Privacy-preserving household load forecasting based on non-intrusive load monitoring: A federated deep learning approach

Xinxin Zhou[1,*], Jingru Feng[1,*], Jian Wang[1] and Jianhong Pan[2]

[1] School of Computer Science, Northeast Electric Power University, jilin, China
[2] State Grid Jilin Electric Power Company Limited, Changchun, China
* These authors contributed equally to this work.

ABSTRACT

Load forecasting is very essential in the analysis and grid planning of power systems. For this reason, we first propose a household load forecasting method based on federated deep learning and non-intrusive load monitoring (NILM). As far as we know, this is the first research on federated learning (FL) in household load forecasting based on NILM. In this method, the integrated power is decomposed into individual device power by non-intrusive load monitoring, and the power of individual appliances is predicted separately using a federated deep learning model. Finally, the predicted power values of individual appliances are aggregated to form the total power prediction. Specifically, by separately predicting the electrical equipment to obtain the predicted power, it avoids the error caused by the strong time dependence in the power signal of a single device. In the federated deep learning prediction model, the household owners with the power data share the parameters of the local model instead of the local power data, guaranteeing the privacy of the household user data. The case results demonstrate that the proposed approach provides a better prediction effect than the traditional methodology that directly predicts the aggregated signal as a whole. In addition, experiments in various federated learning environments are designed and implemented to validate the validity of this methodology.

INTRODUCTION

Since AC power cannot be stored, the power generation capacity of the power plant must reach a stable dynamic balance with the power consumption (*López et al., 2015*), in order to guarantee that the power grid can operate stably and efficiently, reasonable power supply plans and power dispatching methods need to be developed, and accurate power forecasting is a necessary condition. With the continuous growth of the world's population, the demand for energy from mankind is increasing day by day. At present, domestic dwellings account for about 20%–40% of the world's total energy use, and the household demand load often has a significant impact on the peak load of seasonal and daily electricity (*Wu et al., 2014*). Due to the inherent uncertainty of distributed renewable

Corresponding author
Jingru Feng, 1665639965@qq.com

energy resources (*Li et al., 2022a*, *2022b*), the renewable integration also brings new challenges to household load forecasting. Accurate forecasting of household electricity demand is not only a prerequisite for guaranteeing the safety and stability of the power grid, but also for peaking and flattening valleys by means of floating electricity prices and balancing the supply and demand relationship at peak electricity consumption.

Literature review

At present, load forecasting methods commonly used abroad can be separated into statistical methods (*e.g.*, the non-linear extrapolation method (*Wang et al., 2012*; *Nazih, Fawwaz & Osama, 2011*; *Sadaei et al., 2019*), time series method (*Aly, 2020*; *Gasparin, Lukovic & Alippi, 2022*; *Wang, Xia & Kang, 2010*, *etc.*) and machine learning forecasting methods (*e.g.*, artificial neural networks (ANN) (*Zeng & Qiao, 2013*; *Yildiz, Bilbao & Sproul, 2017*; *Cui, Wang & Yue, 2019*; *Kumar, Mishra & Gupta, 2016*), support vector machines (SVM) (*Shi, Li & Yu, 2009*; *Li & Gu, 2013*; *Liu et al., 2018*, *etc.*). Since a single forecasting method often has various defects, the combined forecasting method is gradually applied to the research of power system load forecasting, and its performance is better than all independent load forecasting models (*Ziel & Liu, 2016*). *Chen & Chang (2004)* proposed a load forecasting method based on user clustering. After clustering historical load curves, wavelet analysis is proposed to predict the future 24-h load curve. Reference (*Stephen et al., 2015*) compared the performance of seven existing load forecasting technologies, including linear regression, ANN, SVM, and their variants on two commercial buildings and three residential household data sets. The results show that these traditional system-level load forecasting methods can provide good forecasts for commercial buildings, but perform poorly on household data. As for load prediction at the household level, there are still few reference works, and most of the existing forecasting methods directly predict the total power, such as Kalman filter (*Guan et al., 2013*; *Sharma et al., 2020*; *Zheng, Chen & Luo, 2019*), Autoregressive comprehensive moving average model (*Stephen et al., 2015*; *Lee & Ko, 2011*; *Kuster, Rezgui & Mourshed, 2017*), etc.

Research on household or residential electricity demand estimation currently takes several different ideas, including evaluating and improving existing load forecasting techniques, developing and researching new methods, or a combination of the two. *Wang et al. (2018)* introduced a deep recursive neural network to learn and correlate shared information between customers and solve the problem of overfitting. The performance of this method on the Irish housing data set exceeds the autoregressive integrated moving average model (ARIMA), SVM, and classic recurrent neural networks. *Ponoćko & Milanović (2018)* proposed a convolutional-long short-term memory neural network (CNN-LSTM) model with selective autoregressive characteristics to enhance single-step-ahead power load forecasting. This improvement can be observed in the three spatial granularities of apartments, floors, and building floors and achieves higher prediction accuracy. Recently, some scholars have proposed a household prediction method based on NILM, which can be more accurately used for household-level power prediction. The reason is that certain equipment used alone may have a strong time dependence, such as air conditioning. In summer or winter, we turn on the air conditioner but it is not turned

on 24 h a day. In spring or autumn, it is only turned on occasionally, which means that the air conditioner changes over time. When different appliances are aggregated, the time dependence of this single appliance may be broken. Therefore, predicting the total power distribution of the entire family may be less accurate than predicting the power distribution of a single appliance. *Li et al. (2021)* proposed a method for household electricity consumption prediction based on NILM, which combines the correlation of electrical behavior in its state duration, and the experiment has good accuracy, but this method does not consider time comprehensively information. *Fang et al. (2020)* developed a NILM algorithm for analyzing the operating characteristics of different appliances. *Dinesh, Makonin & Bajić (2020)* proposed a new method for single-household electricity consumption prediction based on NILM and aggregated spectral clustering, and used four public data sets for testing. However, this method has a high error, does not take advantage of deep learning, and does not describe the part of data preprocessing. Such deep learning based prediction methods have also been the subject of research in recent years since rich training data yield the expected model performance and deep learning based methods have a great number of trainable parameters. At present, there are very few existing researches on this kind of household-level load forecasting based on NILM.

In addition, if the predictive model is to be applied to a wider range of electricity consumers and appliances, sufficient electricity demand data must be collected from multiple data owners, but multi-source load data may cause privacy disputes (*Wang et al., 2021*). In practice, some households may be reluctant to share data with others, especially when power data information is stored in a central server or during transmission. Due to the frequent exchange of information, it is easy for a third party to obtain power data (*Li et al., 2020*). Currently, many countries have promulgated specific laws. For instance, the "Network Security Law" promulgated in 2017 clearly requires that the collected user information should be strictly confidential (*Li, Li & Wang, 2022*). Thus, to resolve the above-mentioned questions and achieve a balance between consumer privacy protection and data utilization, federated learning is proposed. Federated learning does not need to collect raw data, that is, the private data of family users does not need to be uploaded to the central server for training. Each family participant can conduct model training locally, and finally, just upload the updated model parameters to the cloud server (*Yang et al., 2020*). The framework proposed by federated learning balances the relationship between data utilization and privacy protection, and in the process improves the performance of the co-training model. At present, there is also a very small part of the literature discussing the application of federated learning in electric energy. For example, *Wang et al. (2021)* proposed a federated learning method for power consumer feature identification. This method was composed of privacy protection principal component analysis and FL-ANN model. This model has been proven to perform better. *Briggs, Fan & Andras (2021)* used FL-LSTM for energy demand forecasting, preserving the data privacy of energy consumers, and compared a centrally trained approach with a localized training approach. In addition, how to apply federated learning to the literature at the household load level is also an urgent and challenging issue.

Contribution of this article

In view of the above, most of the current household load forecasts are directly based on the total power forecast as a whole, which will cause large errors. As for the few integrated prediction methods based on NILM, there are also a series of problems such as low prediction accuracy, lack of data preprocessing, not taking advantage of deep learning, or not considering the comprehensive connection of time. In addition, the centralized model is vulnerable to external attacks, which may lead to third parties having the opportunity to obtain load data. In particular, load information may be destroyed when stored in a centralized server or during transmission.

This article presents for the first time a novel federated deep learning prediction model based on NILM. This method integrates federated learning and Bi-directional long short-term memory network-Attention mechanism (BiLSTM-Attention). The following are the main contributions:

(1) This article presents a load prediction method based on federated deep learning (FedDL), which combines federated learning and BiLSTM-Attention. For all we know, this is the first study of federated learning in household load forecasting based on NILM. In this prediction method, under the precondition of guaranteeing the accuracy of the forecast, the data of the users in the house is not leaked, and the data can be isolated to ensure that the data of the users in the house is not leaked, and meet the needs of user privacy protection and data security. On the precondition of guaranteeing the accuracy of the prediction, while reducing the data security risk, it can also learn from the electrical power of other households.

(2) A method based on NILM to predict user household energy consumption is proposed. The NILM technology extracts a single load pattern from the historical total load demand, avoiding the strong time dependence of a single device. Experiments demonstrate that the integrated prediction method based on NILM has better prediction results than the traditional method that directly predicts the aggregated signal as a whole.

(3) Use the data decomposed by the CNN-LSTM hybrid deep learning model to test the proposed federated deep learning prediction model to validate the validity of the proposed model; In addition, experiments in various federated learning environments are designed and implemented to validate the validity of this methodology.

Organization of this article

The main points of this article are as follows: Section 1 introduces the relevant background and relevant work on household load forecasting; Section 2 presents the associated theories of machine learning; Section 3 mainly proposes the federated deep learning model, and introduces the specific process of training this model. Section 4 performs the relevant training and testing of the proposed FedDL model using the data obtained based on the NILM, and analyzes the test results. Section 5 summarizes this article.
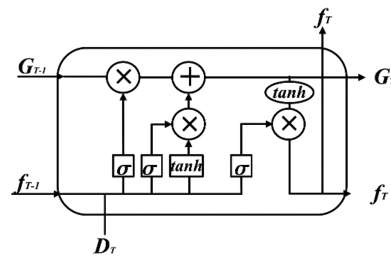
Figure 1 LSTM structure diagram.

RELATED PRELIMINARY WORK

Convolutional neural network (CNN)

Due to their powerful learning and generalization capabilities, artificial neural networks are widely used to deal with pattern recognition problems in the engineering field (*Shi et al., 2008*; *Hou et al., 2020*). This article uses CNN to extract data features in the data preprocessing part. Eq. (1) is the output of the convolutional layer:

$$a_j^l = \text{conv}\left(\sum w_{ij}^l X_i^{l-1} + w_b\right) \tag{1}$$

where w_{ij}^l represents the weight of the convolutional layer; w_b indicates the bias term of the filter; a_j^l and X_i^{l-1} are the input and output in the convolution process (*Lawrence et al., 1997*), separately. The nonlinear activation function is:

$$X_j^l = f\left(a_j^l\right) \tag{2}$$

$$X_j^{l+1} = \text{pooling}\left(X_j^l\right) \tag{3}$$

The pooling function represents the pooling operation, X_j^{l+1} represents the output after the pooling layer.

Bi-directional long short-term memory network-attention mechanism

Long short-term memory network (LSTM)

LSTM is a typical recurrent neural network, which is particularly suitable for dealing with time-series related pattern recognition problems. Fig. 1 shows the LSTM structure.

The input parameters are the memory cell state G_{T-1} and hidden layer state f_{T-1} at time $(T-1)$, and the input sequence element value D_T at time T. The output parameters are the memory cell state G_T and the hidden layer state f_T at time T. The corresponding calculation equation in Eqs. (4)–(8):

$$r_T = \sigma(W_i[f_{T-1}, D_T] + h_i) \tag{4}$$

$$l_T = \sigma\left(W_f[f_{T-1}, D_T] + h_f\right) \tag{5}$$

$$g_T = \tanh(W_c[f_{T-1}, D_T] + h_c) \tag{6}$$

$$G_T = r_T * g_T + l_T * G_{T-1} \tag{7}$$

$$P_T = \sigma(W_o[f_{T-1}, D_T] + h_o) \tag{8}$$

$$f_T = P_T * \tanh(G_T) \tag{9}$$
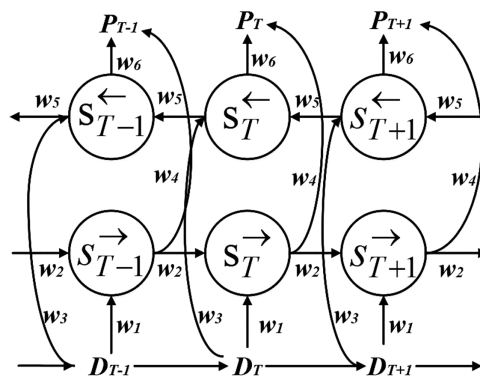
Figure 2 BiLSTM structure diagram.

where W_i, W_f, W_o W_c represent the weight matrix of input gate, forget gate, output gate, and memory unit state respectively; h_i, h_f, h_o h_c represent the bias terms of each gating unit and memory unit (*Lindemann et al., 2021*); σ represents the sigmoid function (*Zhang et al., 2021*); ∗ means that the corresponding elements of the two vectors involved in the operation are multiplied.

BiLSTM

The BiLSTM network has the function of capturing the characteristics of the information before and after (*Liu & Guo, 2019*). Since in the actual power system, the power load data is a time-varying series of data and non-linear, its load at a certain moment is not only influenced by other factors (such as holidays and social environment), but also simultaneously associated with the past input features, while the future input features can also reflect the present load features to some extent (*Liu, Lee & Lee, 2020*). Therefore, in order to effectively obtain the time-series variation features about time, BiLSTM is used to mine the intrinsic connection between the current data and the data of past and future moments, which is shown in Fig. 2.

The BiLSTM hidden layer state P_T at moment T can be found by the forward hidden layer state $\overrightarrow{s_T}$ and the backward hidden layer state $\overleftarrow{s_T}$ in two parts. The forward hidden layer state $\overrightarrow{s_T}$ is decided by the current input D_T and the forward hidden layer state $\overrightarrow{s_{T-1}}$ at moment $(T+1)$. The backward hidden layer state $\overleftarrow{s_T}$ is decided by the current input and the backward hidden layer state $\overleftarrow{s_{T+1}}$ at moment $(T+1)$. The calculation formula is shown below, where $w_i(i=1, 2…, 6)$ is the weight from one cell layer to another cell layer.

$$\overrightarrow{s_T} = f\left(w_1 D_t + w_2 \overrightarrow{s_{T-1}}\right) \tag{10}$$

$$\overleftarrow{s_T} = f\left(w_3 D_T + w_5 \overleftarrow{s_{T+1}}\right) \tag{11}$$

$$P_T = g\left(w_4 \overrightarrow{s_T} + w_6 \overleftarrow{s_T}\right) \tag{12}$$

Attention mechanism

The attention mechanism highlights key information through weight distribution to ensure that features are not lost in the model training process, so it can more effectively mine the long-distance data features of time series with correlation (*Bin et al., 2018*). The
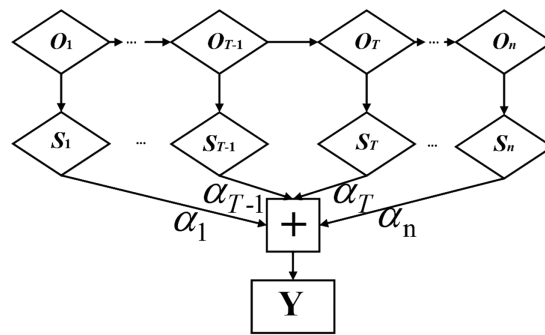
Figure 3 Structure diagram of attention mechanism.

attention mechanism takes the timing sequence of the input power load-related data as the reference object of the neural network (*Ji et al., 2019*). Combined with BiLSTM, the neural network spontaneously selects to retrieve the information from the BiLSTM. Its basic structure is shown in Fig. 3.

O_T is the T-th feature vector of BiLSTM network output, which is input into the hidden layer of attention mechanism to obtain the initial state vector S_T, and then the final output state vector Y is obtained by multiplying and summating the corresponding weight coefficient α_T. The calculation formula is as follows:

$$e_T = \tanh(w_T S_T + b_T) \tag{13}$$

$$\alpha_T = \frac{exp(e_T)}{\sum_{i=1}^{T} e_i} \tag{14}$$

$$Y = \sum_{T=1}^{n} \alpha_T S_T \tag{15}$$

where w_T represents the weight coefficient matrix of the T-th feature vector, and b_T denotes the offset corresponding to the T-th feature vector (*Li, Zhang & Chen, 2022*).

Federated learning

Federated learning is a distributed training method that uses data sets scattered among the participants, integrates data information from multiple parties through privacy protection technology, and collaboratively builds a global model (*Huong et al., 2021*). During the model training process, the relevant information about the model (such as model parameters, model structure, parameter gradient, *etc.*) can be exchanged among participants (The exchange mode can be plaintext, data encryption, adding noise, *etc.*) (*Lu et al., 2019*), but the local training data will not leave the local. This exchange does not expose local user data, decreasing the danger of data leakage. Trained federated learning models can be shared and deployed across data participants.

The training process can be organized into three steps:

(1) Task initialization: the server determines the training target, parameters, and participating terminal devices, and passes the global model ω_G^0 to the participating equipment.

(2) Local model training and update: global model ω_G^t downloaded from the server, where t represents the current training round. The participant appliance i trains the model locally and updates the parameter ω_i^t this time. The updated local model parameter is expressed as ω_i^a, the goal of training is to find the best local model by minimizing the loss function $L(\omega_i^t)$ and upload it to the server:

$$\omega_i^a = arg_{\omega_i^t} min \left[L(\omega_i^t) \right] \tag{16}$$

(3) Global model aggregation and update: The server receives the model parameters uploaded by the participant devices and aggregates them, and finally updates the global model ω_G^{t+1}:

$$L(\omega_i^t) = \frac{1}{N} \sum_{i=1}^{N} L(\omega_i^t) \tag{17}$$

The above training process is iterated until the convergence or termination condition training technique is satisfied. From the classical training model of federated learning, we can find that the local appliance communicates with the server through model parameters during training, so as to protect privacy.

PROPOSED PREDICTION MODEL

In this part, the process of each part is explained in detail. Firstly, the data is preprocessed, and the individual load of household appliances is decomposed from the total load demand through the NILM method based on the CNN-LSTM hybrid model. Then the power of individual appliances is predicted separately by the FedDL prediction model, and the predicted power of individual appliances is aggregated to form the overall power prediction. Finally, the validity of the proposed model is evaluated by some evaluation indicators.

Data decomposition

In this article, the NILM method is based on the CNN-LSTM hybrid model to decompose the total load data of the training sample and obtain the energy consumption value of individual appliances. Fig. 4 presents the structure of the CNN-LSTM.

Firstly, the CNN and LSTM are trained separately using the pre-processed data, and then the feature information extracted from each of the two networks is handled as the equal dimension using the mapping layer, and finally, the two are concatenated. For more specifics, please refer to *Zhou, Feng & Li (2021)*.

FedDL-based prediction model

In order to build a better neural network, a large amount of data containing various devices and responding to consumer behavior habits needs to be fed to the network. However, during model training, the local data owner may encounter the risk of privacy leakage and
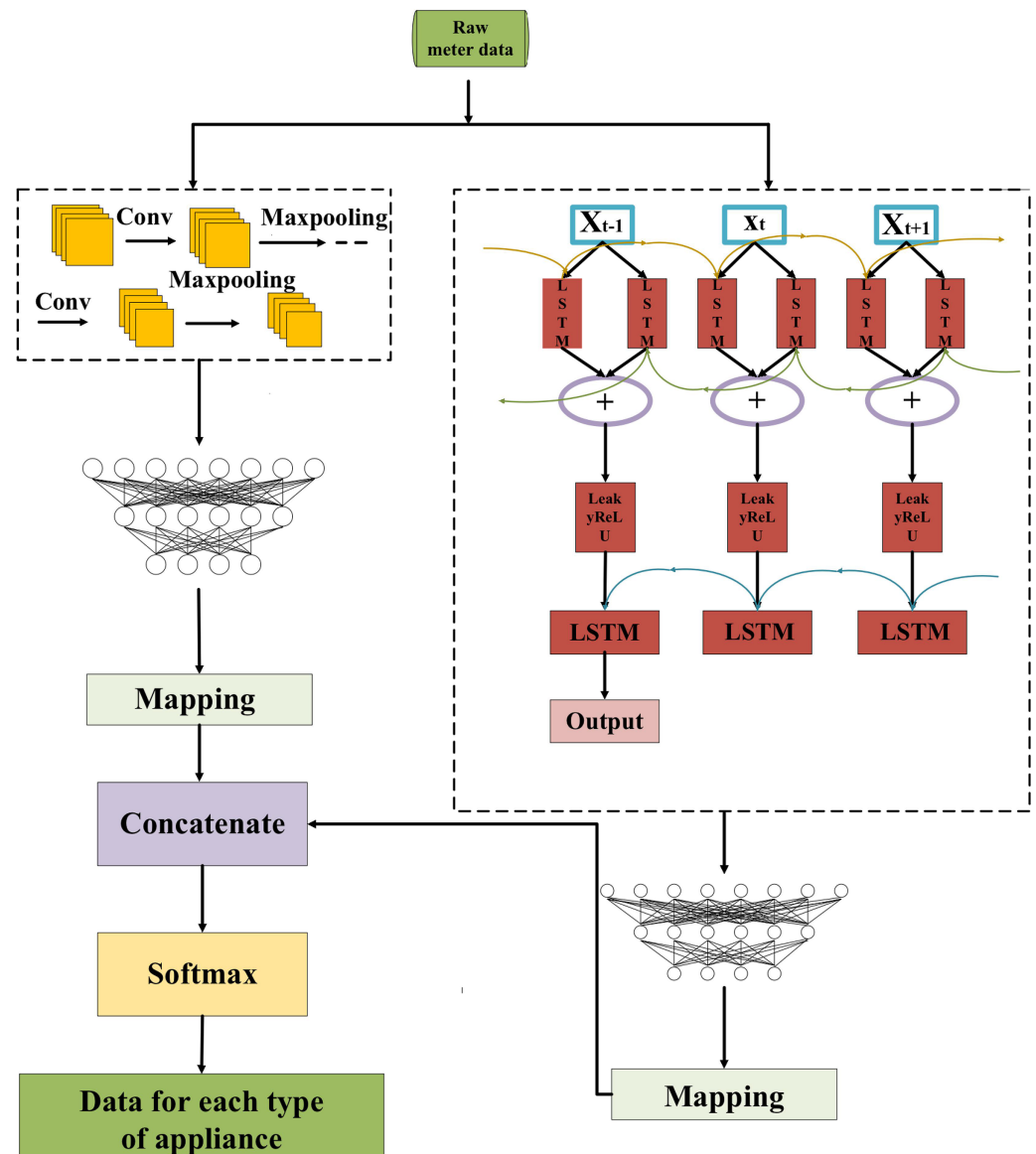
Figure 4 Structure diagram of CNN-LSTM.

loss of control over the data. To address this potential risk, we use federated learning, which is in sharp contrast to traditional centralized machine learning techniques where all sample data is uploaded to a single server. The flow chart of the proposed FedDL prediction model is shown in Fig. 5.

For the purpose of illustrating, we assume that each household has the same computational power, the same hyperparameters (*e.g.*, learning rate, minibatch size) for training in FedDL, and the same optimization algorithm is used.

As shown in Fig. 5, first, the initial global model will deploy its parameters to the client. After receiving the parameters of the global model, the local household will train its local deep learning model based on local data and update its parameters, and then transmit these parameters to the global server. The global deep learning model receives the locally
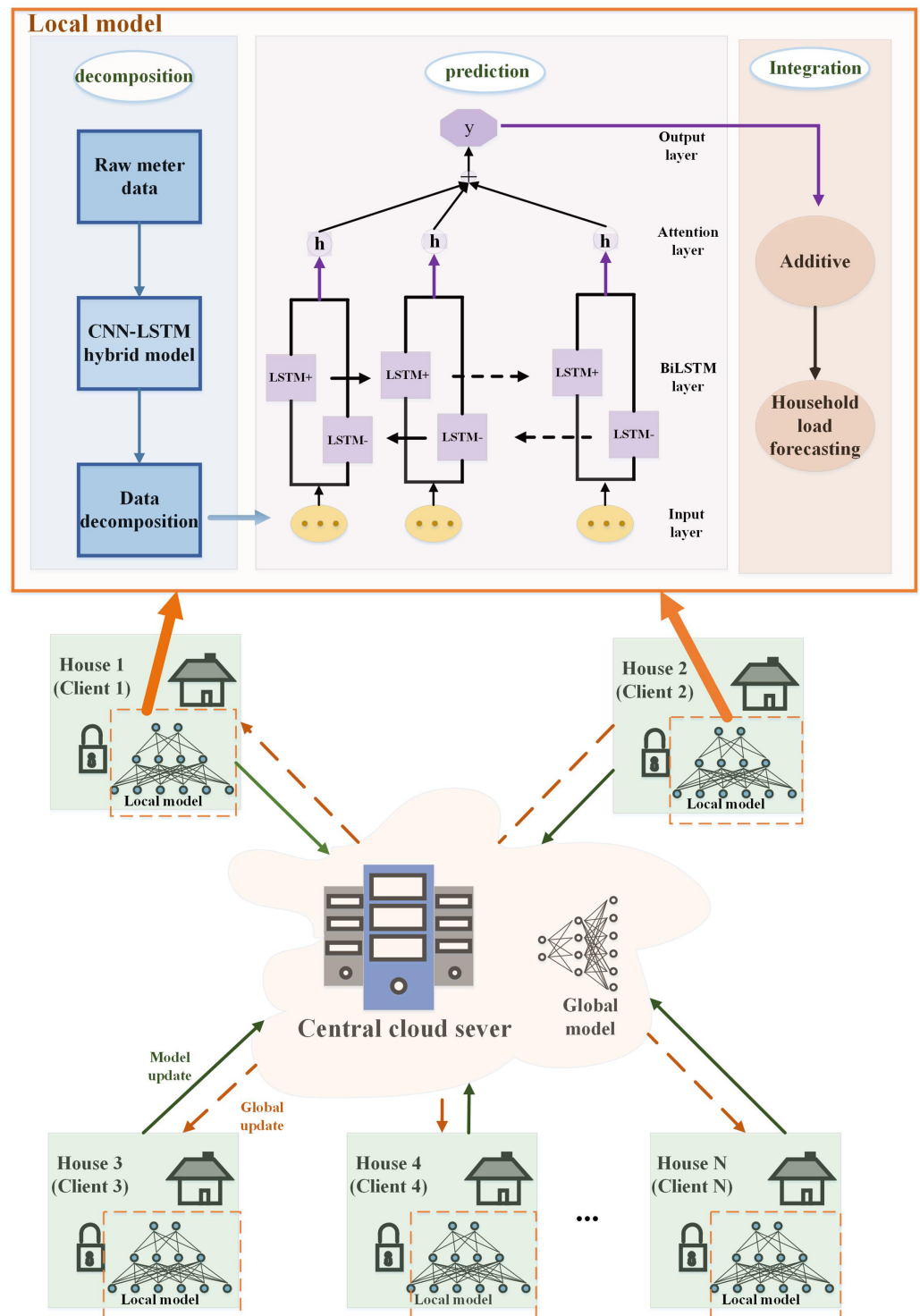
Figure 5 Flowchart of the FedDL prediction model.

updated parameters and adopts a federated average for the parameters, and then updates the global deep learning model with the results based on the federated average. After that, the updated global parameters will be passed to the local model of each house, and the final
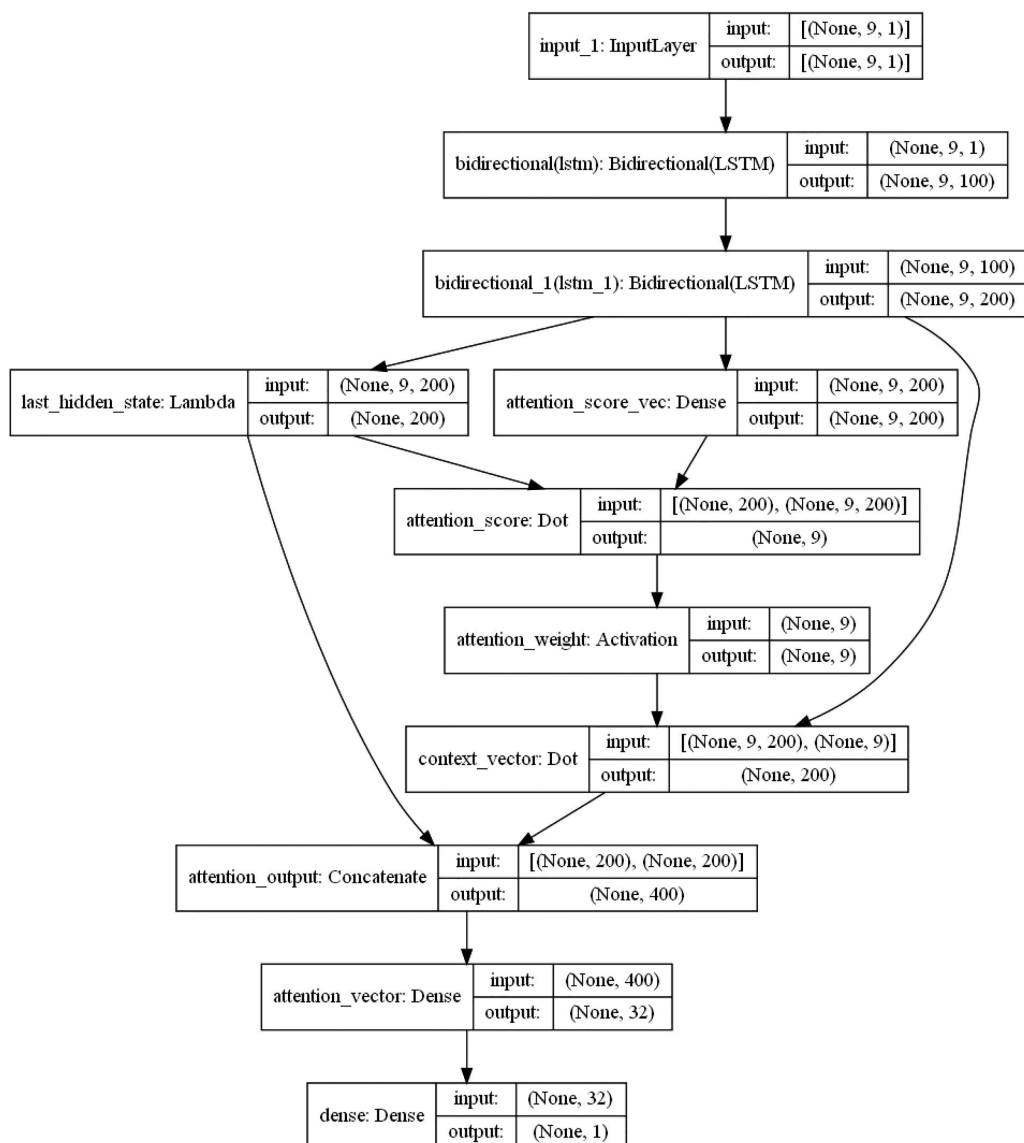
Figure 6 BiLSTM-Attention internal structure diagram.

global deep learning model will be generated after all the global communication rounds have been run between the local residents and the central server.

We explained it step by step under the local model. The individual load of household appliances is decomposed from the total load demand through the NILM method based on the CNN-LSTM hybrid model. Then the power of individual appliances is predicted separately by the prediction model, and the predicted power of individual appliances is aggregated to form the overall power prediction.

Local model structure parameters

The specific internal structure is shown in Fig. 6. First, the model delivers the pre-processed sequence data to the forward and the backward LSTM hidden layer, and the two

are combined with the output vector as the bi-directional timing feature vector of the electric load at the moment *T*. The attention layer is responsible for assigning attention weights, highlighting the key factors, ignoring irrelevant information, and continuously iterating the optimal weight parameter matrix. Then, in order to obtain the new vector of inputs to the fully connected layer, the weight vector is combined with the shallow layer output. Finally, the predicted load values are obtained by the fully connected layer (sigmoid as the activation function).

The LSTM layer can remember or forget significant information. In principle, the more layers, the stronger the learning ability. But too many layers will lead to difficult convergence and a long training time. Therefore, in the model of this article, a two-layer LSTM is set. The number of neurons in the first layer is set as 128, and the number of neurons in the second layer is 68.

Evaluation indicators of federated deep learning model

In machine learning, there are two major problems: classification and regression. The essence of both is to predict the input. The difference between the two lies in the type of output results. The output of the classification problem is a qualitative discrete variable, while the output of the regression problem is a quantitative continuous variable. For these two kinds of problems, there are corresponding evaluation indexes to measure the validity of the models and algorithms. As a typical regression problem, power load prediction mainly has the following evaluation indexes:

(1) Mean Absolute Error (MAE): The advantage is that the calculation is simple and the description error is clear. The specific formula is shown in Eq. (18):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left| y_{tru} - y_{pre} \right| \tag{18}$$

(2) Root Mean Square Error (RMSE): For data whose unit measurement is too large, this operation can retain the original description unit for the data. For example, if the unit of electricity load is KW, the square of the difference will increase by an order of magnitude, which will add trouble to the model description, so RMSE is chosen to maintain the original magnitude. The smaller the original root mean square error, the more accurate the prediction. The specific formula is shown in Eq. (19):

$$\text{RMSE} = \left(\frac{1}{n} \sum_{i=1}^{n} (y_{tru} - y_{pre})^2 \right)^{\frac{1}{2}} \tag{19}$$

In Eqs. (18) and (19), y_{tru} and y_{pre} are the true and predicted values, respectively.

CASE STUDY

In this part, the integrated power is decomposed into individual device power by the NILM method based on the CNN-LSTM hybrid model, and the power of each device is predicted

separately using the federated deep learning model. Finally, the predicted power values of individual appliances are aggregated to form the overall power prediction. The validity of the method is confirmed through qualitative and quantitative evaluation results. In addition, experiments in various federated learning environments are designed and implemented to validate the validity of this methodology. The programming language in the experiment is python and Tensorflow2.0 is used.

In the BiLSTM-Attention model used in this article, a 2-layer LSTM is set. The number of neurons in the first layer is set as 128, and the number of neurons in the second layer is 68. The activation function is tanh, the learning rate is 0.001, Adam is selected as the optimization function, and the batchsize is 512.

Generation of the data set

The data set UKDALE established by British scholars Jack Kelly and William Knottenbelt is used for the experiment (*Kelly & Knottenbelt, 2015*). The load data set contains information about individual loads in five households as well as the total household electricity consumption. And use the load data from July 1, 2013 to July 3, 2013 of five households in the UKDALE dataset as the data set used in this study. The previous 4,070 min of data were used as the training set, and the remaining 250-min data is employed as the testing set. The data in the dataset is sampled every 6 s.

Experimental results

For ease of analysis without loss of generality, the experimental results shown in this article take household 1 as an example. The experimental results are shown below.

Using CNN-LSTM hybrid model decomposition method to decompose the total load data of the training samples to obtain the energy consumption values of each load device, without considering the simultaneous existence of multiple electrical appliances. The load decomposition results of loading equipment for a certain three days in April are shown in Fig. 7. This step of decomposition sets the stage for the next step of prediction.

In Fig. 7, the nine kinds of electrical appliances obtained by decomposition are refrigerators, kettles, home theater computers, microwave ovens, radios, fans, light, active subwoofers, and computers. The power curves of the nine electrical appliances are marked by different colors.

Based on the load decomposition results of nine load devices, nine load prediction models are trained respectively according to the prediction model structure proposed in "Proposed prediction model". Figs. 8–16 show the real values of nine appliances compared to the predicted values, and Fig. 17 shows the total predicted value of the household loads.

From Figs. 8–16, we can see that the results of the true and predicted values of each appliance are very close, and Fig. 17 illustrates the total household load forecasting results. Its total true value is also very close to the predicted power with better results. These results demonstrate that the NILM-based load decomposition approach is effective and the FedDL model manages to accurately forecast the household loads.

To prove that the NILM-based integrated prediction methodology provides a better prediction effect than the traditional methodology that directly predicts the aggregated
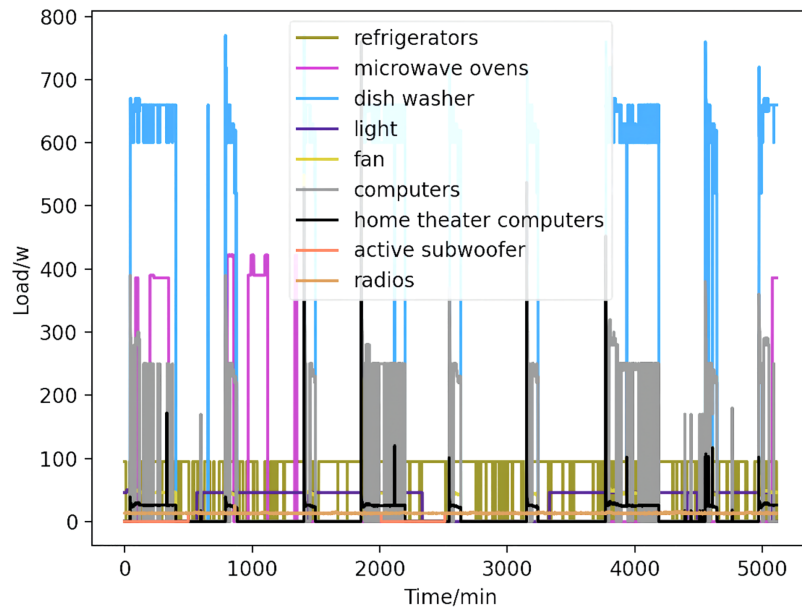
Figure 7 Load curve. The nine kinds of appliances obtained by decomposition (refrigerators, kettles, home theater computers, microwave ovens, radios, fans, light, active subwoofers, and computers) are shown, with power curves delineated by color.
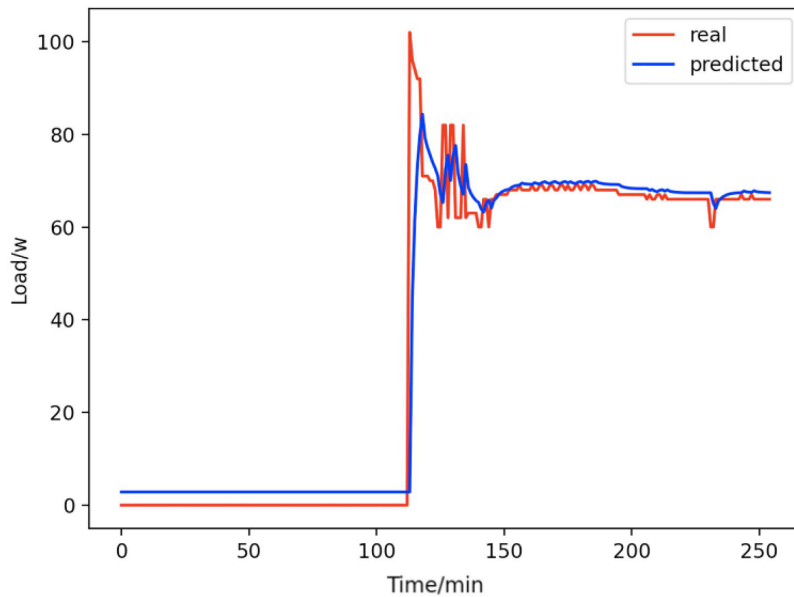
Figure 8 Active subwoofer. Comparison of actual and predicted values of active subwoofers.

signal as a whole, we conducted a comparative experiment based on integrated prediction after load decomposition and overall direct prediction.

Figure 18 shows the comparison between the integrated prediction based on load decomposition and the overall direct prediction. It is observed that both can reflect the electricity consumption trend of household users, but the method based on integrated
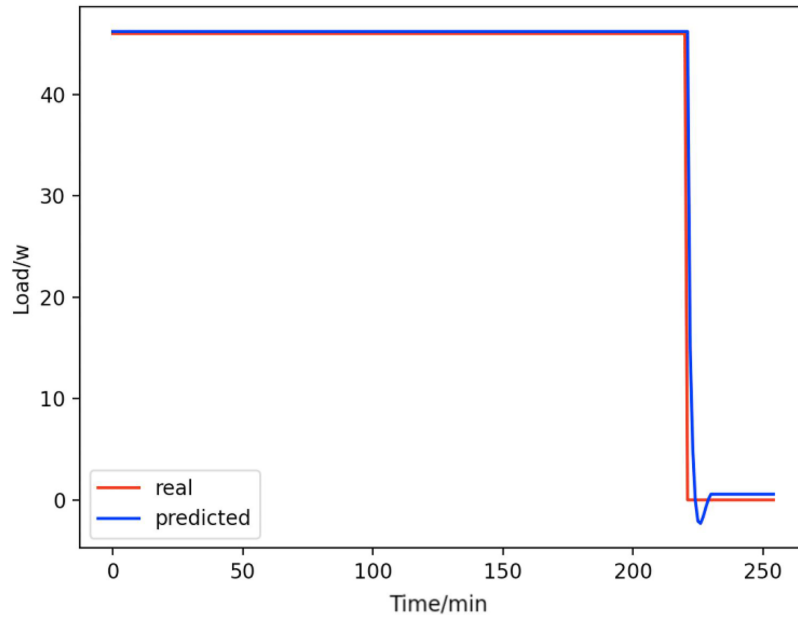
Figure 9 Light. Comparison of actual and predicted values of lights.
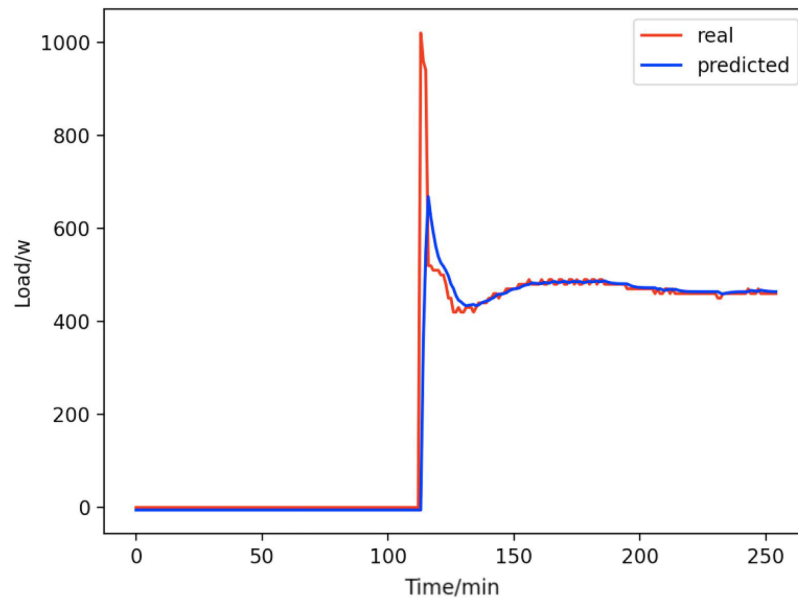
Figure 10 Home theater computers. Comparison of actual and predicted values of home theater computers.

prediction and comprehensive prediction has a better fitting ability to the actual load curve. In contrast to the prediction results based on the overall direct prediction, in the overall equipment, different equipment has different power characteristics. In addition, there may be strong time dependence in the power signal of single equipment, and the overall direct prediction ignores the difference in power characteristics between devices,
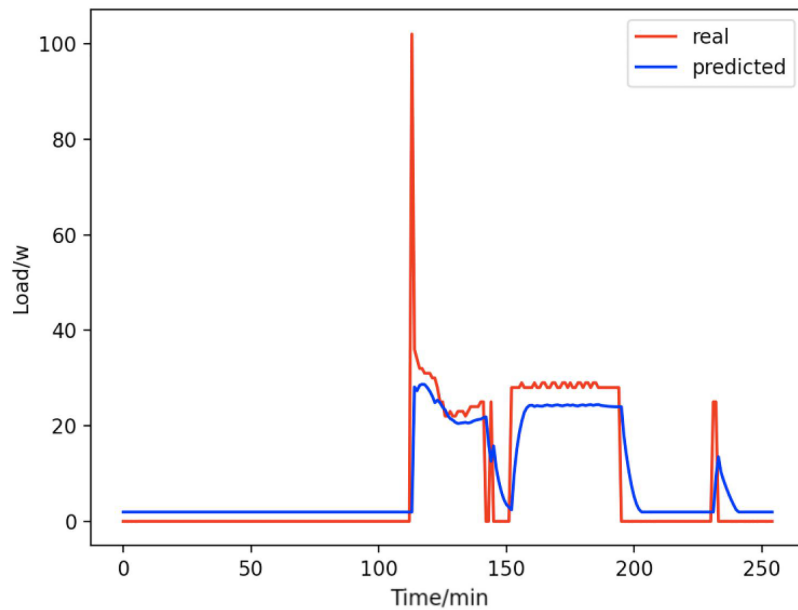
Figure 11 Fans. Comparison of actual and predicted values of fans.
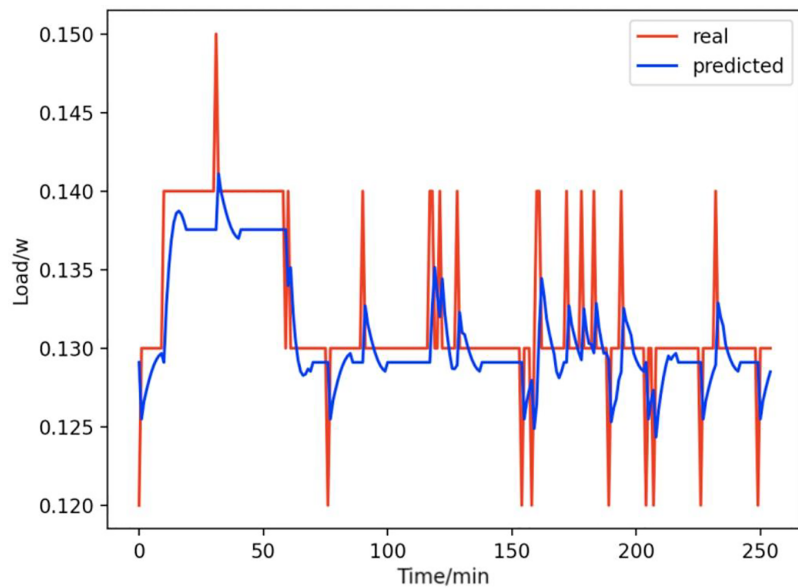
Figure 12 Radios. Comparison of actual and predicted values of radios.

resulting in a large mistake in the overall direct prediction method. However, the load data based on load decomposition can avoid the above problems.

Comparisons with other methods

To reasonably evaluate the performances of the proposed approach, comparisons with other methods have been carried out. For the different models selected in this article,
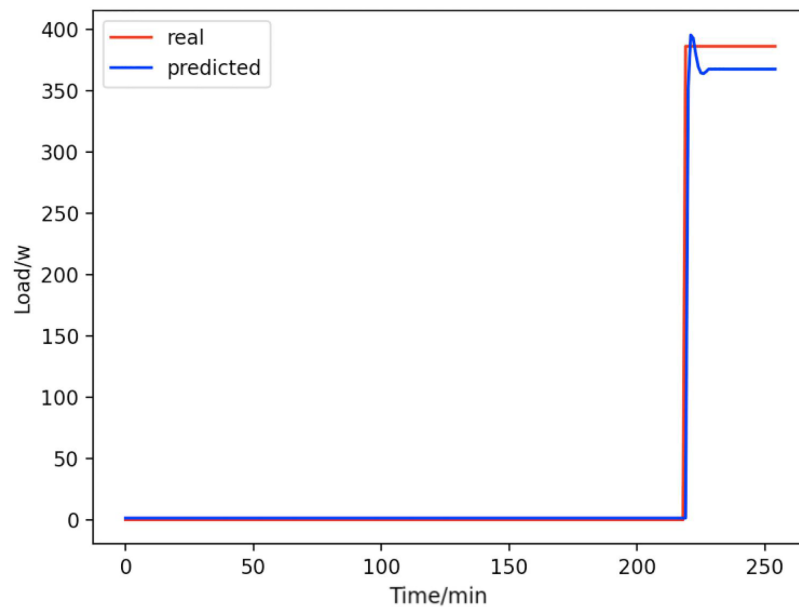
Figure 13 Microwave ovens. Comparison of actual and predicted values of microwave ovens.
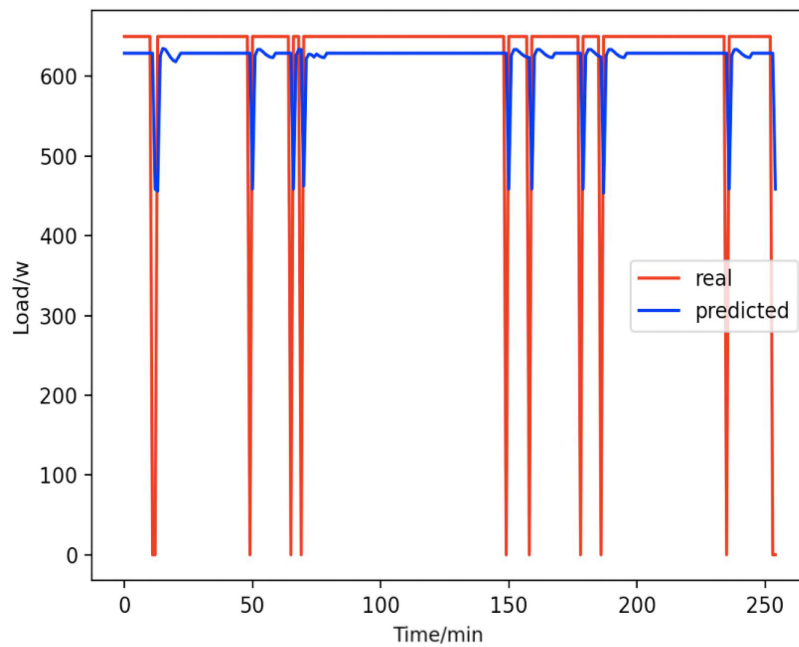
Figure 14 Kettles. Comparison of actual and predicted values of kettles.

related experiments are done on the basis of NILM load decomposition. The same input samples are used for training the model, and the RMSE and MAE are used as the evaluation indexes of the prediction model.
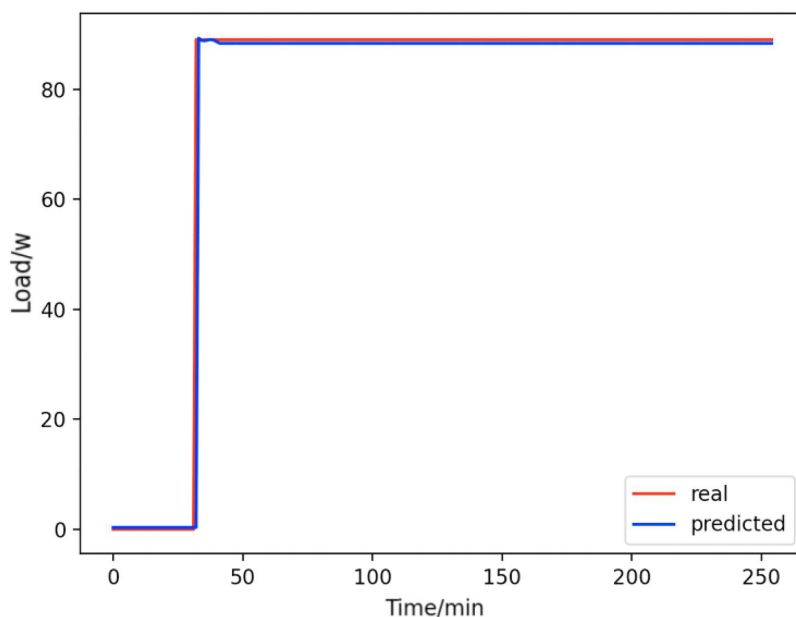
Figure 15 Refrigerators. Comparison of actual and predicted values of refrigerators.
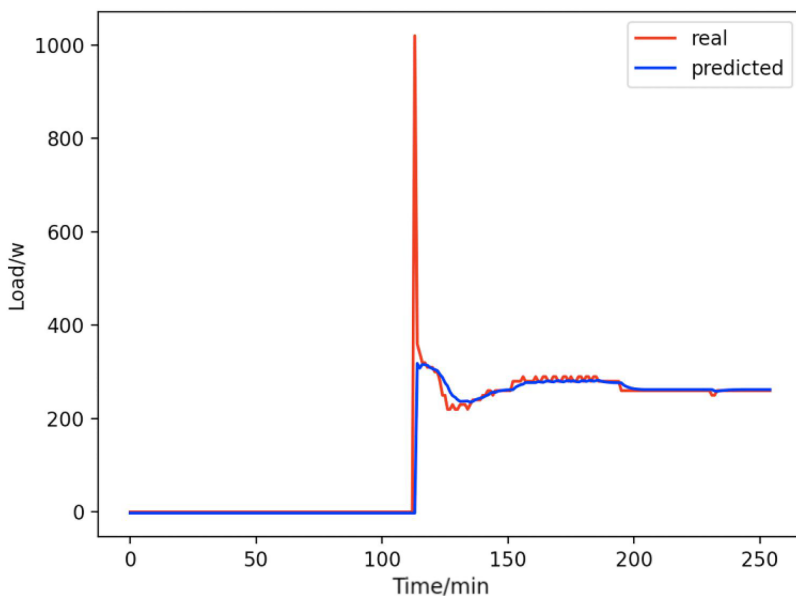
Figure 16 Computers. Comparison of actual and predicted values of computers.

As can be seen from Table 1, the performance of the FedDL is better than that of all other alternative algorithms (besides the BiLSTM-Attention). However, the FedDL provides better privacy protection for co-modeling participants than the BiLSTM-Attention. In this prediction method, under the precondition of guaranteeing the accuracy of the prediction, the data of household users are not leaked.
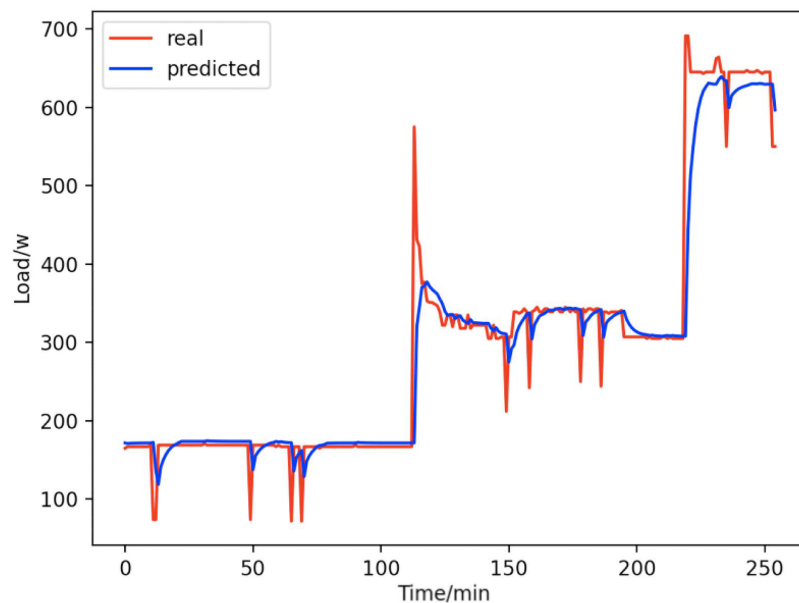
Figure 17 Total forecast. A comparison of the total true value and predicted value of these 9 appliances (refrigerators, kettles, home theater computers, microwave ovens, radios, fans, light, active subwoofers, and computers).
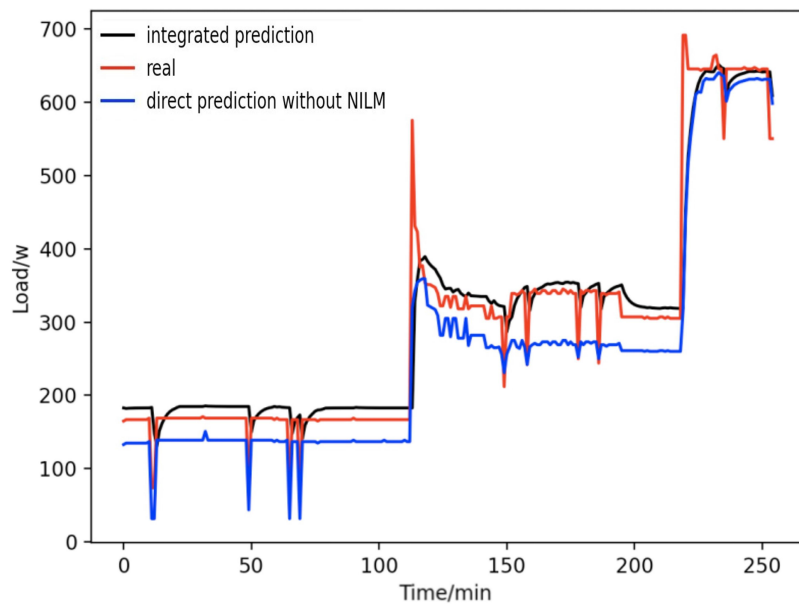
Figure 18 Comparison of overall direct forecast and individual forecast.

To further validate the validity of the proposed model in a more profound way, we also test it on the other four families, where the same amount of appliances and data are selected as household 1 (All the selected models are based on NILM load decomposition). Meanwhile, the first three models in Table 1 are selected for comparative experiments, which are shown in the figure below.
Table 1 Comparison of evaluation indexes.

	MAE	RMSE
Proposed method	0.08141	0.16739
BiLSTM-Attention with NILM	0.07825	0.15956
ANN with NILM	0.28376	0.34675
LSTM with NILM	0.10956	0.18266
FFANN with NILM	0.27869	0.50923
ARIMA with NILM	0.28865	0.55243
SVM with NILM	0.28914	0.52826
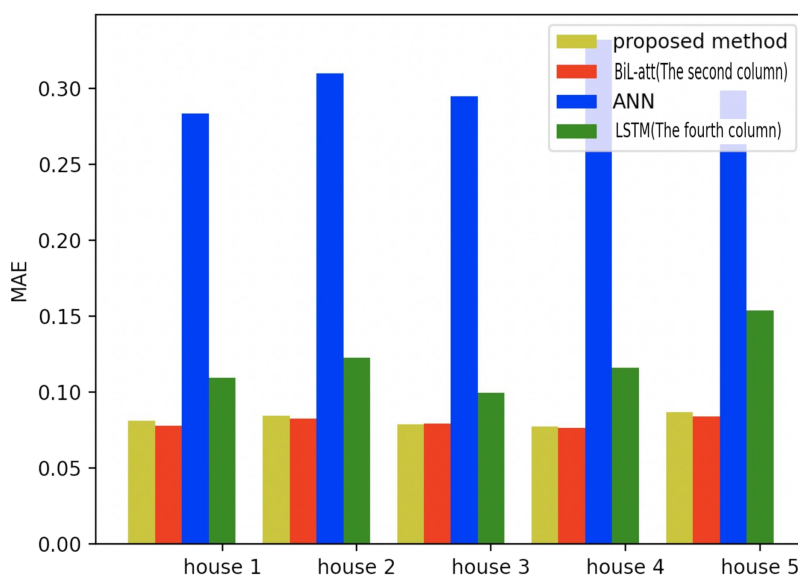
Figure 19 Comparison of five households using MAE using different methods.

Figures 19–21 provide the comparative error analysis of the five households in the UKDALE data set under different models. As for the comparison with the BiLSTM-Attention, the performance of the proposed model is slightly inferior to that of the BiLSTM-Attention, but the FedDL can effectively protect user privacy. In terms of these two indicators, the FedDL outperforms the LSTM and ANN, which further proves the superiority of the proposed approach.

Robustness analysis of federated settings

To evaluate the effect of primary parameters on the robustness of the proposed FedDL prediction model, we design a different experiment with varied local Iteration period E and the client score C that performs calculations during training. Here, the number of local iteration cycles for each round of clients is set to 5, 50, and 80; the scores for the number of clients are set to 50% and 100%. The results are displayed in Fig. 22.
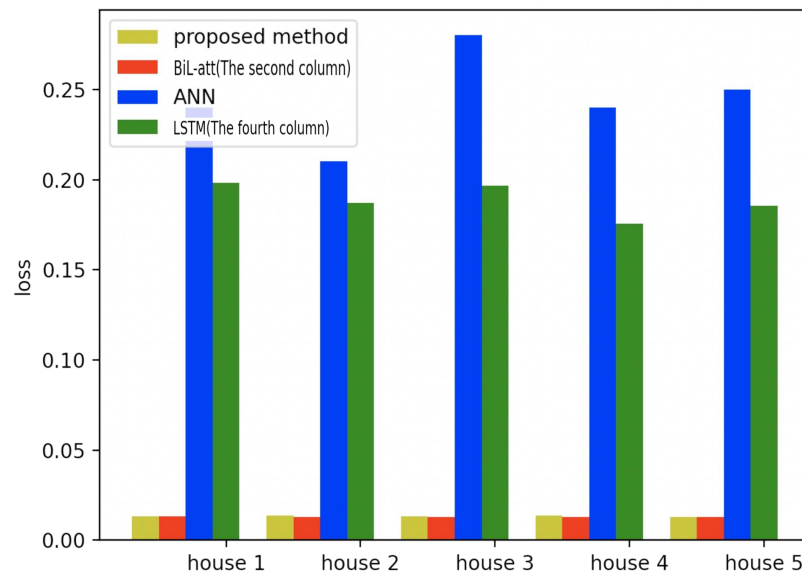
Figure 20 Comparison of five households using loss using different methods.
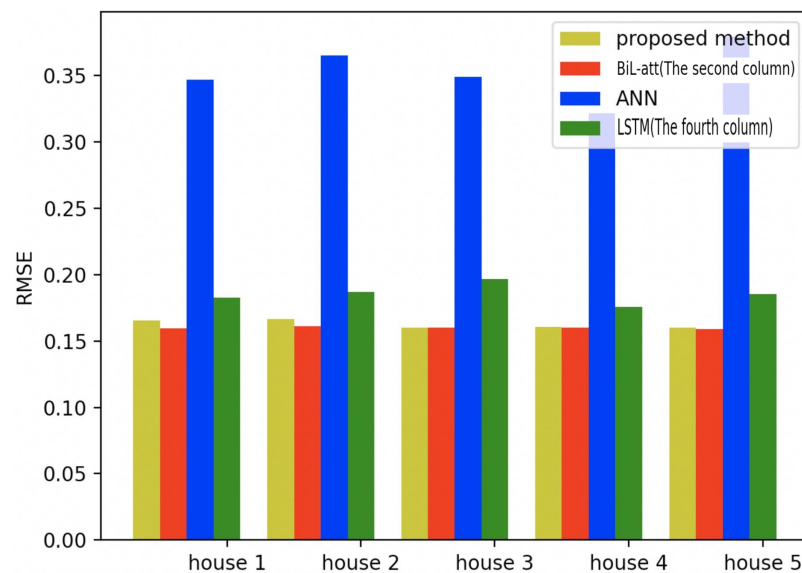
Figure 21 Comparison of five households using RMSE using different methods.
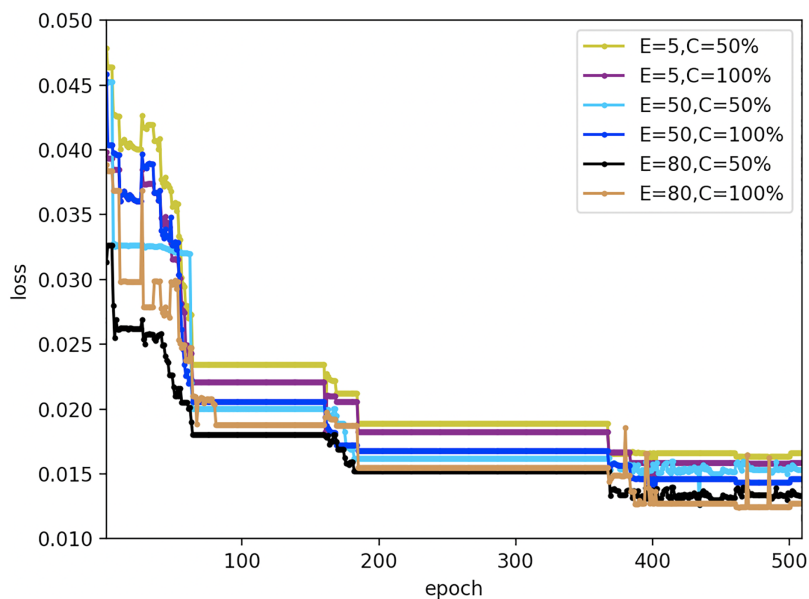
Figure 22 Loss results under different federal settings.

As can be seen from Fig. 22, different federal settings have different effects on FedDL performance. When the local client epochs are fixed, the performance of the model would not change with different local client epochs, which proved that the prediction model has good robustness to different local client epochs. And when the client ratio is fixed, the loss function value of the prediction model reduces slightly with the increase of iterations.

CONCLUSIONS

This article proposes a household load prediction method based on federated deep learning and non-intrusive load monitoring. Based on the simulation results, the following conclusions can be drawn:

(1) By combining federal learning with BiLSTM-Attention, this work develops a household load prediction approach based on non-intrusive load monitoring *via* federated deep learning. In this prediction method, under the precondition of guaranteeing the accuracy of the prediction, the data of household users are not leaked. It can reduce the data security risk and use the electrical power of other households to meet the needs for user privacy protection.

(2) The NILM based on the CNN-LSTM hybrid model is used to decompose the individual load of household appliances from the overall load demand, and the power of each appliance is predicted by the decomposed power of individual appliances, which avoids the strong time dependence of individual equipment. The case results demonstrate that this NILM-based integrated prediction methodology provides a better prediction effect than the traditional methodology that directly predicts the aggregated signal as a whole.

(3) This article uses the data decomposed by the CNN-LSTM hybrid deep learning model to test the proposed federated deep learning prediction model. The load forecasting result

is very close to the real loads. And through comparison with other models, the superiority of the proposed method is verified. In addition, experiments in various federated learning environments are implemented to validate the robustness of this methodology.

In this article, external influence factors, such as weather factors and humidity factors, are not taken into account in the prediction process, but these factors need to be considered in more realistic scenarios. It is interesting to extend the proposed NILM algorithm for supporting the energy management of residential integrated energy systems (*Sun, Deng & Li, 2020*; *Li et al., 2022*). Besides, using automated reinforcement learning for determining the optimal hyper parameters of the used deep learning model (*Li, Wang & Yang, 2022*) is also deserved to be investigated. Another interesting topic is to investigate the NILM issues under cyber-attacks (*Wang & Srikantha, 2021*; *Li, Li & Chen, 2019*; *Iliaee, Liu & Shi, 2021*).

Funding
The authors received no funding for this work.

Competing Interests
The authors declare that they have no competing interests.

Author Contributions
- Xinxin Zhou performed the computation work, authored or reviewed drafts of the article, and approved the final draft.
- Jingru Feng conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Jian Wang performed the experiments, prepared figures and/or tables, and approved the final draft.
- Jianhong Pan conceived and designed the experiments, performed the computation work, authored or reviewed drafts of the article, and approved the final draft.

Data Availability
The code is available in the Supplemental File.
The dataset is available at: https://data.ukedc.rl.ac.uk/browse/edc/efficiency/residential/EnergyConsumption/Domestic/UK-DALE-2015.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.1049#supplemental-information.

REFERENCES

Aly HHH. 2020. A proposed intelligent short-term load forecasting hybrid models of ANN, WNN and KF based on clustering techniques for smart grid. *Electric Power Systems Research* **182(January (1)):**106191 DOI 10.1016/j.epsr.2019.106191.

Bin Y, Yang Y, Shen F, Xie N, Shen HT, Li X. 2018. Describing video with attention-based bidirectional LSTM. *IEEE Transactions on Cybernetics* **49(7):**2631–2641 DOI 10.1109/TCYB.2018.2831447.

Briggs C, Fan Z, Andras P. 2021. Federated learning for short-term residential energy demand forecasting. ArXiv Preprint ArXiv:2105.13325.

Chen BJ, Chang MW. 2004. Load forecasting using support vector machines: a study on EUNITE competition 2001. *IEEE Transactions on Power Systems* **19(4):**1821–1830 DOI 10.1109/TPWRS.2004.835679.

Cui M, Wang J, Yue M. 2019. Machine learning-based anomaly detection for load forecasting under cyberattacks. *IEEE Transactions on Smart Grid* **10(5):**5724–5734 DOI 10.1109/TSG.2018.2890809.

Dinesh C, Makonin S, Bajić IV. 2020. Residential power forecasting based on affinity aggregation spectral clustering. *IEEE Access* **8:**99431–99444 DOI 10.1109/ACCESS.2020.2997942.

Fang Z, Zhao D, Chen C, Li Y, Tian Y. 2020. Nonintrusive appliance identification with appliance-specific networks. *IEEE Transactions on Industry Applications* **56(4):**3443–3452 DOI 10.1109/TIA.2020.2994279.

Gasparin A, Lukovic S, Alippi C. 2022. Deep learning for time series forecasting: the electric load case. *CAAI Transactions on Intelligence Technology* **7(1):**1–25 DOI 10.1049/cit2.12060.

Guan C, Luh PB, Michel LD, Chi Z. 2013. Hybrid Kalman filters for very short-term load forecasting and prediction interval estimation. *IEEE Transactions on Power Systems* **28(4):**3806–3817 DOI 10.1109/TPWRS.2013.2264488.

Hou Y, Jia S, Lun X, Zhang S, Chen T, Wang F, Lv J. 2020. GCNs-net: a graph convolutional neural network approach for decoding time-resolved EEG motor imagery signals. arXiv preprint arXiv:2006.08924.

Huong TT, Bac TP, Long DM, Luong TD, Dan NM, Quang LA, Cong LT, Thang BD, Tran KP. 2021. Detecting cyberattacks using anomaly detection in industrial control systems: a federated learning approach. *Computers In Industry* **132(7):**103509 DOI 10.1016/j.compind.2021.103509.

Iliaee N, Liu S, Shi W. 2021. Non-Intrusive load monitoring based demand prediction for smart meter attack detection. In: *2021 International Conference on Control, Automation and Information Sciences (ICCAIS)*. Xi'an, China: IEEE, 370–374.

Ji Z, Xiong K, Pang Y, Li X. 2019. Video summarization with attention-based encoder-decoder networks. *IEEE Transactions on Circuits and Systems for Video Technology* **30(6):**1709–1717 DOI 10.1109/TCSVT.2019.2904996.

Kelly J, Knottenbelt W. 2015. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific Data* **2(1):**1–14 DOI 10.1038/sdata.2015.7.

Kumar S, Mishra S, Gupta S. 2016. Short term load forecasting using ANN and multiple linear regression. In: *2016 Second International Conference on Computational Intelligence & Communication Technology (CICT)*. Ghaziabad: IEEE, 184–186.

Kuster C, Rezgui Y, Mourshed M. 2017. Electrical load forecasting models: a critical systematic review. *Sustainable Cities and Society* **35(11):**257–270 DOI 10.1016/j.scs.2017.08.009.

Lawrence S, Giles CL, Tsoi AC, Back AD. 1997. Face recognition: a convolutional neural-network approach. *IEEE Transactions on Neural Networks* **8(1)**:98–113 DOI 10.1109/72.554195.

Lee CM, Ko CN. 2011. Short-term load forecasting using lifting scheme and ARIMA models. *Expert Systems with Applications* **38(5)**:5902–5911 DOI 10.1016/j.eswa.2010.11.033.

Li L, Fan Y, Tse M, Lin K-Y. 2020. A review of applications in federated learning. *Computers & Industrial Engineering* **149(5)**:106854 DOI 10.1016/j.cie.2020.106854.

Li Y, Gu X. 2013. Application of online SVR in very short-term load forecasting. *International Review of Electrical Engineering (IREE)* **8(1)**:277–282.

Li Y, Li Z, Chen L. 2019. Dynamic state estimation of generators under cyber attacks. *IEEE Access* **2019(7)**:125253–125267 DOI 10.1109/ACCESS.2019.2939055.

Li Y, Li J, Wang Y. 2022. Privacy-preserving spatiotemporal scenario generation of renewable energies: a federated deep generative learning approach. *IEEE Transactions on Industrial Informatics* **18(4)**:2310–2320 DOI 10.1109/TII.2021.3098259.

Li Y, Li K, Yang Z, Yu Y, Xu R, Yang M. 2022a. Stochastic optimal scheduling of demand response-enabled microgrids with renewable generations: An analytical-heuristic approach. *Journal of Cleaner Production* **330(2)**:129840 DOI 10.1016/j.jclepro.2021.129840.

Li L, Meinrenken CJ, Modi V, Culligan PJ. 2021. Short-term apartment-level load forecasting using a modified neural network with selected auto-regressive features[J]. *Applied Energy* **287(147)**:116509 DOI 10.1016/j.apenergy.2021.116509.

Li Y, Wang R, Yang Z. 2022. Optimal scheduling of isolated microgrids using automated reinforcement learning-based multi-period forecasting. *IEEE Transactions on Sustainable Energy* **13(1)**:159–169 DOI 10.1109/TSTE.2021.3105529.

Li Y, Wang B, Yang Z, Li J, Chen C. 2022. Hierarchical stochastic scheduling of multi-community integrated energy systems in uncertain environments via Stackelberg game. *Applied Energy* **308(3)**:118392 DOI 10.1016/j.apenergy.2021.118392.

Li Y, Wang B, Yang Z, Li J, Li G. 2022b. Optimal scheduling of integrated demand response-enabled community-integrated energy systems in uncertain environments. *IEEE Transactions on Industry Applications* **58(2)**:2640–2651 DOI 10.1109/TIA.2021.3106573.

Li Y, Zhang M, Chen C. 2022. A deep-learning intelligent system incorporating data augmentation for short-term voltage stability assessment of power systems. *Applied Energy* **308(3)**:118347 DOI 10.1016/j.apenergy.2021.118347.

Lindemann B, Maschler B, Sahlab N, Weyrich M. 2021. A survey on anomaly detection for technical systems using LSTM networks. *Computers In Industry* **131(3)**:103498 DOI 10.1016/j.compind.2021.103498.

Liu G, Guo J. 2019. Bidirectional LSTM with attention mechanism and convolutional layer for text classification. *Neurocomputing* **337(5)**:325–338 DOI 10.1016/j.neucom.2019.01.078.

Liu S, Lee K, Lee I. 2020. Document-level multi-topic sentiment classification of email data with bilstm and data augmentation. *Knowledge-Based Systems* **197(1)**:105918 DOI 10.1016/j.knosys.2020.105918.

Liu Q, Shen Y, Wu L, Li J, Zhuang L, Wang S. 2018. A hybrid FCW-EMD and KF-BA-SVM based model for short-term load forecasting. *CSEE Journal of Power and Energy Systems* **4(2)**:226–237 DOI 10.17775/CSEEJPES.2016.00080.

Lu Y, Huang X, Dai Y, Maharjan S, Zhang Y. 2019. Blockchain and federated learning for privacy-preserved data sharing in industrial IoT. *IEEE Transactions on Industrial Informatics* **16(6)**:4177–4186 DOI 10.1109/TII.2019.2942190.

López G, Custodio V, Moreno JI, Sikora M, Moura P, Fernández N. 2015. Modeling smart grid neighborhoods with the ENERsip ontology. *Computers In Industry* **70(8)**:168–182 DOI 10.1016/j.compind.2015.01.008.

Nazih AS, Fawwaz E, Osama MA. 2011. Medium-term electric load forecasting using multivariable linear and non-linear regression. *Smart Grid and Renewable Energy* **2(2)**:126–135 DOI 10.4236/sgre.2011.22015.

Ponoćko J, Milanović JV. 2018. Forecasting demand flexibility of aggregated residential load using smart meter data. *IEEE Transactions on Power Systems* **33(5)**:5446–5455 DOI 10.1109/TPWRS.2018.2799903.

Sadaei HJ, de Lima e Silva PC, Guimarães FG, Lee MH. 2019. Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series. *Energy* **175(12)**:365–377 DOI 10.1016/j.energy.2019.03.081.

Sharma S, Majumdar A, Elvira V, Chouzenoux E. 2020. Blind Kalman filtering for short-term load forecasting. *IEEE Transactions on Power Systems* **35(6)**:4916–4919 DOI 10.1109/TPWRS.2020.3018623.

Shi Z, Li Y, Yu T. 2009. Short-term load forecasting based on LS-SVM optimized by bacterial colony chemotaxis algorithm. In: *International Conference on Information and Multimedia Technology*Jeju Island, South Korea.

Shi Z-B, Yu T, Zhao Q, Li Y, Lan Y-B. 2008. Comparison of algorithms for an electronic nose in identifying liquors. *Journal of Bionic Engineering* **5(3)**:253–257 DOI 10.1016/S1672-6529(08)60032-3.

Stephen B, Tang X, Harvey PR, Galloway S, Jennett KI. 2015. Incorporating practice theory in sub-profile models for short term aggregated residential load forecasting. *IEEE Transactions on Smart Grid* **8(4)**:1591–1598 DOI 10.1109/TSG.2015.2493205.

Sun J, Deng J, Li Y. 2020. Indicator & crowding distance-based evolutionary algorithm for combined heat and power economic emission dispatch. *Applied Soft Computing* **90(2)**:106158 DOI 10.1016/j.asoc.2020.106158.

Wang Y, Bennani IL, Liu X, Sun M, Zhou Y. 2021. Electricity consumer characteristics identification: a federated learning approach. *IEEE Transactions on Smart Grid* **12(4)**:3637–3647 DOI 10.1109/TSG.2021.3066577.

Wang Y, Chen Q, Sun M, Kang C, Xia Q. 2018. An ensemble forecasting method for the aggregated load with subprofiles. *IEEE Transactions on Smart Grid* **9(4)**:3906–3908 DOI 10.1109/TSG.2018.2807985.

Wang J, Jia R, Zhao W, Wu J, Dong Y. 2012. Application of the largest Lyapunov exponent and non-linear fractal extrapolation algorithm to short-term load forecasting. *Chaos, Solitons & Fractals* **45(9–10)**:1277–1287 DOI 10.1016/j.chaos.2012.06.009.

Wang J, Srikantha P. 2021. Stealthy black-box attacks on deep learning non-intrusive load monitoring models. *IEEE Transactions on Smart Grid* **12(4)**:3479–3492 DOI 10.1109/TSG.2021.3062722.

Wang Y, Xia Q, Kang C. 2010. Secondary forecasting based on deviation analysis for short-term load forecasting. *IEEE Transactions on Power Systems* **26(2)**:500–507 DOI 10.1109/TPWRS.2010.2052638.

Wu Z, Zhou S, Li J, Zhang X-P. 2014. Real-time scheduling of residential appliances via conditional risk-at-value. *IEEE Transactions on Smart Grid* **5(3)**:1282–1291 DOI 10.1109/TSG.2014.2304961.

Yang Z, Chen M, Saad W, Hong CS, Shikh-Bahaei M. 2020. Energy efficient federated learning over wireless communication networks. *IEEE Transactions on Wireless Communications* **20(3)**:1935–1949 DOI 10.1109/TWC.2020.3037554.

Yildiz B, Bilbao JI, Sproul AB. 2017. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. *Renewable and Sustainable Energy Reviews* **73(19)**:1104–1122 DOI 10.1016/j.rser.2017.02.023.

Zeng J, Qiao W. 2013. Short-term solar power prediction using a support vector machine. *Renewable Energy* **52(3)**:118–127 DOI 10.1016/j.renene.2012.10.009.

Zhang M, Li J, Li Y, Xu R. 2021. Deep learning for short-term voltage stability assessment of power systems. *IEEE Access* **2021(9)**:29711–29718 DOI 10.1109/ACCESS.2021.3057659.

Zheng Z, Chen H, Luo X. 2019. A Kalman filter-based bottom-up approach for household short-term load forecast. *Applied Energy* **250**:882–894 DOI 10.1016/j.apenergy.2019.05.102.

Zhou X, Feng J, Li Y. 2021. Non-intrusive load decomposition based on CNN-LSTM hybrid deep learning model. *Energy Reports* **7(6)**:5762–5771 DOI 10.1016/j.egyr.2021.09.001.

Ziel F, Liu B. 2016. Lasso estimation for GEFCom2014 probabilistic electric load forecasting. *International Journal of Forecasting* **32(3)**:1029–1037 DOI 10.1016/j.ijforecast.2016.01.001.