IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 18, 2021, accepted June 5, 2021, date of publication June 9, 2021, date of current version June 17, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3087696

Non-Intrusive Load Monitoring Using a CNN-LSTM-RF Model Considering Label Correlation and Class-Imbalance

XIAO ZHOU[1,3], SHUJIAN LI[2], CHENGXI LIU[3], (Senior Member, IEEE), HAOJUN ZHU[1], NAN DONG[1], AND TIANYING XIAO[1]
[1]Energy Development Research Institute, CSG, Guangzhou 510000, China
[2]State Key Laboratory of Advanced Electromagnetic Engineering and Technology, Huazhong University of Science and Technology, Wuhan 430000, China
[3]Department of Electrical Engineering and Automation, Wuhan University, Wuhan 430000, China

Correspondin g author: Chengxi Liu (liuchengxi@whu.edu.cn)

ABSTRACT Non-Intrusive Load Monitoring (NILM) is particularly important for demand response. This paper proposes an innovative method based on a deep learning model to recognize the working state of electrical appliances using low frequency load data. The approach includes a data processing step, a deep learning model and a new accuracy calculation method. The data processing step consists of a multi-feature and high-dimensional method (MFHDM) and a pre-training process. The deep learning model consists of a convolutional neural network (CNN), a long-term short-term memory network (LSTM) and a random-forest (RF) algorithm. The proposed method addresses the label correlation problem and the class-imbalance problem. To test the proposed method, the Reference Energy Disaggregation Dataset (REDD) and the Pecan Street dataset (PSD) are used. A comparative analysis with several models shows that the proposed method can effectively improve electrical appliance recognition accuracy and realize NILM.

INDEX TERMS NILM, CNN, LSTM, RF, multi-label classification, class-imbalance.

I. INTRODUCTION

Today, the application of Non-Intrusive Load Monitoring (NILM) is particularly significant because of its advances in communications technologies and artificial intelligence. With the help of NILM, demand response (DR) schemes can be implemented more efficiently, especially for household consumers. Moreover, consumers can have a more comprehensive understanding of their electricity behaviors and detailed bills, which can help them to develop an energy-aware behavior. With these insights, a reduction in operating costs from grid operators and electricity costs for ending consumers can be achieved.

The methods developed to achieve NILM can be divided into two categories based on the required data. The first category is based on a transient analysis of high-frequency sampling signals. More specifically, these methods identify electrical behaviors through improved measurement methods [1] or feature extracting methods [2]–[4]. Reviews on this category are available in [5]–[7].

The associate editor coordinating the review of this manuscript and approving it for publication was Liangtian Wan.

The other category of NILM methods uses single and low-frequency measurement data (such as active power and voltage). According to the purpose of the analysis, these studies can be further divided into load disaggregation and electrical appliances ON/OFF state analysis, which correspond to a regression problem and a classification problem. The former aims to isolate from the total load the power values of different appliances from the total load, such as the study by Ram [8], which proposes a modified cross-entropy method for classification of events in NILM systems. Ref. [9] compares the effects of three deep neural networks for load disaggregation, which can provide more information but has an unsatisfactory effect of transfer learning. The latter can be viewed as the multi-label classification (MLC) recognition problem. In this approach, the ON state for electrical appliance is marked as 1 and the OFF state, as 0. Many methods have been proposed to solve the MLC recognition problem.

Studies based on deep learning models for classification tasks have shown good performance for NILM. Nevertheless, two important characteristics of NILM should be taken into account, label correlation and class imbalance. In terms of label correlation, there are correlations between the uses of different appliances, for example, washers and dryers are often used in sequence. In addition, electricity consumption behaviors are highly correlated with time, for example, lighting loads always appear in the evening and photovoltaics (PV) can only output in the daytime. All of the above are common sense to us, but the models know nothing about these relationships. It is meaningful to train the model to learn these relationships and characteristics. For class imbalance, some electrical appliances either have a low percentage in use, such as washers, or a high percentage in use, such as refrigerators. For these electrical appliances, the model can fail due to the gradient vanishing problem, which occurs because constant state recognition results can reach a high precision score.

Considering the above characteristics of NILM, this paper proposes a new method based on a deep learning network, which includes a data processing step, a deep learning network model and a new accuracy calculation method. The proposed method considers the characteristics of the household load and the logic of electricity consumption behaviors. Simulation results show the method can effectively recognize electrical consumption behaviors with minimal data requirements.

II. RELATED WORK

Deep learning has made considerable advancements in MLC for computer vision. The application of deep learning for NILM generally achieves similar success. The application of deep learning techniques to the NILM problem was first introduced by Kelly [9]. He compares the ability of the CNN, LSTM and Stacked Denoising Autoencoders models in NILM. More studies utilizing neural networks and deep learning to solve this problem have followed [11]–[13]. Ref. [14] proposes a pinball quantile loss function to guide the deep neural network for NILM. Particle swarm optimization is introduced for parameters optimization of training algorithms in artificial neural networks by [15]. Studies [16]–[18] propose more advanced deep learning models to achieve NILM, these form the state-of-the-art in this area of research. Comparing with studies on load disaggregation, [19] provides electrical appliances ON/OFF state analysis by deep learning, and it has a good transfer learning effect.

LSTM networks are widely applied in time series analyses because of their ability to handle long time dependencies. Therefore, LSTM networks have a comparative advantage in detecting load data changes when applied to NILM problems. Relevant studies have been carried out [20]–[22]. Particularly, Jihyun proposes an LSTM model and additionally a novel signature to improve classification performance [20]. Mauch uses a generic two-layer bidirectional LSTM architecture [21]. Kaselimi proposes a Bayesian optimized bidirectional LSTM regression model for NILM [22].

The CNN is good at extracting features of two-dimensional (2D) data and is widely used in many applications, especially in the fields of image and video recognition. However, because load data are one-dimensional (1-D) time series data, it is difficult to directly apply CNNs in load disaggregation. However, many studies propose applying CNNs to achieve NILM. Lan proposes a method to convert electric current data into a grayscale image with a resolution of 64×64, and then designs a CNN model to identify the load [23]. Yang proposes an imaging rule to convert current waveforms to greyscale image and constructs a CNN architecture premised upon VGG-16 to tackle the issue of NILM [24]. Medeiros transforms the values of active and reactive power into a matrix form for CNN training, and reaches a high accuracy in NILM [25]. Fabrizio proposes a new CNN-based system for NILM applications that achieves encouraging results on the public-use dataset BLUED [11].

Apart from deep learning-based methods for NILM, there have been many studies focused on realizing NILM by machine learning methods. Support Vector Machines (SVM) [26], Decision Trees (DT) [27], combinatorial optimization (CO) [28] and Random Forest (RF) [29] have been used to solve event-based NILM. The Hidden Markov Model (HMM) and its modification have been designed to solve state-based NILM. [30]–[32].

Compared to the above approaches, this paper presents a novel method based on a CNN-LSTM-RF model approach for NILM. Three innovative aspects have been introduced to address challenges in NILM. First, this paper proposes a multi-feature and high-dimensional method (MFHDM) to directly extract features from low-frequency measurement data. With this method, the proposed model can learn and utilize expert knowledge efficiently. Second, the label correlation and class-imbalance problems are considered in this study. A compound re-weighting method is proposed to solve the problem of class-imbalance problem. The label correlation problem is solved by pre-training, and transforming the MLC problems into sequence generation problems. Finally, a deep learning model with a CNN-LSTM-RF structure is designed to solve the NILM problem.

III. METHODOLOGY

The NILM method proposed in this paper is composed of three components: a data preprocessing step, a deep learning model and an accuracy calculation. For the data preprocessing step, the MFHDM is used to extract features and a pre-training process is adopted to elicit correlations in electrical consumption behaviors. For the deep learning model, a deep learning network based on a CNN, an LSTM network and an RF is constructed. We select a CNN because it is good at extracting data features and an LSTM network because it can address the long-term dependence of time series data. The RF is introduced to decode and output results. For the accuracy calculation, a new method is developed that trains and evaluates the model. Furthermore, a label relevance analysis is used to address the low recognition performance when low-frequency or unobvious features are present in the electricity behavior. The structure of our proposed model is shown in Fig. 1.

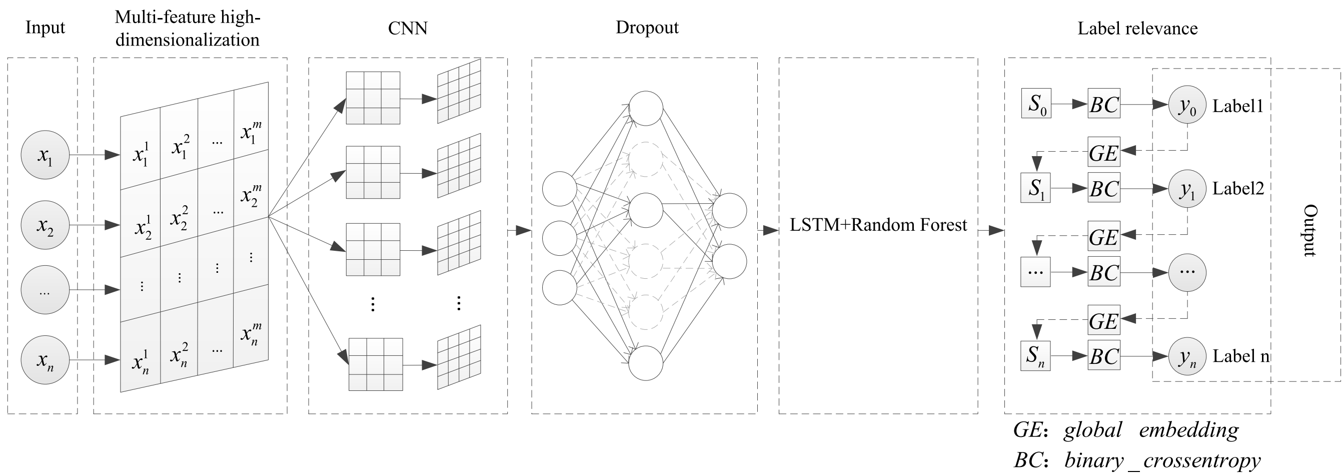

FIGURE 1. The structure of proposed model.

A. MULTI-FEATURE AND HIGH-DIMENSIONAL METHOD FOR LOAD DATA

Low-frequency power data are a one-dimensional time series sequence. Many methods have been proposed to extract information from this type of time series data. These methods, which we call expert knowledge, can improve the accuracy of a model if it can be learned.

An MFHDM is proposed to transform the 1-D load data into 2-D matrix data. Specifically, the multi-feature method uses a mathematical transformation s to extract features, specifically, the mean, variance, and distribution parameters of the load sequence are calculated under different time period lengths. The high-dimensional method isolates the load data according to different time periods. Because the parameters are selected according to the operating characteristics of household appliances, the MFHDM combines a deep learning model with expert knowledge. A schematic diagram of load curve processing using MFHDM is shown in Fig. 2.

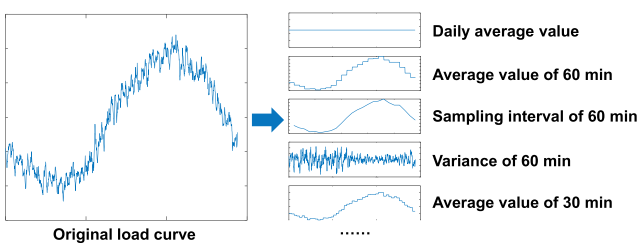

FIGURE 2. Schematic diagram of multi-feature and high-dimensional approach for load data.

Let X(i) be the measured aggregate load over all appliances, where i represents the i th number in the data set. The time interval between two adjacent load data points is the sampling interval.

The formula of the MFHDM is shown in (1), where T is the period length and subscript j corresponds to data in T. Subscripts mean, var, diff, and sum represent the mean, variance, difference, and sum of the data, respectively. In our model, $T = [1, 5, 30, 60, 1440]$, that is 1 minute, 5 minutes, half an hour, an hour and a day in sequence corresponding to different electrical operation characteristics.

$$\mathrm{X}_{DBSCAN,Tj} = \begin{bmatrix} X_{mean} \\ X_{var} \\ X_{diff} \\ X_{sum} \end{bmatrix}_{Tj} = \begin{cases} \begin{bmatrix} X_i \\ 0 \\ 0 \\ X_i \end{bmatrix}, & i < T_j \\ \begin{bmatrix} \sum_{i-T_j}^{i} \frac{X_i}{T_j} \\ \sum_{i-T_j}^{i} \frac{(X_i - X_{mean})^2}{T_j} \\ X_i - X_{i-T_j} \\ \sum_{i-T}^{i} X_i \end{bmatrix}, & i \geq T_j \end{cases} \tag{1}$$

B. PRE-TRAINING

Pre-training is adopted to address the label correlation problem. Because the superposition of loads will override the load characteristics of many of the appliances, considering label correlation can reduce the covering of individual load characteristics. Pre-training can determine the sequence of appliances. Assume that there are L electrical appliances. From the perspective of sequence generation, the MLC task can be modeled as finding an optimal label sequence y that maximizes the conditional probability p(y|x), which is calculated as follows:

$$p(y|\,x) = \prod_{i=1}^{L} p(\,y_i|\,y_1, y_2, \cdots y_{i-1}, x) \tag{2}$$

where x is the output of the upper network. Labels that have obvious load characteristics are placed in the front. Thus, label correlation is considered sequentially in this paper. In other words, electrical appliances with obvious load characteristics will be recognized first. A diagram of this process is shown in Fig. 3. Then, the recognition is taken as the input to recognize the state of other electrical appliances, to reduce the override of load characteristics by appliances with obvious load characteristics. Pre-training recognizes the working state of appliances individually, the detection sequence in the model is determined based on independent recognition accuracy.

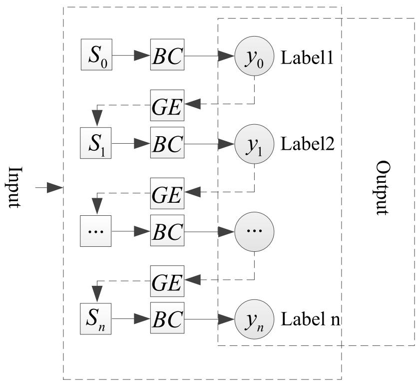

FIGURE 3. **A diagram of the procedure for transforming the MLC problem into a sequence generation problem.**

Pre-training is also adopted to consider correlations in electrical consumption behaviors and to improve the accuracy of NILM. In particular, the density-based spatial clustering of applications with noise (DBSCAN) is applied to classify the daily load curve in pre-training process. DBSCAN does not need to set the number of clusters beforehand and has superior performance in anomaly detection. A classification schematic diagram of each point of DBSCAN is shown in Fig. 4.

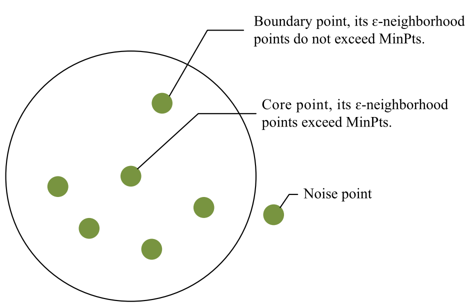

FIGURE 4. **A schematic diagram of DBSCAN points.**

DBSCAN looks for clusters by calculating the ε-neighborhood of each point in the data set. If the ε-neighborhood of a point P contains M points, a new cluster with P as the core point will be created, otherwise P is marked as a noise point. Then, DBSCAN repeatedly looks for points that are directly density-accessible to these core points, a process that may involve merging density-accessible clusters. The process ends when no new points can be added to any cluster. The parameters ε and M of the algorithm are given according to prior knowledge.

C. ACCURACY CALCULATION METHOD

A new accuracy method is developed to evaluate the recognition accuracy of residential NILM and to address the class-imbalance problem. The accuracy calculation method comprises two steps. The first step uses F1 score (5) to calculate the accuracy of appliance recognition instead of the traditional accuracy (6). The harmonic average algorithm [35] is used to calculate precision and recall, which are shown in (3) and (4), respectively.

$$precision = \frac{TP}{FP + TP} \tag{3}$$

$$recall = \frac{TP}{FN + TP} \tag{4}$$

where TP is the correct positive prediction count, FP is the incorrect positive prediction count, and FN is the incorrect negative prediction count. The F1 score is used to evaluate the accuracy of the model, which is shown in (5)

$$F1\_score = \frac{2}{\left[\left({}^{1}/_{precision}\right) + \left({}^{1}/_{recall}\right)\right]} \tag{5}$$

$$t\_accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

where TN is the correct negative prediction count.

The second step is to re-weight values between labels, where label weights are set according to the recognition sequence of the electrical appliances. More specifically, the higher the accuracy in pre-training, the lower the weight value. The calculation method for model accuracy is shown in (7).

$$\begin{cases} accuracy = \sum_{i=1}^{n} \beta_i \cdot f1\_score(i) \\ s.t. \sum \beta_i = 1 \end{cases} \tag{7}$$

where f1_score(i) represents F1 score of appliance i and β_i is the accuracy coefficient of electrical appliance i, which is calculated according to the recognition accuracy in pre-training.

D. CNN-LSTM-RF MODEL

The deep learning model consists of a CNN, an LSTM network and an RF. CNNs have shown excellent performance in tasks such as image identification and natural language processing. A significant advantage of CNNs is that they can mine information from high-dimensional data such as that generated by MFHDM. At the same time, compared with Recurrent Neural Networks (RNNs), CNNs can greatly reduce the computational burden. The LSTM is an improved network that is structure based on the RNN. This structure maintains the advantages of an RNN while addressing the gradient disappearance problem that often occurs during RNN model training. More importantly, the LSTM network can remember the previous state for short or long period of time. A typical LSTM network structure includes input gates, forget gates, and output gates. This unique structure eliminates the need for layer-by-layer conversion of information during the propagation process. Part of the information can be directly transmitted to the subsequent network layer through the "gate" structure. This structure ensures effective information transmission, ensures that there is no gradient disappearance no matter the network depth, and has better convergence.

The output from the CNN-LSTM network needs to be decoded to determine the recognition running state of various household appliances. The RF, a combinatorial classifier, is used to decode the outputs of the LSTM and to determine the classification results. The RF determines the most appropriate threshold value as the output of the LSTM network is a probability value between 0 and 1. More importantly, the RF model determines the threshold value for different times. The detailed structure and parameters of the model are given below.

1) The daily load data are collected at 1 minute intervals and contains 1440 data points. After data processing with the MFHDM, the 11440 time series data are transformed to a 51440 matrix.

2) Through the pre-training process, the daily load curves are clustered, and each daily load curves is labeled with a different label and provided as an input for the CNN. Therefore, each input to the CNN is a 61440 matrix.

3) The input data are first processed through the *convolution* layers of the CNN. In this work, we applied two convolutional layers, each with a different number of filters (128 in the first convolution layer, 32 in the second).

4) A dropout layer can be inserted between each layer to improve the robustness of the model.

5) The output from the CNN is passed to the LSTM network. In this study, the LSTM has 128 hidden layers. The mathematical expression for the LSTM is

$$\begin{aligned} Y_t &= H_t W_{hq} + b_q \\ H_t &= O_t \odot \tanh(C_t) \\ C_t &= F_t \odot C_{t-1} + I_t \odot \overline{C}_t \qquad (8) \\ I_t &= \sigma\left(X_t W_{xi} + H_{t-1} W_{hi} + b_i\right) \\ F_t &= \sigma\left(X_t W_{xf} + H_{t-1} W_{hf} + b_f\right) \\ O_t &= \sigma\left(X_t W_{xo} + H_{t-1} W_{ho} + b_o\right) \qquad (9) \end{aligned}$$

where I_t, F_t, O_t are the outputs of input gate, forget gate and output gate of LSTM respectively, W is the weight coefficient, b is the bias parameter, σ is activation function, H is the hidden state of the previous time step. The LSTM activation function is sigmoid, which makes the output bounded between 0 and 1. The expression for the sigmoid function is

$$s(x) = 1/(1 + \exp(-x)) \qquad (10)$$

6) The output of the LSTM is fed into the RF. The maximum depth of the RF is set to 3, and the estimator number is set to 60. The RF output is either 0 or 1, indicating that the appliance is in the non-running or running state, respectively. The mathematical expression for the RF is

$$Gain(D, a) = Ent(D) - \sum_{v=1}^{V} \frac{|D^v|}{|D|} Ent(D^v) \qquad (11)$$

$$Ent(D) = -\sum_{k=1}^{|y|} p_k \log_2 p_k \qquad (12)$$

where *Ent(D)* is the entropy of input D, p_k is the probability distribution of variable.

IV. SIMULATION AND RESULTS

A. DATASET DESCRIPTION

There are several publicly available datasets that can be used for NILM research. This study is based on the Reference Energy Disaggregation Data Set (REDD) [36] and the Pecan Street dataset (PSD) [37]. The REDD is one of the most popular datasets in NILM research. The data were collected from six households for several weeks and contains low-frequency power data and high-frequency voltage and current data. In this manuscript, we consider the data measured from air conditioners, refrigerators, dishwashers, washers/dryers, microwaves and stoves. The Pecan Street dataset measures circuit-level electricity use and generation from nearly 1,000 volunteer homes every minute of every day. It has been collecting data continuously for several years. In this manuscript, data measured from refrigerators, air conditioners, electric vehicles, washers, dish washers and photovoltaics are used for the simulation. The interval of the data used in this study is 1 minute, including the training data and test data.

In the following, the Pecan Street dataset is used to illustrate the complete process of our method from training to testing. At the same time, we verify the validity of our proposed method using the REDD.

B. SIMULATION PROCESS

Corresponding to the three components of our proposed method, the simulation process can be divided into three stages.

Stage I: Overview of data and pre-training.

We analyze the distribution of load for several different 24 hours periods. We then apply DBSCAN to classify the daily load curves. In this simulation, the cluster number, $N_{cluster}$, was determined by DBSCAN to be 5. The typical daily load curves based on classification results are shown in Fig. 5. The ordinate represents the normalized load data value. Different clusters correspond to different combinations of electrical appliances which reflects label correlation. The blue curve (Cluster 1) indicates that the air conditioner was in operation and the photovoltaic operate at high power output in this day. The yellow curve (Cluster 4) indicates that the air

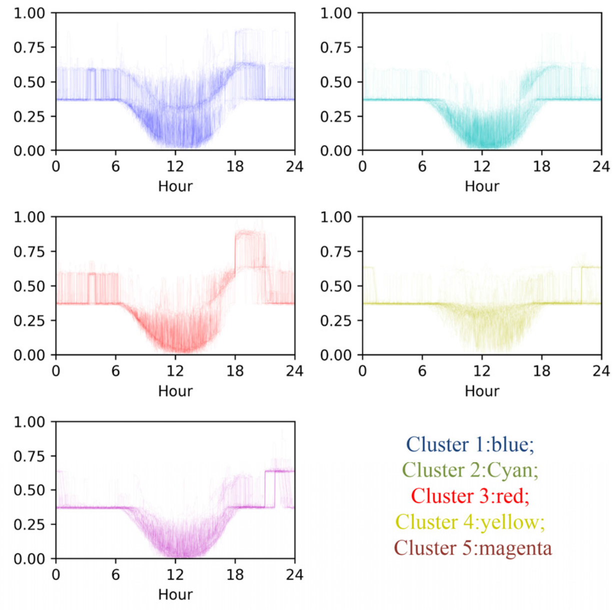

FIGURE 5. Typical daily load curves.

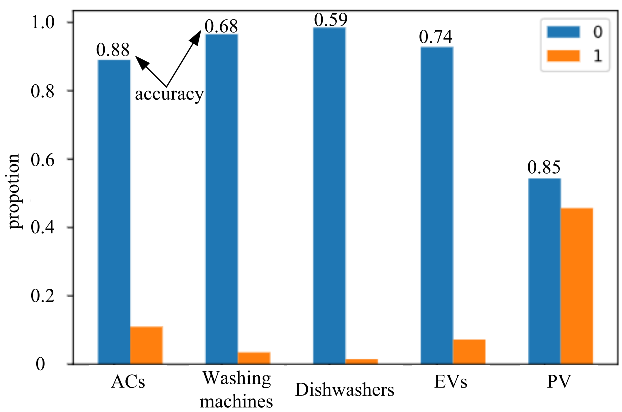

FIGURE 6. The operation state statistics of each electrical appliance.

conditioner was not in operation and photovoltaic operated at low power output. The classification results will be used as model input. The operating state proportion for each electrical appliance is shown Fig. 6. As the specific number of "0" labels and "1" labels is lopsided, the traditional accuracy calculation method (6) is not suitable in this model. The number above the bar chart represents independent recognition accuracy.

Another pre-training task is to arrange the sequence of recognition order for the appliances. According to the independent recognition accuracy of the CNN-LSTM-RF model, the sequence is air-conditioner, photovoltaic, electric vehicle, washing machine and dishwasher in descending order, from highest to lowest.

Stage II: Apply the compound re-weighting method.

The compound re-weighting is calculated using (2)-(5). Parameter β_i is calculated using (13), where acc_i is the

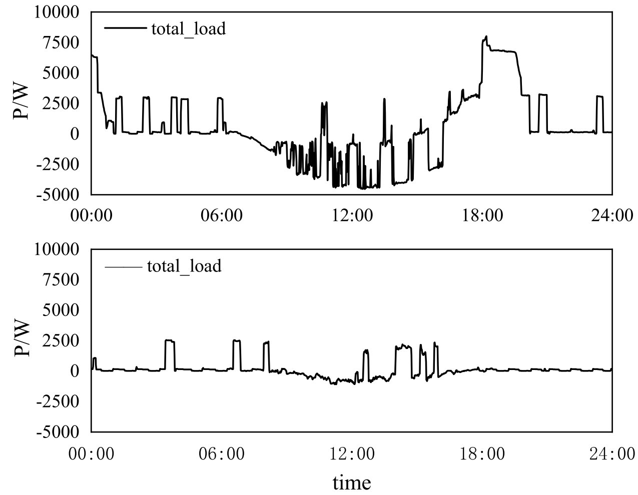

FIGURE 7. Typical daily load curves of a household consumer. The upper subgraph represents Day 1, and the lower subgraph represents Day 2.

recognition accuracy in pre-training.

$$\beta_i = \frac{1}{acc_i}, \quad \forall i \tag{13}$$

Stage III: Build model and verify.

According to the METHODOLOGY section and with the results of pre-training, the CNN-LSTM-RF model is built in Python. The dataset is divided into a training set, a test set and a verification set. The model is first trained by the training set. Then, the model parameters are adjusted using the test set. Finally, the performance of the model is verified by the verification set.

C. RESULTS AND DISCUSSION

Two typical days with different load characteristic were chosen to facilitate a detailed comparison of different models. Day 1 is shown in the upper frame of Fig. 7; Day 2, in the lower. Because the house has installed distributed photovoltaic power generation, the total power is negative during some periods.

Fig. 8 (upper) illustrates the power curve of each appliance for Day 1. The running state curve of each electrical appliance on a typical day is given in Fig. 8 (lower). Because the refrigerator is normally in the ON state, we do not show the state in the results.

Several classic NILM methods are used for this comparative study. The alternative methods considered are the KNN [38], the CNN-LSTM [11], [20]–[25], the combinatorial optimization (CO) [28], the FHMM [31] and the priori biased NILM method (PBN) [39]. In the KNN model, the k value is set to 5, the weights method is uniform, and the algorithm is auto. In the CNN-LSTM model, CNN has two convolutional layers with 128 filters in the first convolution layer and 32 filters in the second. The LSTM has 128 hidden layers. The activation function for the LSTM is sigmoid. The parameters of CO, the FHMM and the PBN are taken from the references accordingly.

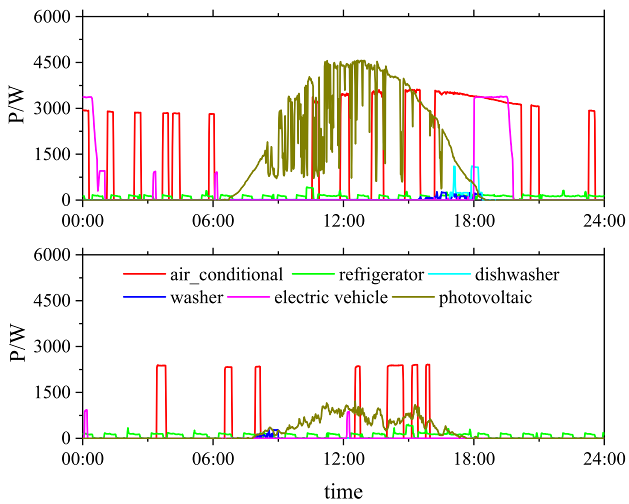

FIGURE 8. Load curve of different household electrical appliances on a typical day.

Taking air conditioner and electric vehicle as examples, the recognition results of the different models are presented in Fig.s 9 and 10. Fig. 9 illustrates the results for the air conditioner. The first column shows the recognition results of Day 1, the second column shows the results of Day 2. The first row is the real state of the appliance, the second through the seventh rows are the recognition results of the different models, corresponding to the KNN, CNN-LSTM, CO, FHMM, PBN and proposed model, respectively.

As Fig. 9 shows, the air conditioner has been used during these two days, the start-stop frequency of the compressor during Day 1 is higher than that on Day 2. Contrasting the recognition results of different models, we see that the recognition accuracy of the proposed model is highest overall. In general, the recognition accuracy for the air conditioner is satisfactory for all of these models.

Fig. 10 shows the recognition results for the electric vehicle. The electric vehicle is charged during Day 1, but it is not charged during Day 2. The recognition accuracy of the KNN is the lowest. Obviously, the recognition ability for the charging state is higher than that for the uncharged state, which means the precision of EVs is larger than the recall. In general, for these two days, the proposed model has the highest recognition accuracy, and the recall score of the proposed model is significantly higher than that of the other models.

Several experiments were conducted on different datasets to compare the recognition ability and generalization ability of each model. Table 1 shows the accuracy scores of the different models calculated using (6) for the REDD and PSD. Table 2 lists the F1 score of the different models on REDD dataset and Table 3 lists the F1 score of different models for the PSD. For the REDD, the system was developed to identify six specific types of electrical appliances: air conditioners, washers/dryers, dishwashers, refrigerators, microwaves and stoves. For the PSD, the system was developed to identify six specific types of electrical appliances: air conditioners,

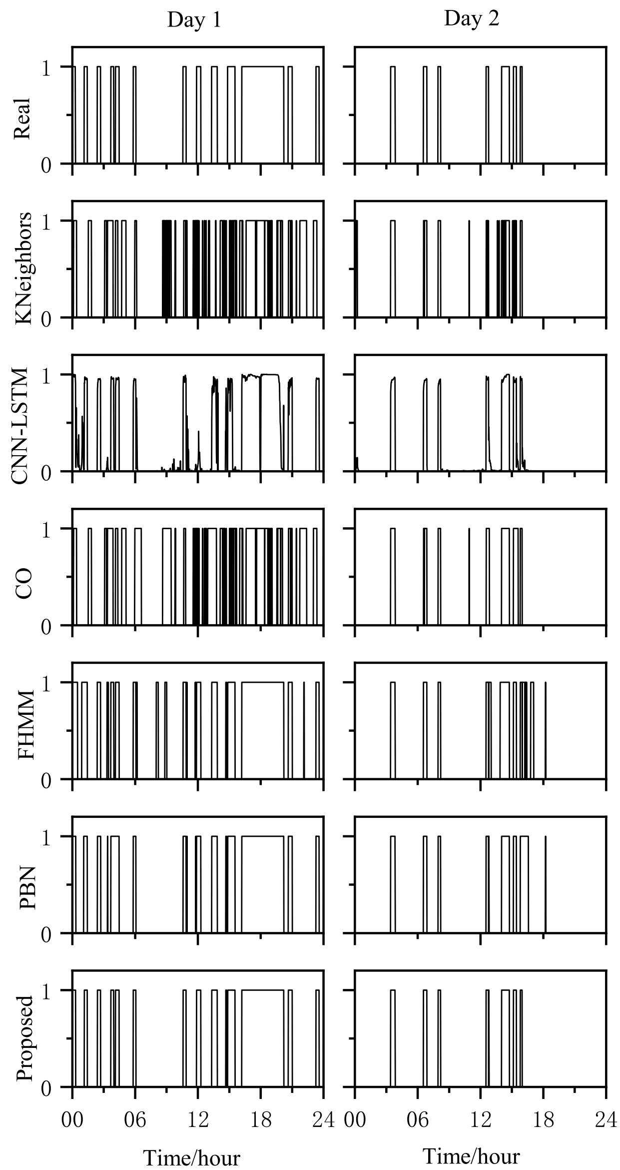

FIGURE 9. Recognition results for the air conditioner under different models.

washing machines, dishwashers, refrigerators, electrical vehicles and photovoltaics.

Training data and test data are divided 1:1. In the REDD, the training set is data from April 18 to April 30, and the test set is data from May 1 to May 12. In the PSD, the former 200 days data were used as the training set, the next 200 days data as the test set.

Table 1 shows that the proposed model was successful in the identification process of the electrical appliances with an accuracy of 95% in the REDD and 94% in the PSD. It can be observed from the table that the proposed method based on the CNN-LSTM-RF model proved to be efficient in the different scenarios, having the highest accuracy scores in all the experiments.

To carry out a detailed comparative study of the different models, the F1 score of each appliance was calculated

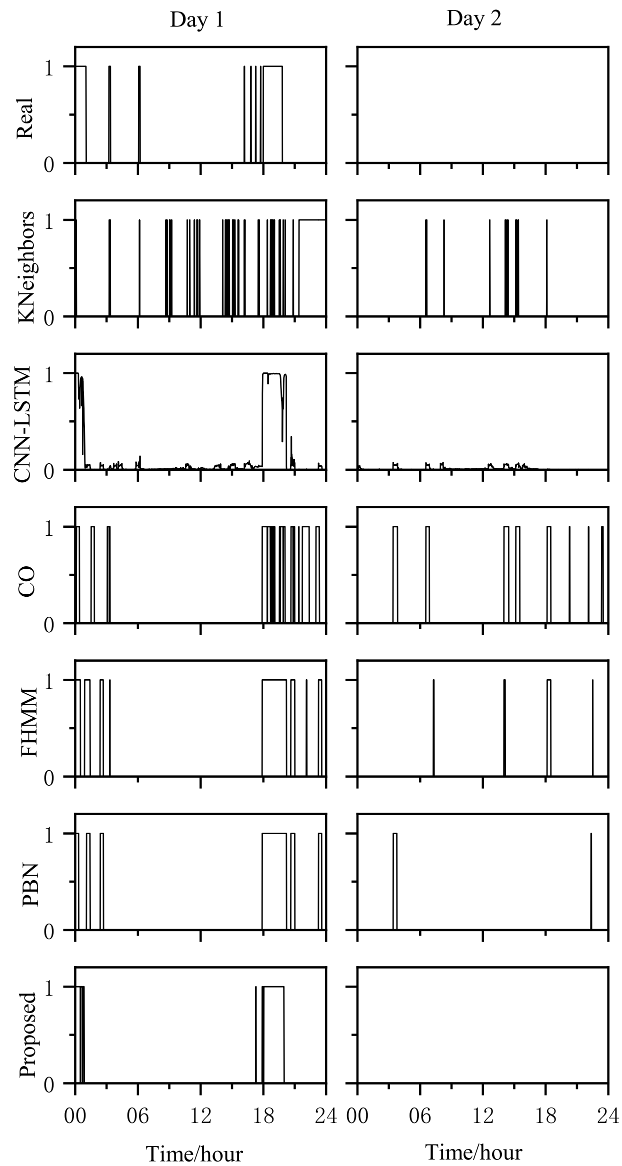

FIGURE 10. Recognition results for EV charging under different models.

TABLE 1. Accuracy of Models on Testing with Different Datasets.

Model / Dataset	KNN [38]	CNN-LSTM [11], [20]-[25]	CO [28]	FHMM [31]	PBN [39]	Proposed
REDD	58%	70%	45%	70%	89%	95%
PSD	55%	82%	64%	79%	91%	94%

under each model, these are shown in Table 2 and Table 3. In Table 2, all of the considered models utilized the REDD. Overall, no matter the electrical appliance, the proposed model obtains the highest F1 score, followed by the PBN. The F1 score for air conditioners is the highest, followed by stoves and washers/dryers.

In Table 3, all the considered models utilized the PSD dataset. The proposed model obtains the highest F1 scores for all appliances, followed by PBN. Observing the F1 scores for different electrical appliances, air conditioners is the highest, followed by PVs and EVs.

TABLE 2. F1 Scores of Different Appliances (REDD dataset).

Model / Appliances	KNN	CNN-LSTM	CO	FHMM	PBN	Proposed
ACs	80%	93%	68%	72%	97%	98%
Washer/Dryer	61%	83%	74%	78%	88%	91%
Dish washers	38%	49%	35%	55%	75%	80%
Microwave	42%	55%	42%	65%	72%	82%
Stove	70%	74%	75%	75%	92%	92%

TABLE 3. F1 Scores of Different Appliances (PSD dataset).

Model / Appliances	KNN	CNN-LSTM	CO	FHMM	PBN	Proposed
ACs	78%	91%	65%	70%	98%	98%
Washing machines	38%	51%	31%	62%	85%	90%
Dish washers	34%	45%	25%	36%	74%	80%
EVs	70%	92%	82%	90%	92%	95%
PVs	75%	80%	85%	92%	94%	96%

V. CONCLUSION AND FUTURE WORK

In this paper, a new method based on a CNN-LSTM-RF deep learning model is proposed to achieve non-intrusive load monitoring. First, a data preprocessing method, DBSCAN, is developed to extract features from the load data. Second, a pre-training process is executed to consider correlations in electrical consumption behaviors and to improve the NILM accuracy. In addition, a compound re-weighting method is proposed to improve the training model and to avoid gradient disappearance. Finally, a combined deep learning network based on CNN-LSTM-RF is constructed to output the final recognition results. The main idea behind the proposed method is to combine expert knowledge and deep learning models. DBSCAN references data processing expert knowledge within NILM. Using DBSCAN, the one-dimensional load data are converted into a two-dimensional matrix, which can be input into the CNN directly.

Two public datasets were used to test the proposed method. Experiments show that the proposed method achieved a satisfactory effect. Both the accuracy scores and F1 scores of the proposed method were the highest among the compared models.

The advantages of the proposed method are that the system can be implemented by utilizing low-cost meters for the acquisition of the aggregated load data because the method needs only low frequency data. In fact, the proposed method can achieve NILM based on one minute interval data, which makes it a promising prospect for application.

Verification of the generalization ability of the model is a direction of our future research. In addition, the deep learning model requires a large amount of labeled data, which is difficult to obtain. An unsupervised model is under consideration to realize the recognition of unlabeled data through pre-training.

REFERENCES

[1] H.-H. Chang, K.-L. Chen, Y.-P. Tsai, and W.-J. Lee, "A new measurement method for power signatures of nonintrusive demand monitoring and load identification," *IEEE Trans. Ind. Appl.*, vol. 48, no. 2, pp. 764–771, Mar. 2012.

[2] L. Guo, S. Wang, H. Chen, and Q. Shi, "A load identification method based on active deep learning and discrete wavelet transform," *IEEE Access*, vol. 8, pp. 113932–113942, 2020.

[3] P. Held, S. Mauch, A. Saleh, D. Ould Abdeslam, and D. Benyoucef, "Frequency invariant transformation of periodic signals (FIT-PS) for classification in NILM," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5556–5563, Sep. 2019.

[4] T.-T.-H. Le, H. Kang, and H. Kim, "Household appliance classification using lower odd-numbered harmonics and the bagging decision tree," *IEEE Access*, vol. 8, pp. 55937–55952, 2020.

[5] Y.-H. Lin and M.-S. Tsai, "Development of an improved time–frequency analysis-based nonintrusive load monitor for load demand identification," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 6, pp. 1470–1483, Jun. 2014.

[6] F. D. Garcia, W. A. de Souza, and F. P. Marafao, "Embedded NILM as home energy management system: A heterogeneous computing approach," *IEEE Latin Amer. Trans.*, vol. 18, no. 2, pp. 360–367, Feb. 2020.

[7] C. Dinesh, B. W. Nettasinghe, R. I. Godaliyadda, M. P. B. Ekanayake, J. Ekanayake, and J. V. Wijayakulasooriya, "Residential appliance identification based on spectral information of low frequency smart meter measurements," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2781–2792, Nov. 2016.

[8] R. Machlev, Y. Levron, and Y. Beck, "Modified cross-entropy method for classification of events in NILM systems," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 4962–4973, Sep. 2019.

[9] J. Kelly and W. Knottenbelt, "Neural NILM: Deep neural networks applied to energy disaggregation," in *Proc. 2nd ACM Int. Conf. Embedded Syst. Energy-Efficient Built Environ.*, Nov. 2015.

[10] F. M. Wittmann, J. C. Lopez, and M. J. Rider, "Nonintrusive load monitoring algorithm using mixed-integer linear programming," *IEEE Trans. Consum. Electron.*, vol. 64, no. 2, pp. 180–187, May 2018.

[11] F. Ciancetta, G. Bucci, E. Fiorucci, S. Mari, and A. Fioravanti, "A new convolutional neural network-based system for NILM applications," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–12, 2021.

[12] G. Zhou, Z. Li, M. Fu, Y. Feng, X. Wang, and C. Huang, "Sequence-to-sequence load disaggregation using multiscale residual neural network," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 25033100.

[13] Y. Yang, J. Zhong, W. Li, T. A. Gulliver, and S. Li, "Semisupervised multilabel deep learning based nonintrusive load monitoring in smart grids," *IEEE Trans. Ind. Informat.*, vol. 16, no. 11, pp. 6892–6902, Nov. 2020.

[14] E. Gomes and L. Pereira, "PB-NILM: Pinball guided deep non-intrusive load monitoring," *IEEE Access*, vol. 8, pp. 48386–48398, 2020.

[15] H.-H. Chang, L.-S. Lin, N. Chen, and W.-J. Lee, "Particle-swarm-optimization-based nonintrusive demand monitoring and load identification in smart meters," *IEEE Trans. Ind. Appl.*, vol. 49, no. 5, pp. 2229–2236, Sep. 2013.

[16] V. Singhal, J. Maggu, and A. Majumdar, "Simultaneous detection of multiple appliances from smart-meter measurements via multi-label consistent deep dictionary learning and deep transform learning," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2969–2978, May 2019.

[17] D. Li and S. Dick, "Residential household non-intrusive load monitoring via graph-based multi-label semi-supervised learning," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4615–4627, Jul. 2019.

[18] S. Singh and A. Majumdar, "Non-intrusive load monitoring via multi-label sparse representation-based classification," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1799–1801, Mar. 2020.

[19] M. D'Incecco, S. Squartini, and M. Zhong, "Transfer learning for non-intrusive load monitoring," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1419–1429, Mar. 2020.

[20] K. Jihyun, T. Le, and K. Howon, "Nonintrusive load monitoring based on advanced deep learning and novel signature," *Comput. Intell. Neurosci.*, vol. 2017, Oct. 2017, Art. no. 4216281.

[21] L. Mauch and B. Yang, "A new approach for supervised power disaggregation by using a deep recurrent LSTM network," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2015, pp. 63–67.

[22] M. Kaselimi, N. Doulamis, A. Doulamis, A. Voulodimos, and E. Protopapadakis, "Bayesian-optimized bidirectional LSTM regression model for non-intrusive load monitoring," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 2747–2751.

[23] Z. Lan, B. Yin, T. Wang, and G. Zuo, "A non-intrusive load identification method based on convolution neural network," in *Proc. IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, China, Beijing, Nov. 2017, pp. 1–5.

[24] D. Yang, X. Gao, L. Kong, Y. Pang, and B. Zhou, "An event-driven convolutional neural architecture for non-intrusive load monitoring of residential appliance," *IEEE Trans. Consum. Electron.*, vol. 66, no. 2, pp. 173–182, May 2020.

[25] A. P. Medeiros, L. N. Canha, D. P. Bertineti, and R. M. de Azevedo, "Event classification in non-intrusive load monitoring using convolutional neural network," in *Proc. IEEE PES Innov. Smart Grid Technol. Conf.-Latin Amer. (ISGT Latin Amer.)*, Sep. 2019, pp. 1–6.

[26] H. Altrabalsi, V. Stankovic, J. Liao, and L. Stankovic, "Low-complexity energy disaggregation using appliance load modelling," *AIMS Energy*, vol. 4, no. 1, pp. 884–905, 2016.

[27] J. Liao, G. Elafoudi, L. Stankovic, and V. Stankovic, "Non-intrusive appliance load monitoring using low-resolution smart meter data," in *Proc. IEEE Int. Conf. Smart Grid Commun. (SmartGridComm)*, Nov. 2014, pp. 535–540.

[28] G. W. Hart, "Nonintrusive appliance load monitoring," *Proc. IEEE*, vol. 80, no. 12, pp. 1870–1891, Dec. 1992.

[29] P. R. Z. Taveira, C. H. V. De Moraes, and G. Lambert-Torres, "Non-intrusive identification of loads by random forest and fireworks optimization," *IEEE Access*, vol. 8, pp. 75060–75072, 2020.

[30] J. Z. Kolter and T. S. Jaakkola, "Approximate inference in additive factorial HMMs with application to energy disaggregation," in *Proc. AISTATS*, Mar. 2012, pp. 1472–1482.

[31] R. Bonfigli, E. Principi, M. Fagiani, M. Severini, S. Squartini, and F. Piazza, "Non-intrusive load monitoring by using active and reactive power in additive factorial hidden Markov models," *Appl. Energy*, vol. 208, pp. 1590–1607, Dec. 2017.

[32] W. Kong, Z. Y. Dong, D. J. Hill, J. Ma, J. H. Zhao, and F. J. Luo, "A hierarchical hidden Markov model framework for home appliance modeling," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3079–3090, Jul. 2018.

[33] P. Yang, H. Sun, and H. Wang, "SGM: Sequence generation model for multi-label classification," in *Proc. 27th Int. Conf. Comput. Linguistics*, 2018, pp. 3915–3926.

[34] A. Garg, J. Noyola, R. Verma, A. Saxena, and A. Jami, "Exploring correlation between labels to improve multi-label classification," *Comput. Sci.*, 2015. [Online]. Available: http://arxiv.org/abs/1511.07953v1

[35] C. Huang, Y. Li, C. Change Loy, and X. Tang, "Deep imbalanced learning for face recognition and attribute prediction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 11, pp. 2781–2794, Nov. 2020.

[36] J. Z. Kolter and J. M. Johnson, "REDD: A public data set for energy disaggregation research," in *Proc. SustKDD Workshop Data Mining Appl. Sustainability*, 2011, pp. 59–62.

[37] (Mar. 2021). *Pecan Street Dataset*. [Online]. Available: http://www.pecanstreet.org/

[38] S. M. Tabatabaei, S. Dick, and W. Xu, "Toward non-intrusive load monitoring via multi-label classification," *IEEE Trans. Smart Grid*, vol. 8, no. 1, pp. 26–40, Jan. 2017.

[39] S. Welikala, C. Dinesh, M. P. B. Ekanayake, R. I. Godaliyadda, and J. Ekanayake, "Incorporating appliance usage patterns for non-intrusive load monitoring and load forecasting," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 448–461, Jan. 2019.

XIAO ZHOU received the B.S. degree in Water Conservancy and Hydropower Project and the Ph.D. degree in electrical engineering from the Huazhong University of Science and Technology, in 2014 and 2019, respectively. Since 2020, he holds a postdoctoral position with the Energy Development Research Institute, CSG. His research interests include demand response, energy management, and energy storage.

SHUJIAN LI received the B.S. degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2018, where he is currently pursuing the M.Sc. degree in electrical engineering. His current research interests include load forecasting and demand response in smart grid.

CHENGXI LIU (Senior Member, IEEE) received the B.Eng. and M.Sc. degrees from the Huazhong University of Science and Technology, Wuhan, China, in 2005 and 2007, respectively, and the Ph.D. degree from the Department of Energy Technology, Aalborg University, Aalborg, Denmark, in 2013. He was with Energinet, the Danish TSO, until 2016. He was a Research Associate with the Department of Electrical Engineering and Computer Science, The University of Tennessee, USA, from 2016 to 2018. He was an Associate Professor with the Department of Energy Technology, Aalborg University, from 2018 to 2019. He is currently a Professor with the Department of Electrical Engineering and Automation, Wuhan University. His research interests include stability, dynamics, and control of power grids, simulation methods of power systems, and big data and artificial intelligence applied in power systems.

HAOJUN ZHU received the B.S. and Ph.D. degrees in electrical engineering from the South China University of Technology, in 2001 and 2006, respectively. He is a Researcher with the Energy Development Research Institute, China Southern Power Grid. His research interests include demand response and energy supply-demand balance and flexibility.

NAN DONG received the B.S. degree in electrical engineering from Northeast Electric Power University, Jilin, China, in 2009, and the M.S. degree in electrical engineering from North China Electric Power University, Beijing, China, in 2012. From 2012 to 2021, she was an Engineer with the Energy Development Research Institute, CSG, Guangzhou, China. Her research interests include energy supply and demand, power economy, and electricity demand forecasts.

TIANYING XIAO received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2013, and the Ph.D. degree from the Institute of Electrical Engineering, Chinese Academy of Sciences, Beijing, in 2019. She is currently a Researcher with the Energy Development Research Institute, China Southern Power Grid. Her research interests include active distribution network operation and planning with large-scale renewable energy integration, demand response, and energy supply-demand balance and flexibility.