediction model.

**Table 8.** Results of ablation experiments.

Model	MAE	MAPE
CNN and Bi-LSTM	28.31	22.23%
Bi-LSTM and attention	26.22	20.49%
CNN, Bi-LSTM, and Attention	23.10	18.55%

Figures 3 and 4 are examples of a normal electricity usage pattern detection, from which it is clear that the real-time threshold curve follows the true trend of the sequence, indicating that the resulting model can effectively reflect the user's electricity usage habits. In addition, the real power consumption curve in the figure does not exceed the threshold range, indicating that the user's power consumption is within the normal usage range.

Figure 5 shows the sequence of curves where anomalous consumption patterns occur, with actual values significantly greater than the threshold at the markers in the graph. As the model uses past electricity consumption behavior as a guide, the model attempts to predict electricity consumption in the same way as past predictions. When the actual value is greater than the threshold, the system detects abnormal energy consumption behavior in real-time and can record it. In addition, the method can accurately distinguish between normal and abnormal consumption behavior.
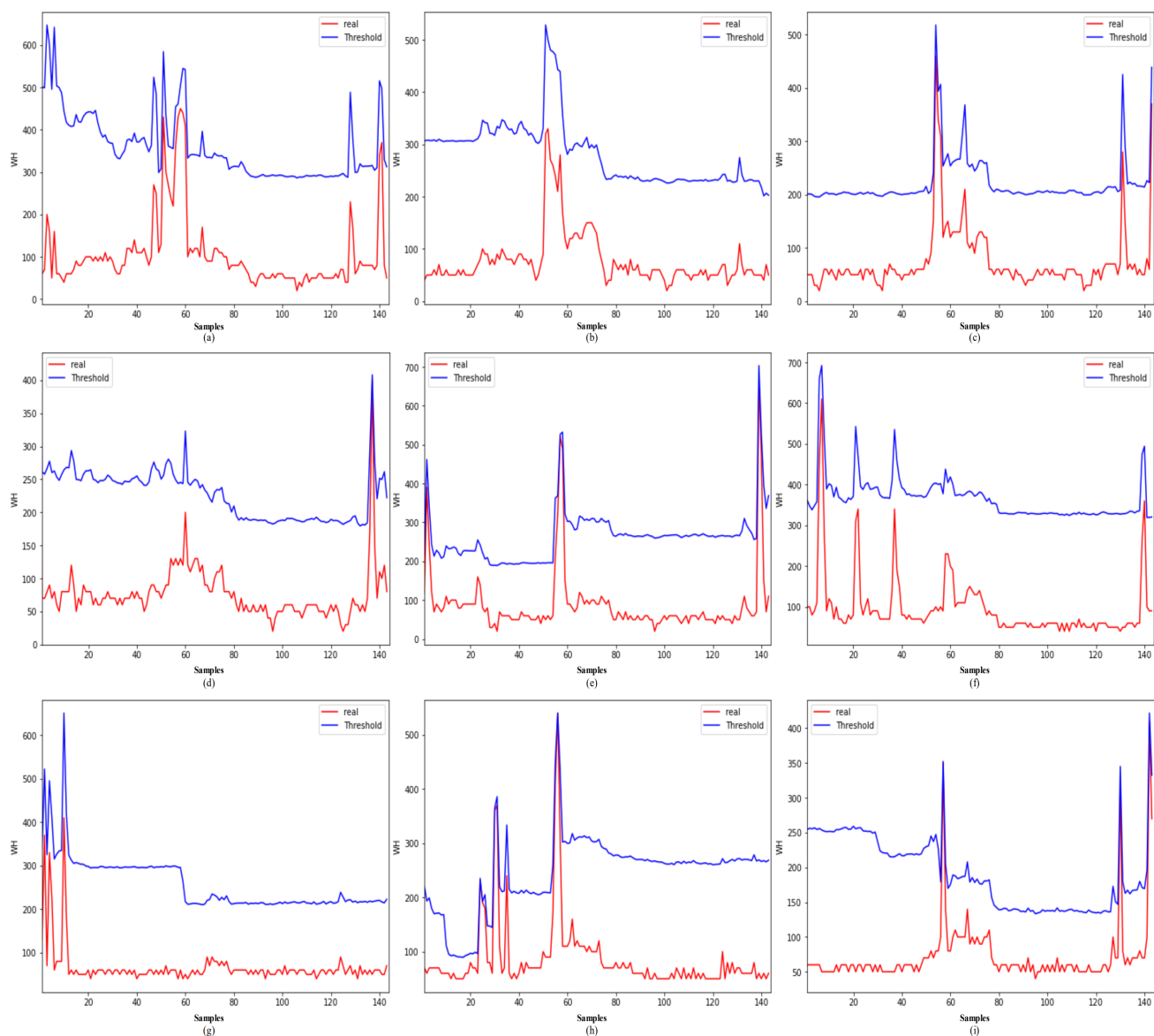
Figure 3. Example of the first group of normal energy consumption behavior: (**a**) electricity consumption on 1 May; (**b**) electricity consumption on 2 May; (**c**) electricity consumption on 3 May; (**d**) electricity consumption on 4 May; (**e**) electricity consumption on 5 May; (**f**) electricity consumption on 6 May; (**g**) electricity consumption on 8 May; (**h**) electricity consumption on 9 May; (**i**) electricity consumption on 10 May.
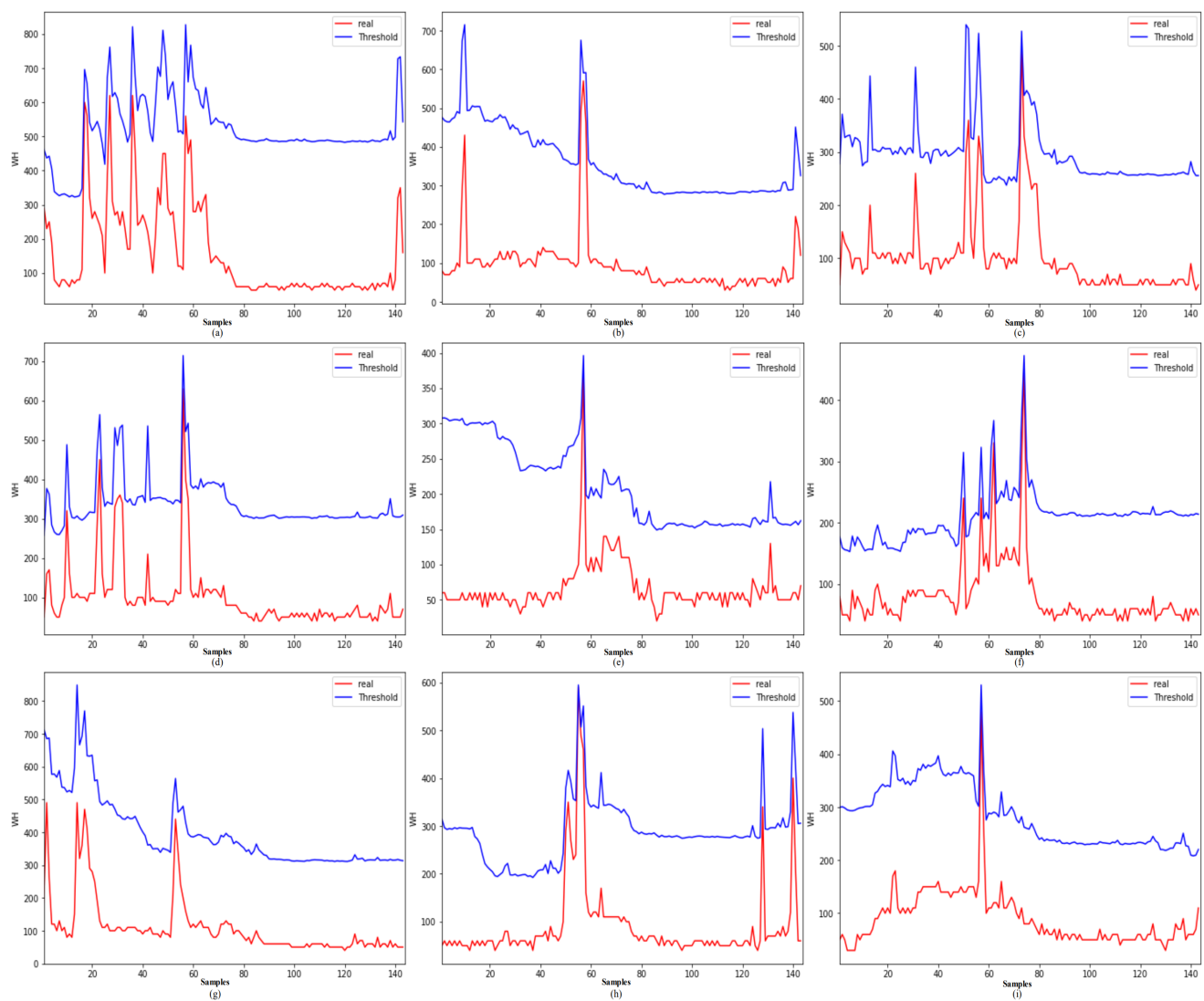
Figure 4. Example of the second group of normal energy consumption behavior: (**a**) electricity consumption on 13 May; (**b**) electricity consumption on 14 May; (**c**) electricity consumption on 15 May; (**d**) electricity consumption on 16 May; (**e**) electricity consumption on 17 May; (**f**) electricity consumption on 18 May; (**g**) electricity consumption on 22 May; (**h**) electricity consumption on 23 May; (**i**) electricity consumption on 24 May.
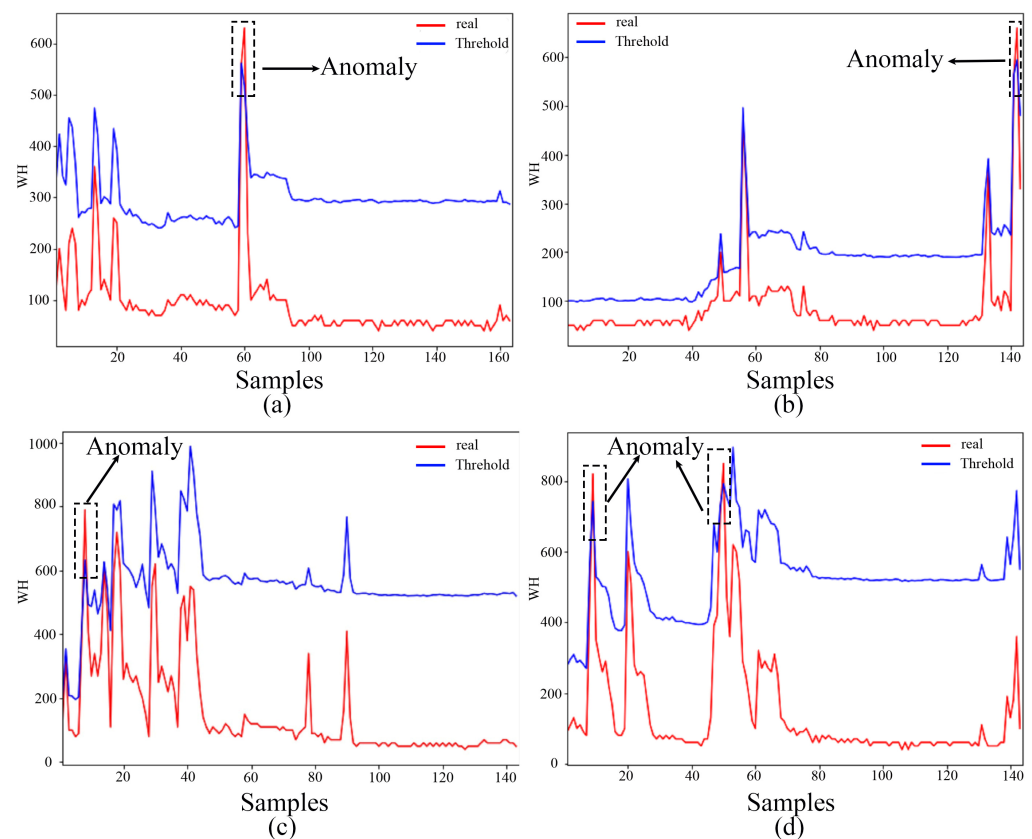
Figure 5. Example of abnormal energy consumption behavior: (**a**) electricity consumption on 7 May; (**b**) electricity consumption on 12 May; (**c**) electricity consumption on 21 May; (**d**) electricity consumption on 26 May.

4. Discussion

Compared with traditional approaches to energy consumption prediction detection using machine learning methods, our method can predict customers' electricity consumption information from high-dimensional energy consumption history data without a priori knowledge and can detect customers' electricity consumption anomalies in real-time. As for the method of energy consumption anomaly detection using LSTM networks, our method combines the features of CNN, Bi-LSTM, and an attention mechanism, and has the advantage of obtaining contextual information on time series data. It will consider the impact of different time dimensions in the input sequence on the energy consumption and make the hidden layer more focused on the key information in the sequence data.

The main objective is to use high-dimensional energy consumption data to identify abnormal electricity consumption behavior of users. However, there are corresponding limitations, such as the need to obtain the user's historical energy consumption information in advance to train the model as a way to capture the user's electricity consumption habits and to obtain a day's worth of energy consumption information first at the time of use. Secondly, the system is not well placed to point out a user's high energy consumption behavior if the user's past electricity habits are poor, and their electricity consumption is high. In addition, the practical implementation of the system depends on the availability, privacy constraints, and computing power of the data streams concerning the different user profiles.

After our neural network model is developed and deployed, the data distribution will change for various reasons, and then the model needs to be updated. We use the old model as the base model, combine the old data with the new data, and re-train the model

to update the model. The model update interval will be changed according to the actual situation, usually once every six months.

The COVID-19 pandemic and the energy crisis are anomalies that will lead to changes in customers' electricity consumption habits. Since our approach uses historical energy consumption-related data to build the model, past models will become unreliable. To deal with this anomaly, the best solution is to update the model to reduce the error.

5. Conclusions

In this study, a high-dimensional energy consumption anomaly detection method based on CNN, Bi-LSTM, and attention mechanism is proposed.

The experimental results show that the model obtained by training with historical high-dimensional energy consumption information can effectively reflect the electricity consumption behavior of users. In addition, comparisons in the ablation experiments fully illustrate that the combination of CNN, Bi-LSTM, and attention mechanisms has a better performance compared to using isolated components. In anomaly detection, the resulting model was trained to identify abnormal electricity usage behavior of users in real time. Therefore, it confirms the suitability of the model for anomaly detection.

At the same time, the research helps to establish a real-time anomaly detection system in buildings, through which building managers can plan energy consumption rationally and identify abnormal electricity usage by users. In addition, users can use the system to understand their electricity consumption and reduce energy waste.

In practical applications, information on various parameters required for model training and prediction can be obtained by reading from various sensors installed in the building. These data are transmitted through IoT devices to the cloud for calculation and storage. In addition, due to the development of IoT and communication technologies, our approach is highly applicable by simply training and deploying models in the cloud and sending the energy consumption anomaly detection results to individual building managers and users through the network. In terms of cost, the main cost is the purchase and installation of IoT devices. Since this study uses an open source system, the software licensing cost is effectively reduced.

There are two directions for future research. The first is to continue to improve model accuracy and reduce prediction errors so that the model can better analyze users' electricity consumption behavioral habits; the second is that customers' electricity consumption habits may change over time and the applicability of the model decreases, so addressing conceptual drift is a priority in future work.

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

References

1. Energy Performance of Buildings Directive. Available online: https://ec.europa.eu/energy/en/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en (accessed on 15 December 2021).
2. IEA. Tracking Buildings 2021. Available online: https://www.iea.org/reports/tracking-buildings-2021 (accessed on 30 November 2021).
3. Depuru, S.S.S.R.; Wang, L.; Devabhaktuni, V.; Gudi, N. Smart meters for power grid—Challenges, issues, advantages and status. In Proceedings of the IEEE/PES Power Systems Conference and Exposition, Phoenix, AZ, USA, 20–23 March 2011.
4. Himeur, Y.; Alsalemi, A.; Bensaali, F.; Amira, A. Robust event-based non-intrusive appliance recognition using multi-scale wavelet packet tree and ensemble bagging tree. *Appl. Energy* **2020**, *267*, 114877. [CrossRef]
5. Rashid, H.; Singh, P.; Stankovic, V.; Stankovic, L. Can non-intrusive load monitoring be used for identifying an appliance's anomalous behavior? *Appl. Energy* **2019**, *238*, 796–805. [CrossRef]
6. Rashid, H.; Singh, P. Monitor: An abnormality detection approach in buildings energy consumption. In Proceedings of the IEEE 4th International Conference on Collaboration and Internet Computing (CIC), Philadelphia, PA, USA, 18–20 October 2018.
7. Himeur, Y.; Alsalemi, A.; Bensaali, F.; Amira, A. A novel approach for detecting anomalous energy consumption based on micro-moments and deep neural networks. *Cogn. Comput.* **2020**, *12*, 1381–1401. [CrossRef]
8. Fenza, G.; Gallo, M.; Loia, V. Drift-aware methodology for anomaly detection in smart grid. *IEEE Access* **2019**, *7*, 9645–9657. [CrossRef]
9. Sial, A.; Singh, A.; Mahanti, A. Detecting anomalous energy consumption using contextual analysis of smart meter data. *Wirel. Netw.* **2021**, *27*, 4275–4292. [CrossRef]
10. Ghanbari, M.; Kinsner, W.; Ferens, K. Anomaly detection in a smart grid using wavelet transform, variance fractal dimension and an artificial neural network. In Proceedings of the 2016 IEEE Electrical Power and Energy Conference (EPEC), Ottawa, ON, Canada, 12–14 October 2016.
11. Giannoni, F.; Mancini, M.; Marinelli, F. Anomaly detection models for IoT time series data. *arXiv* **2018**, arXiv:1812.00890.
12. Zhou, Y.; Hu, W.; Min, Y.; Zheng, L.; Liu, B.; Yu, R.; Dong, Y. A semi-supervised anomaly detection method for wind farm power data preprocessing. In Proceedings of the 2017 IEEE Power & Energy Society General Meeting, Chicago, IL, USA, 16–20 July 2017.
13. Huang, T.; Sethu, H.; Kandasamy, N. A new approach to dimensionality reduction for anomaly detection in data traffic. *IEEE Trans. Netw. Serv. Manag.* **2016**, *13*, 651–665.
14. Kudo, T.; Morita, T.; Matsuda, T.; Takine, T. PCA-based robust anomaly detection using periodic traffic behavior. In Proceedings of the 2013 IEEE International Conference on Communications Workshops (ICC), Budapest, Hungary, 9–13 June 2013.
15. Dai, X.; Gao, Z. From model, signal to knowledge: A data-driven perspective of fault detection and diagnosis. *IEEE Trans. Ind. Inform.* **2013**, *9*, 2226–2238. [CrossRef]
16. Pereira, J.; Silveira, M. Unsupervised anomaly detection in energy time series data using variational recurrent autoencoders with attention. In Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018.
17. Munir, M.; Siddiqui, S.A.; Chattha, M.A.; Dengel, A.; Ahmed, S. Fusead: Unsupervised anomaly detection in streaming sensors data by fusing statistical and deep learning models. *Sensors* **2019**, *19*, 2451. [CrossRef]
18. Zheng, Z.; Yang, Y.; Niu, X.; Dai, H.N.; Zhou, Y. Wide and deep convolutional neural networks for electricity-theft detection to secure smart grids. *IEEE Trans. Ind. Inform.* **2017**, *14*, 1606–1615.
19. Li, S.; Han, Y.; Yao, X.; Yingchen, S.; Wang, J.; Zhao, Q. Electricity theft detection in power grids with deep learning and random forests. *J. Electr. Comput. Eng.* **2019**, *2019*, 4136874. [CrossRef]
20. Hollingsworth, K.; Rouse, K.; Cho, J.; Harris, A.; Sartipi, M.; Sozer, S.; Enevoldson, B. Energy anomaly detection with forecasting and deep learning. In Proceedings of the IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018. [CrossRef]
21. Wang, X.; Zhao, T.; Liu, H.; He, R. Power consumption predicting and anomaly detection based on long short-term memory neural network. In Proceedings of the IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, 12–15 April 2019. [CrossRef]
22. Xu, C.; Chen, H. Abnormal energy consumption detection for GSHP system based on ensemble deep learning and statistical modeling method. *Int. J. Refrig.* **2020**, *114*, 106–117.
23. Ergen, T.; Kozat, S.S. Unsupervised anomaly detection with LSTM neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2019**, *31*, 3127–3141.
24. Keras. Available online: https://keras.io (accessed on 19 November 2015). [CrossRef]
25. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Available online: https://www.tensorflow.org/ (accessed on 19 November 2015). [CrossRef] [PubMed]
26. Kamal, M.B.; Mendis, G.J.; Wei, J. Intelligent soft computing-based security control for energy management architecture of hybrid emergency power system for more-electric aircrafts. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 806–816.
27. Li, H.; Min, M.R.; Ge, Y.; Kadav, A. A context-aware attention network for interactive question answering. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017.
28. Brown, M.; Barrington-Leigh, C.; Brown, Z. Kernel regression for real-time building energy analysis. *J. Build. Perform. Simul.* **2012**, *5*, 263–276. [CrossRef]