IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.Doi Number

Predicting Residential Electricity Consumption Using CNN-BiLSTM-SA Neural Networks

MENG-PING WU [1], Fan WU [1]

[1]School of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

Corresponding author: Mengping WU. Email: xomxp8@163.com

ABSTRACT As global population growth and the use of household appliances increase, residential electricity consumption has surged, leading to challenges in maintaining a balanced electrical load. This surge often results in localized and intermittent power outages, adversely affecting residential electricity reliability and the profitability of power supply companies. Addressing this, we propose a novel CNN-BiLSTM-SA model, combining Convolutional Neural Network (CNN), Bidirectional Long Short-Term Memory (BiLSTM), and Self-Attention (SA) mechanisms, to accurately predict residential electricity consumption. The model integrates temporal feature extraction through the CNN module, correlation capture via BiLSTM, and intrinsic feature analysis using SA mechanisms, enhancing predictive accuracy. Our experimental results demonstrate the model's precision, outperforming existing methods in key performance metrics. Additionally, ablation studies affirm the synergistic design of CNN-BiLSTM-SA's network components, contributing to its overall efficacy.

INDEX TERMS electricity consumption; prediction; CNN; BiLSTM; self attention

I. INTRODUCTION

The contemporary world faces a dual challenge in the energy sector: the growing scarcity of global energy resources and the critical need for stable power supply [1]. This situation is exacerbated by increasing industrialization and urbanization, leading to escalated energy demands [2]. Ensuring a stable and sustainable power supply, amidst these escalating demands and dwindling resources, has emerged as a paramount concern for both energy policy makers and consumers. The global energy crisis underscores the necessity for efficient energy management and consumption strategies, highlighting the urgency of developing innovative approaches to address these challenges.

Household electricity consumption forms a significant portion of the global electricity usage, with residential sectors accounting for a substantial share of total energy consumption [3]. This consumption pattern is not just a reflection of the growing number of household appliances but also of lifestyle changes and economic development. As households continue to be major consumers of electricity, understanding and managing their energy consumption patterns becomes a pivotal aspect of global energy management.

Analyzing and predicting household electricity consumption is vital for several reasons. First, it enables utility companies to optimize power generation and distribution, ensuring a stable power supply. Second, it aids in the identification of user-specific consumption patterns, allowing for personalized energy-saving strategies. Moreover, accurate predictions of electricity demand can assist in balancing supply and demand, thereby reducing the risk of power outages and enhancing the overall efficiency of the energy sector. Thus, the study of household electricity consumption patterns is not only a matter of economic interest but also of strategic importance in maintaining energy stability and sustainability.

Several seminal studies have laid the foundation for the analysis of household electricity consumption. For instance, Pappas [4] explored the application of traditional time-series models in forecasting electricity demand, highlighting the importance of seasonal and daily patterns in consumption. Another notable work by Andoh [5] employed machine learning techniques to segment users based on consumption patterns, providing insights into energy usage behaviors. These studies exemplify the diverse methodologies employed in understanding and forecasting household electricity consumption, each contributing uniquely to the field.

The long-term characteristics of household electricity consumption, such as seasonality, trend, and user-specific patterns, necessitate models capable of capturing both spatial and temporal dependencies. CNNs excel in identifying spatial hierarchies in data, making them suitable for extracting features from consumption patterns. BiLSTMs are adept at capturing long-term temporal dependencies, essential for understanding consumption trends over time. SA mechanism further enhances the model's ability to focus on relevant parts of the data, providing a more nuanced understanding of electricity usage. The synergy of these techniques makes the CNN-BiLSTM-SA model particularly apt for tackling the intricacies of household electricity consumption prediction.

This paper introduces the CNN-BiLSTM-SA model, a novel framework designed to effectively predict household electricity consumption. The model harnesses the strengths of CNN for feature extraction, BiLSTM for capturing temporal relationships, and SA for emphasizing crucial aspects of the consumption data. This integrated approach enables a comprehensive analysis of electricity usage patterns, offering more accurate and reliable predictions. The contribution of this paper lies not only in the development of the CNN-BiLSTM-SA model but also in its application to household electricity consumption, providing valuable insights for energy management and policy-making. This work, therefore, stands as a significant advancement in the field of energy consumption analysis, paving the way for more sustainable and efficient energy utilization strategies. The main contributions of this paper are as follows:

- We present a novel hybrid deep learning framework aimed at accurately forecasting electricity consumption within an actual residential house.
- We demonstrate the best performance across all prediction metrics compared to previous studies.
- We conduct an analysis of the variables pertaining to household appliances that exert influence on the prediction of energy consumption.

The rest of this paper is organized as follows. In section 2, the related work on residential electricity consumption prediction is discussed. The prediction of residential electricity consumption (PREC) is described in section 3. Section 4 details the proposed CNN-BiLSTM-SA neural networks architecture. Section 5 presents experimental results and section 6 concludes this study.

II. RELATED WORK

The endeavor to predict household electricity consumption has rapidly evolved into a vibrant research area, spurred by global sustainability goals and technological advancements. The methodologies employed in this realm predominantly align with three distinct categories: traditional statistical methods, machine learning models, and hybrid approaches. Each category, while distinct in its approach, converges on the common goal of enhancing the accuracy and efficiency of electricity consumption predictions. This multiplicity of approaches signifies both the complexity of the problem and the interdisciplinary nature of the solutions being developed. Table 1 shows the studies related to power consumption prediction according to three categories.

Traditional statistical methods serve as the foundation of early research endeavors in this field. These methods are lauded for their simplicity and interpretability, which is crucial for policy formulation and strategic planning. Key studies in this domain include seasonal prediction model with periodical fluctuation [6], which predicts monthly power consumption in a family, showcasing better precision compared to other models. Building on this work, the seminal study offers an intelligent data-analysis method for modeling and predicting daily electricity consumption in buildings [7]. Another line of research uses a Genetic Algorithm-Based Numerical Moment Matching method for predicting electricity [8]. In contrast, Zolfaghari [9] combines the adaptive wavelet neural network with the ARIMA-GARCH family models for forecasting. A different perspective in this domain is provided by Arghira [10], which focuses on predicting the next day electricity consumption for various services in homes. Similarly, Amber [11] compares various techniques for forecasting electricity consumption in an administration building. Moreover, Geetha [12] proposed a random forest supervised learning model for forecasting power consumption. Additionally, Syah [13] introduces a test method for selecting the correct data set for electricity consumption prediction. Furthermore, Farzana [14] compares different demand models to improve estimation efficiency for future energy consumption projections. Akyol [15] defines and systematically analyzes the Persistence Forecast Effect in forecasting domestic electricity consumption. Spiliotis [16] proposed a statistical and machine learning method to predict week-ahead hourly electricity prices, finding that machine learning methods provided better forecasts in terms of accuracy and bias. They also noted the value-added of explanatory variables in extrapolation tasks for both statistical and machine learning approaches. Following this trend, Kim [17] compared linear regression (a traditional statistical method) and artificial neural network algorithms for predicting electricity consumption in a building, finding that the artificial neural network model was more accurate and stable for working days. Yang [18] adopted the support vector machine method to analyze statistical data influencing electricity consumption and developed a corresponding forecasting model, indicating its accuracy. In addition, Tso [19] compared regression analysis, decision tree, and neural networks for electricity consumption prediction, suggesting that decision tree and neural network models could be viable alternatives to traditional regression analysis.

The integration of machine learning models marked a paradigm shift in predicting household electricity consumption. These models are adept at deciphering complex, non-linear relationships within data. One of the pioneering works in this category is automated machine learning for short-term electric load forecasting by Wang [20], which uses automated machine learning systems for load prediction. Extending this idea, Rahman [21] analyzes household electricity consumption data using various models including ARIMA and LSTM. Complementing this work, Yan [22] proposes a hybrid deep learning neural network framework combining CNN and LSTM. Moreover, a influential study of Aurangzeb [23], evaluates eight regression models for predicting power consumption of single households. Furthermore, Guo [24] utilizes a deep feedforward network for short-term electricity load forecasting. Alhussein [25] combines convolutional neural network and long short-term memory for forecasting. Another aspect of this category is the analysis and prediction of energy use in smart homes using machine learning models. One of the pioneering works in this category is by Xiong [26], which analyzes and predicts energy use in smart homes using machine learning models. Dlamini [27] proposed a deep quantile regression model and a gradient boosting model for power consumption forecasting. The final stream of research in this category is the electrical energy consumption prediction using machine learning. One of the pioneering works in this category is by Hyeon [28], utilized deep learning models, including vanilla LSTM and sequence to sequence with attention mechanism, for predicting electric energy consumption in households. They observed that the vanilla LSTM showed the best performance based on the root-mean-square error metric. Following this direction, Irankhah [29] proposed a new hybrid network consisting of Auto-Encoder LSTM layer, Bi-LSTM layer, and Fully connected layer for short-term load forecasting. They demonstrated that the proposed network achieved the smallest value in terms of RMSE, MAE, and MAPE compared with other approaches. In another influential study by Hajjaji [30], different algorithms for daily power consumption prediction in a university campus context are evaluated using time series, machine learning, and deep learning approaches. They found that deep learning approaches, especially the RNN-LSTM hybrid model, achieved better results than time series and machine learning forecasting models.

Hybrid approaches, blending the strengths of statistical and machine learning models, represent the cutting edge of research in this area. These methods aim to synergize the interpretability of statistical models with the predictive power of machine learning algorithms. Alkhatib [31] develops a hybrid model using conditional inference trees and linear regression for forecasting. Velasco [32] uses a hybrid forecasting model using ARIMA and ANN for a power utility's dataset to predict the next day's electric load consumption. This hybrid model delivered a Mean Absolute Percentage Error (MAPE) of 4.09%, demonstrating the synergy of statistical and machine learning approaches. Additionally, Saxena [33] proposed a hybrid model based on ARIMA, logistic regression, and artificial neural networks to forecast peak load days for a billing period, achieving significant savings for an American university during a one-year testing period. Similarly, Kesornsit [34] presented a hybrid model integrating dimensionality reduction and feature selection algorithms with a backpropagation neural network to predict electricity consumption in Thailand. This model consistently outperformed others in their study. Irankhah [35] proposed a new hybrid network consisting of Auto-Encoder LSTM layer, Bi-LSTM layer, and Fully connected layer for short-term load forecasting. They demonstrated that the proposed network achieved the smallest value in terms of RMSE, MAE, and MAPE compared with other approaches. Following this direction, Hajjaji [36] evaluated different algorithms for daily power consumption prediction in a university campus context using time series, machine learning, and deep learning approaches. They found that deep learning approaches, especially the RNN-LSTM hybrid model, achieved better results than time series and machine learning forecasting models. These hybrid models underscore the progressive nature of the field, striving for models that are not only accurate but also interpretable and robust.

In synthesizing the literature, it is evident that considerable progress has been made in the realm of household electricity consumption prediction. From the foundational statistical methods to the advanced hybrid models, each category of

Table 1.Related works on electricity consumption

Category	Researcher	Method	Description
Traditional statistical method	Xiuhua C, 2007, [6]	Seasonal prediction model	Analysis and prediction of monthly power consumption in a family
	Zolfaghari M, 2019, [9]	ARIMA-GARCH	Prediction of household electricity consumption for various services
	Kim M. K, 2020, [17]	Linear regression	Predicting electricity consumption in a building
Traditional machine learning method	Yang S, 2006, [18]	Support vector machine	Analyze statistical data influencing electricity consumption and developed a corresponding forecasting model
	Tso G. K. F, 2007, [19]	Neural network	Compared regression analysis, decision tree, and neural networks for electricity consumption prediction
	Guo Z, 2018, [24]	Feedforward network	Short-term electricity load forecasting
Deep learning method	Hyeon J, 2020, [28]	Long short-term memory	predicting electric energy consumption in households
	Hajjaji I, 2021, [30]	RNN-LSTM	Comparing performance of time series, machine learning, and deep learning approaches on power consumption prediction
	Velasco L, 2019, [32]	ARIMA-CNN	Using a hybrid forecasting model for electric load consumption prediction
	Irankhah A, 2021, [35]	BiLSTM	A hybrid network consisting of Auto-Encoder LSTM layer, Bi-LSTM layer, and Fully connected layer for short-term load forecasting

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3400972

research has contributed to a deeper understanding and improved forecasting capabilities. Nevertheless, challenges persist, particularly in handling the vast heterogeneity of data and the dynamic nature of consumption patterns. Addressing these challenges, this paper introduces the novel CNN-BiLSTM-SA framework, which amalgamates the feature extraction capabilities of CNNs, the temporal learning strengths of BiLSTMs, and the contextual focus of the SA mechanism. This model is designed to transcend the limitations of previous approaches, offering a comprehensive solution that is both robust and adaptable. In doing so, this research contributes a significant leap forward in the pursuit of more efficient and sustainable energy management practices.

III. PROBLEM DESCRIPTION

The goal of the PREC is to predict current and possible trends in the consumption of electricity, based on the analysis, modeling, and prediction of current data characteristics. Formally, the input of single data array i at certain statistical time in PREC is represented by two parts: the fully available feature vector $x_i = (f_1, f_2, \ldots, f_i \ldots, f_n), n \in N^*$, where n is the number of selected features, $f_i, i \in [1, n], i \in N^*$ denotes one of the n factors; the corresponding electricity consumption y_i. Furthermore, the input of the data within all data collection time frames are represented as $\{x_t\}_1^T = \{x_1, x_2, \ldots, x_T\}$, and $\{y_t\}_1^T = \{y_1, y_2, \ldots, y_T\}$, where T is the total amount of data collection time steps. The PREC problem can be defined as how to establish an effective model F that can analyze the time series of $\{x_t\}_1^T$ and $\{y_t\}_1^T$, so as to predict the future electricity consumption of residential users. That is, $\{y_t'\}_T^{T+\Delta} = F(\{x_t\}_1^T, \{y_t\}_1^T)$.

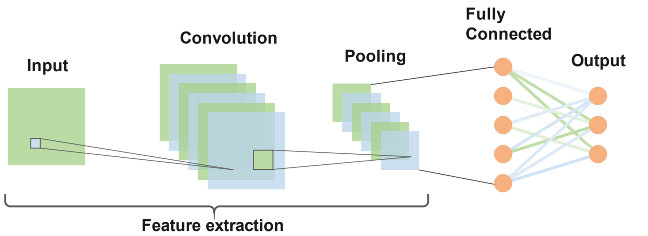
IV. METHODOLOGY

In this section, the CNN, BiLSTM, and SA used in the algorithm in this article are first briefly introduced, followed by a detailed description of the proposed CNN BILSTM-SA framework and its principles. In this section, CNN, BiLSTM, and SA related to the methodology proposed in this article are first briefly introduced, followed by a detailed description of the proposed CNN BILSTM-SA framework and its principles.

A. CNN

CNN is a feedforward neural network that includes convolution and deep structure, widely used in image classification, object detection, face recognition and other fields, and can also be combined with LSTM or other neural networks for solving prediction problems. The basic principle of CNN is to first perform convolution operations on the input data, extract features, and compress and output the number of effective parameters. Subsequently, stepwise sliding calculations are performed using convolutional kernels to further extract local features. By rolling calculations, complete features that balance both global and local aspects are ultimately obtained.

FIGURE 1. The structure of Convolutional Neural Network



The structure of CNN is shown in Fig. 1. It consists of an input layer, multiple hidden layers, and an output layer. The hidden layers typically include convolutional layers, pooling layers, fully connected layers, and sometimes normalization layers. Convolutional layers apply filters to the input, capturing local dependencies and extracting features like edges and textures. Pooling layers reduce the spatial dimensions of the data, lowering the computational load and the model's sensitivity to the location of features in the input. Fully connected layers integrate these features, which are then used for tasks such as classification or regression in the output layer. The activation functions, especially ReLU, introduce non-linearity, enabling the CNN to learn complex patterns. CNNs are renowned for their effectiveness in image and video recognition, image classification, and other visual data-driven tasks.

Convolutional Layer: The convolutional layer extracts features from the input data through the convolution operation. The formula for this operation is:

$$F_{ij}^{l} = \sigma\left(\sum_{m}\sum_{n} W_{mn}^{l} \cdot X_{(i+m)(j+n)} + b^{l}\right) \tag{1}$$

Where F_{ij}^{l} is the feature map at position (i, j) in layer l, W_{mn}^{l} is the convolution kernel, $X_{(i+m)(j+n)}$ is the input data, b^{l} is the bias term, and σ is the activation function.

Pooling Layer: The pooling layer reduces the dimensionality of the feature maps and commonly includes max pooling and average pooling. The formula for this operation is:

$$P_{ij} = max(X_{kl}) \tag{2}$$

where P_{ij} is the output after pooling, and X_{kl} are the input values within the pooling window.

Fully Connected Layer: The fully connected layer transforms the output of the previous layer into class scores. The formula for this operation is:

$$O_i = \sigma\left(\sum_{j} W_{ij} \cdot F_j + b_i\right) \tag{3}$$

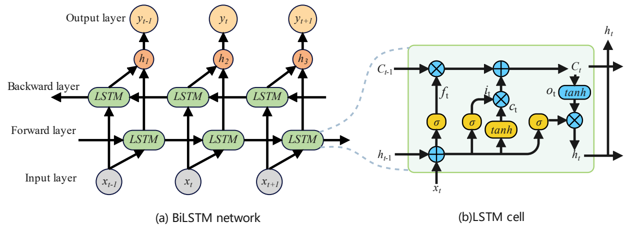
where O_i is the output, W_{ij} are the weights and F_j is the flattened feature map.

B. BI LSTM

LSTM is an improved recurrent neural network used to compensate for the vanishing or exploding gradient problem of RNN. The basic principle of LSTM is to filter invalid information and add valid information through two structures, gate and cell, to achieve real-time updating of storage units, thereby ensuring reasonable gradients. Specifically, the input gate controls the inflow of new information, the forget gate regulates the forgetting of old information, and the output gate determines the output of information. The cell state is updated through the control of the forget gate and the input gate, thereby retaining or discarding information. Compared with RNN, although LSTM can handle longer time series data, its training efficiency and effect are significantly reduced when faced with high sampling frequency and long-period time series problems represented by residential electricity consumption analysis. In addition, due to the existence of gradients, the cumulative loss of LSTM will be too large as the sequence length increases. BiLSTM is a network more suitable for long-term prediction.

BiLSTM, an advancement of traditional LSTMs, processes sequential data bidirectionally, comprehensively understanding temporal dynamics and enabling effective handling of complex dynamics in high-frequency time series data. Comprising two LSTM layers working oppositely, BiLSTM mitigates the risk of losing information in long sequences through bidirectional information processing. Additionally, BiLSTM captures past and future information, enhancing performance in tasks like text classification and sentiment analysis. This dual-direction processing makes BiLSTMs ideal for complex sequence learning, particularly in natural language processing, by addressing limitations like the vanishing gradient problem inherent in standard RNNs. Its structure is shown in Fig. 2.

FIGURE 2. The structure of BiLSTM Network



A BiLSTM processes an input sequence $x = (x_1, x_2, \ldots, x_T)$ with both a forward LSTM and a backward LSTM.The forward LSTM processes the input sequence from left to right. The update rules for its hidden state $h_t^{(f)}$ at each time step t are as follows:

Forget Gate:

$$f_t^{(f)} = \sigma\left(W_f^{(f)} \cdot \left[h_{t-1}^{(f)}, x_t\right] + b_f^{(f)}\right) \tag{4}$$

Input Gate:

$$i_t^{(f)} = \sigma\left(W_i^{(f)} \cdot \left[h_{t-1}^{(f)}, x_t\right] + b_i^{(f)}\right) \tag{5}$$

Candidate Cell State:

$$\tilde{C}_t^{(f)} = tanh\left(W_C^{(f)} \cdot \left[h_{t-1}^{(f)}, x_t\right] + b_C^{(f)}\right) \tag{6}$$

Update Cell State:

$$C_t^{(f)} = f_t^{(f)} * C_{t-1}^{(f)} + i_t^{(f)} * \tilde{C}_t^{(f)} \tag{7}$$

Output Gate:

$$o_t^{(f)} = \sigma\left(W_o^{(f)} \left\{\left[h_{t-1}^{(f)}, x_t\right] + b_o^{(f)}\right)\tag{8}$$

Update Hidden State:

$$h_t^{(f)} = o_t^{(f)} * tanh\left(C_t^{(f)}\right) \tag{9}$$

Here, σ is the sigmoid function, $*$ denotes element-wise multiplication, and $\left[h_{t-1}^{(f)}, x_t\right]$ denotes the concatenation of $h_{t-1}^{(f)}$ and x_t .The backward LSTM processes the input sequence from right to left. Its update rules are similar to those of the forward LSTM, but with $h_{t-1}^{(f)}$ replaced by $h_{t+1}^{(b)}$.Finally, the hidden states of the forward and backward LSTMs are averaged to produce the final output at each time step. This way, the output at each time step incorporates information from both the past and the future.The formula for this is given by:

$$h_t = \frac{1}{2}\left(h_t^{(f)} + h_t^{(b)}\right) \tag{10}$$

C. SELF-ATTENTION

The self-attention mechanism, a key component in modern neural network architectures like Transformers, allows each part of a sequence to consider and weigh the importance of other parts in the same sequence. It calculates attention scores to represent how much focus to put on other parts of the input for each element in the sequence. This mechanism enhances the model's ability to capture context, identify relationships, and understand dependencies in data, especially in tasks involving sequential information like language modeling and machine translation. Self-attention's flexibility and effectiveness in handling long-range dependencies make it a powerful tool in deep learning, offering significant improvements over traditional sequence processing methods. Let Q, K, and V represent query, key, and value respectively. These are matrices obtained by linear transformation of the input sequence. The correlation score is calculated as:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{11}$$

where Q, K, V are matrices representing Queries, Keys, and Values, respectively. These are derived through linear transformations of the input. QK^T is the dot product of the Query matrix Q and the transpose of the Key matrix K. This dot product measures the alignment or similarity between queries and keys. $\sqrt{d_k}$ is a scaling factor where d_k is the dimension of the key vectors. This scaling helps in stabilizing the gradients during training. The softmax function is applied across the rows, turning the scores into probabilities that sum to 1. This function determines the weightage given to each value. The output is a weighted sum of the Value vectors V, where the weights are the attention probabilities. This formula enables the model to focus on different parts of the input sequence, capturing contextual relationships within the data.

D. CNN-BILSTM-SA

In view of the characteristics of the PREC problem, this study proposes a long-term prediction framework that integrates CNN, BILSTM and SA, namely CNN-BILSTM-SA. As shown in Fig. 3, CNN-BILSTM-SA consists of five layers, namely input layer, CNN, BILSTM, SA and output layer. The functions of each sub-layer are as follows：

Input Layer. Receive electricity consumption data and its associated feature vectors, and normalize the data.

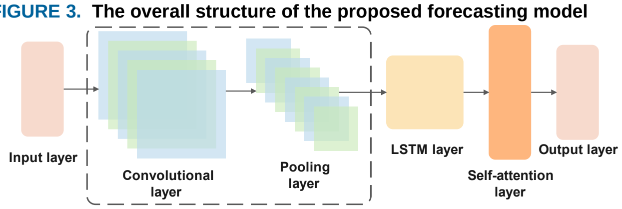
CNN Layer. Connect the input layer and BiLSTM layer. Receive data from the input layer, extract and retain effective mapping features between electricity consumption data and its related feature vectors, through the cooperation of convolutional layers and pooling layers.

BiLSTM Layer. Connect the BiLSTM layer and SA layer. Receive data from CNN and further process and capture more detailed timing features through forward and reverse bidirectional operations.

SA Layer. Connect the SA layer and output layer. Receive data processed by BiLSTM and capture the intrinsic correlation of feature information to help improve the accuracy of the network.

Output Layer. Receive the data processed by CNN-BiLSTM-SA and output the prediction results of residents' future electricity consumption.

FIGURE 3. The overall structure of the proposed forecasting model



To optimize the architecture for forecasting energy consumption, a comprehensive understanding of the input data characteristics is essential. The energy usage data, presented as a multifaceted time series, is preprocessed into 60-minute segments using the sliding window technique to preserve temporal dynamics effectively. Employing a 2x1 kernel aids in preserving temporal intricacies. The input data for the CNN-BiLSTM-SA model is structured as a 60x10 matrix, representing 10 variables over a 60-minute period. This data undergoes convolutional and pooling operations prior to BiLSTM processing. The CNN-BiLSTM-SA architecture parameters, detailed in Table 2, encompass filter counts, convolutional layer dimensions and strides, pooling kernel specifications, and the overall parameter count, encompassing the BiLSTM layer. Additionally, the attention head count is 2, and the hidden layer size is the same as the number of BiLSTM units, the learning rate for both BiLSTM and the optimizer is set to 0.005, the dropout rate is 0.3, and the number of iterations is configured to 100 epochs.

TABLE 2: THE PROPOSED CNN-BiLSTM-SA ARCHITECTURE

Type	Filter	Kernel size	Stride	param
Convolution	64	(2,1)	1	192
Acticvation (softmax)	-	-	-	0
Pooling	-	(2,1)	2	0
Convolution	64	(2,1)	1	8256
Activation (softmax)	-	-	-	0
Pooling	-	(2,1)	2	0
TimeDistributed	-	-	-	0
BiLSTM (64)	-	-	-	180480
Activation	-	-	-	0
Dense (32)	-	-	-	2080
Dense (64)	-	-	-	1980
Total number of parameter				191988

V. EXPERIMENTS

In this section, we first introduce the experimental settings, including dataset description, feature selection, normalization, evaluation metrics, and parameter settings of CNN-BiLSTM-SA. Subsequently, we conducted three sets of experiments, namely model feasibility verification experiment, prediction results comparison experiment and indicator comparison experiment. We use python as the programming language and run it on a PC configured with Intel i7-9700F, 16Gb running memory, and NVIDIA GTX2080Ti.

A. DATA DESCRIPTION

In this paper, the publicly available household electricity consumption dataset provided by the UCI Machine Learning Laboratory was selected as the experimental data source. The dataset can be accessed at https://archive.ics.uci.edu/datasets.

In this dataset, the continuous electricity consumption of a household in Paris, France, over 47 consecutive months was collected, with a sampling frequency of minute-level and a total of 2,075,259 records. Each record includes three groups of 12 parameters sampled during the sampling period. The first group consists of five time variables, namely minute, hour, day, month, and year. The second group comprises four instantaneous electrical parameters, namely global minute-averaged active power (GMAP), global minute-averaged reactive power (GMRP), minute-averaged voltage (MV), and minute-averaged global intensity (MGI). The third group consists of three parameters for the electricity consumption of three functional zones: sub_metering 1-3. Specifically, sub_metering 1 represents the electricity consumption of kitchen appliances including microwave ovens, induction cookers, and dishwashers. Sub_metering 2 represents the electricity consumption of laundry appliances including washing machines, tumble-dryers, and refrigerators. Sub_metering 3 represents the electricity consumption of electric water heaters and air conditioners.

From the standpoint of data volume, parameter richness, and sampling frequency, this dataset proves to be conducive to the training and validation of predictive models. Additionally, empirical studies conducted by numerous researchers have affirmed the suitability of this dataset as an exemplary source for the analysis and forecasting of household electricity consumption patterns.

In this paper, the dataset is partitioned into three subsets: the training set, validation set, and test set. To facilitate clearer presentation and comparison, the proportions of the three subsets in each set of experiments are divided into 7:1:2 and

8:1:1, respectively. Building upon this framework, three sets of experiments are devised to validate the predictive accuracy of the proposed CNN-BiLSTM-SA network in describing the ground truth, compare its performance with other mainstream forecasting algorithms, and assess the contributions of key modules within the network to its overall performance.

B. FEATURE SELECTION

According to the information collected in the electricity data set, the characteristics that affect power consumption are mainly divided into three categories: one is the time window, the other is the time-sharing power consumption of each device, and the household voltage. In terms of time windows, the minute-level sampling data in the data set are preprocessed and expanded to minutes, hours, days and months; in terms of equipment power, according to functions, it is divided into laundry electricity, lighting electricity, heating electricity and kitchen electricity. Electricity; in terms of household voltage, the average voltage within the time window is used as characteristic information. Table 1 lists the characteristic information that contributes to electricity consumption and may affect electricity consumption.

To clarify the correlation between various features and electricity consumption, we employed the Maximal Information Coefficient (MIC) method for calculation. Additionally, to mitigate potential biases inherent in the MIC method, we concurrently utilized the Spearman's rank correlation coefficient and Kendall's rank correlation coefficient for our analysis. The results of the correlation analysis are presented in Table 3. The data from the table indicate that the results obtained from all three methods consistently show a positive correlation between electricity consumption and three features: heater, kitchen, and laundry, with the strength of the correlation decreasing in that order. Conversely, voltage shows a negative correlation with electricity consumption. We are using the results from the MIC method as the basis for our feature selection.

TABLE 3: CORRELATION ANALYSIS OF ELECTRICITY CONSUMPTION WITH FEATURES

	Voltage	Kitchen	Laundry	Heater
MIC	-0.325	0.229	0.183	0.624
Spearman	-0.729	0.316	0.028	0.572
Kendall	-0.551	0.219	0.067	0.301

The correlation analysis reveals associations between electricity consumption and various features. Voltage exhibits a strong negative correlation with electricity consumption, suggesting that consumption tends to increase as voltage decreases. This may indicate greater power demand to maintain normal operations during voltage drops. Consequently, special attention is warranted in modeling to understand the voltage's impact on consumption. Conversely, kitchen and heater usage demonstrates robust positive correlations with electricity consumption, indicating increased consumption during appliance use or heater activation. Therefore, prioritizing these positively correlated features in modeling is crucial, as they likely influence consumption patterns significantly and may serve as key predictors for forecasting and managing electricity consumption. In summary, correlation analysis provides insights into the relationships between electricity consumption and diverse features, guiding subsequent modeling efforts and optimizing predictions and management strategies for electricity consumption.

C. NORMALIZATION

The maximum-minimum method is used to normalize data of different dimensions and orders of magnitude such as time, voltage, power consumption, etc. The specific calculation formula of MM is as follows：

$$x' = \frac{x - min}{max - min} \tag{12}$$

where x is the original value of a single data in the electricity consumption data, min and max are respectively the minimum value and maximum value in the column to which the data belongs, and x' is the normalized value of the data.

D. EVALUATION METRICS

In order to evaluate the prediction accuracy of the model, RMSE (Root Mean Square Error), MAE (Mean Absolute Error) and R^2 (Coefficient of Determination) are used to measure how well the models explain the data. Specifically, RMSE calculates the square root of the mean of squared differences between predicted and true values, while MAE computes the mean of absolute differences between predicted and true values. R^2 indicates the proportion of the variance in the dependent variable explained by the model. The comprehensive utilization of these three metrics enables a more thorough evaluation of model fitting.

The specific formulas are as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - y_t')^2} \tag{13}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|y_t - y_t'| \tag{14}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(y_t - y_t')^2}{\sum_{t=1}^{n}\left(y_t' - \overline{y_t'}\right)^2} \tag{15}$$

Where y_t and y_t' are the collected real power consumption data and predicted power consumption data respectively; n is the amount of electricity consumption data in the electricity data set, $\overline{y_t'}$ denotes the average value of the predicted data. Smaller MAE and RMSE values indicate more accurate predictions, while a larger R2 value signifies predictions that are closer to the actual data.

Furthermore, to distinctly analyze the comparative effectiveness of CNN-BiLSTM-SA and other models, the enhancement percentages of RMSE and MAE are employed:

$$EP_{RMSE} = \frac{RMSE_C - RMSE_P}{RMSE_C} \times 100\% \tag{16}$$

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3400972

$$EP_{MAE} = \frac{MAE_C - MAE_P}{MAE_C} \times 100\% \qquad (17)$$

where EP_{RMSE} and EP_{MAE} represent the enhancement percentages for RMSE and MAE, $RMSE_P$ and MAE_P denote the evaluation metrics for CNN-BiLSTM-SA, and $RMSE_C$ and MAE_C represent the evaluation metrics for other models.

E. EXPERIMENTS

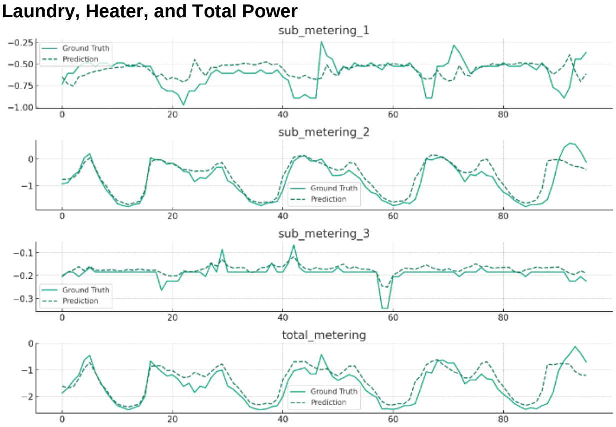
1) MODEL VALIDATION

To ascertain the efficacy of the CNN-BiLSTM-SA model, we conducted a comparative analysis of the predicted and actual values for kitchen, laundry, and heater electricity consumption, as well as the total electricity usage, within the prediction set. This analysis followed the model's training on the training set and its subsequent parameter optimization on the validation set. The results of this experiment are depicted in Figure XX. It is crucial to highlight that the total household electricity consumption is not merely the aggregate of these three electricity usage components. Rather, it is derived through a specific formula, detailed as follows:

$$TM = \frac{GAP \times 1000}{60} - (SM_1 + SM_2 + SM_3) \qquad (18)$$

where TM denotes the total household electricity consumption, GAP represents the global active power, and SM_1, SM_2, SM_3 respectively signify the electricity consumption of the kitchen, laundry, and heater.

FIGURE 4. Actual vs Predicted Electricity Consumption for Kitchen, Laundry, Heater, and Total Power



In the diagram, "sub-metering-1," "sub-metering-2," and "sub-metering-3" represent the electricity usage data for the kitchen, laundry room, and water heater, respectively, while "total-metering" indicates the total electricity consumption, TM. The data presented in the graph indicates that across all four comparisons, the predicted results achieved a highly accurate fit with the electricity consumption data. Notably, the predictions for laundry, heater, and total electricity consumption were more accurate than those for kitchen electricity usage.

2) COMPARISON WITH OTHER MODELS

To further elucidate the performance of the CNN-BiLSTM-SA model, we analyzed various evaluation metrics such as RMSE, MAE, and R2 under the standard electricity test set, comparing the proposed model with other models including SVM (Support Vector Machine), GRU (Gated Recurrent Unit), and LR (Linear Regression). The parameters for the comparison models were directly adopted from the settings outlined in [37].

In the original electricity dataset, the sampling frequency was set at once per minute. For the comparative experiments, we preprocessed the data by performing secondary statistics on the minute-level data, aggregating it into per-minute, per-hour, and per-day intervals. This approach was implemented to facilitate an analysis and comparison of the model's performance across various time periods.

The comparisons of the evaluation metrics are respectively presented in Table 4.

TABLE 4: EVALUATION METRICS COMPARISON FOR ELECTRICITY CONSUMPTION PREDICTION MODELS

Method	Frenquency	RMSE	MAE	R2
SVM	Per-minute	0.9836	0.6080	0.4831
	Per-hour	0.7713	0.5800	0.5100
	Per-day	0.5100	0.5249	0.3917
GRU	Per-minute	0.5402	0.4040	0.8213
	Per-hour	0.4603	0.3329	0.8570
	Per-day	0.2825	0.2412	0.7732
LR	Per-minute	6.5197	5.9100	0.6943
	Per-hour	8.2380	4.4700	0.5758
	Per-day	4.0700	4.3020	0.6385
CNN-LSTM	Per-minute	0.6114	0.3493	0.7643
	Per-hour	0.5957	0.3317	0.7925
	Per-day	0.3221	0.2569	0.9148
Proposed	Per-minute	0.2389	0.2890	0.9910
	Per-hour	0.1615	0.2046	0.9936
	Per-day	0.1554	0.2280	0.9894

The table compares the performance of the proposed CNN-BiLSTM-SA model with other models such as SVM, GRU, and LR, using evaluation metrics such as RMSE, MAE, and R2. The results are presented for three different sampling frequencies: per-minute, per-hour, and per-day.

The worst-performing model in terms of RMSE and MAE is LR, while the second-worst is SVM. The best-performing model is the proposed CNN-BiLSTM-SA model, followed by GRU.

The effect of different sampling frequencies on the evaluation metrics is significant. For instance, the RMSE and MAE values for all models are lower when the sampling frequency is per-day compared to per-minute.

The proposed CNN-BiLSTM-SA model is the best model in terms of RMSE, MAE, and R2. The model with the smallest difference from the proposed model is GRU.

The R2 values for all models are high, indicating that the models can fit the data well.

From minute-level, hourly-level, and daily-level data, it is evident that the proposed method outperforms the contrast algorithms significantly across all three metrics, demonstrating a remarkable lead in fitting electricity consumption data. Furthermore, the proposed method performs comparably well at the hourly and daily levels, indicating that CNN-BiLSTM-SA can achieve quite accurate predictions at the hourly level. At the minute level, it also exhibits outstanding performance.

In summary, the proposed CNN-BiLSTM-SA model outperforms other models in terms of RMSE, MAE, and R2 across all sampling frequencies. The effect of different sampling frequencies on the evaluation metrics is significant, with lower RMSE and MAE values for all models when the sampling frequency is per-day compared to per-minute. The R2 value for our model is the highest, indicating that the model can fit the data well. This indicates that the proposed model can achieve precise predictions of household electricity consumption, thereby providing theoretical and technical support for electricity optimization tasks such as power supply strategies and equipment output for power companies.

3) ABLATION

To further clarify the contribution of each component within the proposed model to its overall performance, ablation experiments are we conducted. The EP_{RMSE} and EP_{MAE} are utilized as valuation metrics (see Eq.16 and Eq.17). The term EP_{RMSE} and EP_{MAE} represent the difference between the performance metrics of the model after removing specific components and the complete CNN-BiLSTM-SA model metrics. The larger the metric values, the greater the contribution of the removed components to the overall performance of the network.

The results are presented in Table 5.

TABLE 5: ABLATION EXPERIMENT RESULTS FOR MODEL COMPONENTS

Method	EP_{RMSE}	EP_{MAE}
BiLSTM	82.70	76.71
BiLSTM-SA	43.24	50.38
CNN-BiLSTM	17.93	22.80
CNN-BiLSTM-SA	0.00	0.00

The data from the table indicates that the RMSE values of the complete model are respectively 82.70%, 43.24%, and 17.93% lower than those of BiLSTM, BiLSTM-SA, and CNN-BiLSTM. Similarly, the MAE values of the complete model are 76.71%, 50.38%, and 22.80% lower in comparison to BiLSTM, BiLSTM-SA, and CNN-BiLSTM.

This indicates that each component of the model significantly impacts the prediction effectiveness. Specifically, considering the BiLSTM and SA components, compared to the model without SA, the complete model shows a reduction of 82.7% and 76.71% in RMSE and MAE metrics, respectively. Upon integrating the SA mechanism, the complete model exhibits a decrease of 43.24% and 50.38% in these two metrics, respectively, highlighting the substantial impact of the SA mechanism. In the case of the CNN and BiLSTM, compared to the scenario with the addition of the CNN network, the complete model shows a reduction of 17.93% and 22.80% in RMSE and MAE metrics, respectively, indicating that the CNN module also contributes to the model's stability. Furthermore, data from the table reveals that while each module significantly enhances model performance, SA's contribution to stability surpasses that of CNN. The ablation study demonstrates the proposed model's validity.

VI. CONCLUSION

We propose a CNN-BiLSTM-SA model for predicting household electricity consumption. By extracting, analyzing, and processing relevant features that may influence electricity usage, our model is capable of rapid, efficient, and accurate predictions. To validate the performance of the proposed model, we designed a series of experiments, including validation experiments, comparative experiments, and ablation studies. In the validation experiments, we compared the predicted results with actual values in the test set, and the results were satisfactory, demonstrating that CNN-BiLSTM-SA is suitable for electricity prediction tasks. In the comparative experiments, we contrasted our model with three popular networks, and the results indicated a dominant advantage in all evaluation metrics across all test scales. Additionally, to clarify the contribution of each module to the model, we conducted ablation studies, which yielded reasonable and satisfying results. We have verified the superior performance of the proposed model on the standard electricity test set. However, we still aim to conduct further experiments on more recent and fresh datasets. Due to resource limitations, there are no public test sets available other than the electricity dataset to date. Therefore, our next step is to attempt independent data collection to establish a dataset, so as to facilitate more in-depth research into household electricity consumption prediction.

ACKNOWLEDGEMENTS: This study Supported by the Undergraduate Training Progrom on Innovation and Entreprenshipgrant X202310251165

AUTHOR CONTRIBUTIONS: The authors confirm contribution to the paper as follows: study conception and design: MENG-PING WU; data collection: Fan Wu; analysis and interpretation of results: Fan Wu; draft manuscript preparation: MENG-PING WU. All authors reviewed the results and approved the final version of the manuscript.

CONFLICTS OF INTEREST: The authors declare that they have no conflicts of interest to report regarding the present study

DATA AVAILABILITY STATEMENT: Not applicable.

REFERENCES

[1] Duffy P , Fitzpatrick C , Conway T ,et al.Energy Sources and Supply Grids – The Growing Need for Storage[J].Issues in environmental science and technology, 2018, 46:1-41.DOI:10.1039/9781788015530-00001.
[2] Cuicui D , Economics S O .The Dynamic Effect and Regional Difference of the Industrialization and Urbanization on the Energy Consumption Intensity in China[J].Journal of Hebei University of Economics and Business, 2015.DOI:10.1021/ac60060a023.
[3] Hojjati B. Global energy consumption driven by more electricity in residential, commercial buildings[J]. US Energy Information Administration, 2019.
[4] Pappas S S, Ekonomou L, Karampelas P, et al. Electricity demand load forecasting of the Hellenic power system using an ARMA model[J]. Electric Power Systems Research, 2010, 80(3): 256-264.

[5] Andoh P Y A, Sekyere C K K, Mensah L D, et al. Forecasting electricity demand in Ghana with the SARIMA model[J]. Journal of Applied Engineering and Technological Science (JAETS), 2021, 3(1): 1-9.

[6] Xiuhua C , Xiren M .Periodical Fluctuant Prediction Model in Digital Household Power Consumption[J].Low Voltage Apparatus, 2007.

[7] Li X, Bowers C P, Schnier T. Classification of energy consumption in buildings with outlier detection[J]. IEEE Transactions on Industrial Electronics, 2009, 57(11): 3639-3644.

[8] Jahani E, Cetin K, Cho I H. City-scale single family residential building energy consumption prediction using genetic algorithm-based Numerical Moment Matching technique[J]. Building and Environment, 2020, 172: 106667.

[9] Zolfaghari M, Sahabi B. A hybrid approach to model and forecast the electricity consumption by NeuroWavelet and ARIMAX-GARCH models[J]. Energy Efficiency, 2019, 12: 2099-2122.

[10] Arghira N, Hawarah L, Ploix S, et al. Prediction of appliances energy use in smart homes[J]. Energy, 2012, 48(1): 128-134.

[11] Amber K P, Ahmad R, Aslam M W, et al. Intelligent techniques for forecasting electricity consumption of buildings[J]. Energy, 2018, 157: 886-893.

[12] Geetha R, Ramyadevi K, Balasubramanian M. Prediction of domestic power peak demand and consumption using supervised machine learning with smart meter dataset[J]. Multimedia Tools and Applications, 2021, 80: 19675-19693.

[13] Syah I F, Abdullah M P, Syadli H, et al. Data Selection Test Method for Better Prediction Of Building Electricity Consumption[J]. J. Teknol, 2016, 78: 67-72.

[14] Farzana S, Liu M, Baldwin A, et al. Multi-model prediction and simulation of residential building energy in urban areas of Chongqing, South West China[J]. Energy and Buildings, 2014, 81: 161-169.

[15] Akyol H B, Preist C, Schien D. Avoiding Overconfidence in Predictions of Residential Energy Demand Through Identification of the Persistence Forecast Effect[J]. IEEE Transactions on Smart Grid, 2022, 14(1): 228-238.

[16] Spiliotis E, Doukas H, Assimakopoulos V, et al. Forecasting week-ahead hourly electricity prices in Belgium with statistical and machine learning methods[M]//Mathematical modelling of contemporary electricity markets. Academic Press, 2021: 59-74.

[17] Kim M K, Kim Y S, Srebric J. Predictions of electricity consumption in a campus building using occupant rates and weather elements with sensitivity analysis: Artificial neural network vs. linear regression[J]. Sustainable Cities and Society, 2020, 62: 102385.

[18] Yang S, Wang Y. Applying support vector machine method to forecast electricity consumption[C]//2006 International Conference on Computational Intelligence and Security. IEEE, 2006, 1: 929-932.

[19] Tso G K F, Yau K K W. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks[J]. Energy, 2007, 32(9): 1761-1768.

[20] Wang C, Bäck T, Hoos H H, et al. Automated machine learning for short-term electric load forecasting[C]//2019 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, 2019: 314-321.

[21] Rahman S, Alam M G R, Rahman M M. Deep learning based ensemble method for household energy demand forecasting of smart home[C]//2019 22nd International Conference on Computer and Information Technology (ICCIT). IEEE, 2019: 1-6.

[22] Yan K, Wang X, Du Y, et al. Multi-step short-term power consumption forecasting with a hybrid deep learning strategy[J]. Energies, 2018, 11(11): 3089.

[23] Aurangzeb K. Short Term Power Load Forecasting using Machine Learning Models for energy management in a smart community[C]//2019 International Conference on Computer and Information Sciences (ICCIS). IEEE, 2019: 1-6.

[24] Guo Z, Zhou K, Zhang X, et al. A deep learning model for short-term power load and probability density forecasting[J]. Energy, 2018, 160: 1186-1200.

[25] Alhussein M, Aurangzeb K, Haider S I. Hybrid CNN-LSTM model for short-term individual household load forecasting[J]. Ieee Access, 2020, 8: 180544-180557.

[26] **ong X, Wei Y. The Analysis and Predication of Energy Use in Smart Homes Based on Machine Learning[C]//2020 International Conference on Computing and Data Science (CDS). IEEE, 2020: 381-386.

[27] Dlamini G, Megha S. Predicting Household power consumption: Using Gradient Boosting and Deep Quantile Regression Model[C]//Journal of Physics: Conference Series. IOP Publishing, 2020, 1694(1): 012031.

[28] Hyeon J, Lee H Y, Ko B, et al. Deep learning-based household electric energy consumption forecasting[J]. The Journal of Engineering, 2020, 2020(13): 639-642.

[29] Irankhah A, Rezazadeh S, Moghaddam M H Y, et al. Hybrid deep learning method based on lstm-autoencoder network for household short-term load forecasting[C]//2021 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS). IEEE, 2021: 1-6.

[30] Hajjaji I, El Alami H, El-Fenni M R, et al. Evaluation of Artificial Intelligence Algorithms for Predicting Power Consumption in University Campus Microgrid[C]//2021 International Wireless Communications and Mobile Computing (IWCMC). IEEE, 2021: 2121-2126.

[31] Alkhatib W, Alhamoud A, Böhnstedt D, et al. Hybrid Model for Large Scale Forecasting of Power Consumption[C]//Advances in Computational Intelligence: 14th International Work-Conference on Artificial Neural Networks, IWANN 2017, Cadiz, Spain, June 14-16, 2017, Proceedings, Part I 14. Springer International Publishing, 2017: 661-672.

[32] Velasco L, Polestico D, Macasieb G, et al. A Hybrid Model of Autoregressive Integrated Moving Average and Artificial Neural Network for Load Forecasting[J]. International Journal of Advanced Computer Science and Applications, 2019, 10(11): 14-16.

[33] Saxena H, Aponte O, McConky K T. A hybrid machine learning model for forecasting a billing period's peak electric load days[J]. International Journal of Forecasting, 2019, 35(4): 1288-1303.

[34] Kesornsit W, Sirisathitkul Y. Hybrid Machine Learning Model for Electricity Consumption Prediction Using Random Forest and Artificial Neural Networks[J]. Applied Computational Intelligence and Soft Computing, 2022, 2022.

[35] Irankhah A, Rezazadeh S, Moghaddam M H Y, et al. Hybrid deep learning method based on lstm-autoencoder network for household short-term load forecasting[C]//2021 7th International Conference on Signal Processing and Intelligent Systems (ICSPIS). IEEE, 2021: 1-6.

[36] Hajjaji I, El Alami H, El-Fenni M R, et al. Evaluation of Artificial Intelligence Algorithms for Predicting Power Consumption in University Campus Microgrid[C]//2021 International Wireless Communications and Mobile Computing (IWCMC). IEEE, 2021: 2121-2126.

[37] Kim T Y, Cho S B. Predicting residential energy consumption using CNN-LSTM neural networks[J]. Energy, 2019, 182: 72-81.

MENG-PING WU is currently studying in Computer Science and Technology for a bachelor's degree at the School of Information Science and Engineering, East China University of Science and Technology, China. Her research interests include deep learning and data analytics.

Fan Wu is currently studying for a bachelor's degree in Automation at the School of Information Science and Engineering, East China University of Science and Technology, China. His research interests include deep learning and data mining.