Research Article

Household Electricity Load Forecasting Based on Multitask Convolutional Neural Network with Profile Encoding

Mingxin Wang [ID],[1] **Yingnan Zheng** [ID],[2] **Binbin Wang** [ID],[2] **and Zhuofu Deng** [ID][2]

[1]*College of Water Conservancy, ShenYang Agricultural University, Shenyang 110866, China*
[2]*College of Software, Northeastern University, Shenyang 110004, China*

Correspondence should be addressed to Yingnan Zheng; software_zheng@126.com

Household load forecasting provides great challenges as a result of high uncertainty in individual consumption of load profile. Traditional models based on machine learning tried to explore uncertainty depending on clustering, spectral analysis, and sparse coding with hand craft features. Recently, deep learning skills like recurrent neural network attempt to learn the uncertainty with one-hot encoding which is too simple and not efficient. In this paper, for the first time, we proposed a multitask deep convolutional neural network for household load forecasting. The baseline of one branch is built on multiscale dilated convolutions for load forecasting. The other branch based on deep convolutional autoencoder is responsible for household profile encoding. In addition, an efficient encoding strategy for household profile is designed that serves a novel feature fusion mechanism integrated into forecasting branch. Our proposed network serves an end-to-end manner in training and inference process. Sufficient ablation studies were conducted to demonstrate effectiveness of innovations and great generalization in point and probabilistic load forecasting at household level, which provides a promising prospect in demand response.

1. Introduction

Smart grid is considered as an electric grid that specializes in delivering electricity in a controlled and intelligent approach from points of generation to consumers, both of which form an integral part of the smart grid when customers are able to modify their purchasing patterns and behavior according to the received information, incentives, and disincentives [1–3]. Attractions of smart grid depend on its capability of improving reliability performance spontaneously, encouraging customers' responsiveness and advanced efficiency decisions between customers and utility providers [3, 4]. Consequently, demand side management (DSM) occupies an essential integral part of smart grid [5–7]. Meanwhile, the smart meter plays a crucial role in DSM that is able to achieve energy savings, exploit renewable energy resources, and encourage customers' participation in energy market depending on deep cognition to residential load profiles or behaviors [8].

As a crucial component of demand response (DR), load forecasting is categorized with different horizon: very short-term load forecasting (VSTLF), short-term load forecasting (STLF), medium-term load forecasting (MTLF), and long-term load forecasting (LTLF) [9]. Related cut-off horizons involve days, weeks, and years [10]. A large number of comprehensive researches pay more attentions on point forecasting at high aggregated level [11–24]. However, there are relatively few studies on household load forecasting that provides great uncertainties or volatilities as a result of various lifestyle and randomness of consumption behavior in residential lives [25–29]. Some researches have evaluated popular models for VSTLF and STLF at household level [27, 29–37], most of which rely on machine learning approaches of time series analysis such as autoregressive integrated moving average model (ARIMA), fuzzy clustering, sparse coding, and support vector regression (SVR).

Particularly, deep learning skills, for example, recurrent neural networks (RNNs) or convolutional neural networks (CNNs), have been confirmed to have superior performance

[27, 29, 37] rather than traditional methods at load forecasting in different aggregated level due to excellent capability of extracting discriminative correlation features in sequence. Chitalia et al. [38] presented a robust short-term electrical load forecasting framework that can capture variations in building operation, regardless of building type and location using advanced deep recurrent neural networks. Deng et al. [9, 29, 39] devised a novel convolutional neural model with multiscale dilated kernel for short-term load or price forecasting, which provides more predominance of extracting significant features in time-series analysis. Sideratos et al. [40] gave an advanced fuzzy-based ensemble model for load forecasting using hybrid deep neural networks following a two-stage architecture including radial basis function neural network (RBFNN) and CNN. Li et al. [41] developed a convolutional long short-term memory-based neural network with selected autoregressive features to improve short-term household electricity load forecasting accuracy by employing three strategies: autoregressive features selection, exogenous features selection, and a "default" state to avoid overfitting at times of high load volatility. Dong et al. [42] designed a deep learning approach based on K-nearest neighbors to capture uncertainty and reflect the range of electrical load fluctuation. Dudek et al. [43] presented a hybrid and hierarchical deep learning model for midterm load forecasting. The model combines exponential smoothing (ETS), advanced long short-term memory (LSTM) and ensembling. ETS extracts dynamically the main components of each individual time series and enables the model to learn their representation. Multilayer LSTM is equipped with dilated recurrent skip connections and a spatial shortcut path from lower layers to allow the model to better capture long-term seasonal relationships and ensure more efficient training.

Although deep learning skills have achieved better performances in short-term load forecasting, proper lag selection and hyperparameters setting in deep learning are required for searching optimal training results. It is rather a hard problem for which an optimal solution cannot be found in a polynomial time. This hardness is accentuated by the complexity of electricity-consumption data patterns [44]. One effective strategy to acquire an optimal configuration in forecasting model depends on metaheuristics approaches [45–56] with excellent capability of finding near-optimal solutions in a very large space. However, there are some great challenges in specific field of household short-term load forecasting.

1.1. Reducing Uncertainty.

Generally, load forecasting concentrates on different aggregation levels, such as system, feeder, and regional ones. In some level, individuals could be inferred from one forecasting model with shared parameters as a result of their similar behaviors. For instance, load patterns of commercial buildings serve a larger granularity, where different rooms provide regular consumptions under external factors such as weather or central air conditions. Load profiles of manufactory enterprises reflect little variations with relatively stable plans of production. However, the load profiles of residences exhibit more volatilities and uncertainties because of their different lifestyles and randomness of behaviors providing great challenges in forecasting accuracy. Researches [29, 37] demonstrated that original household load profile can be decomposed into regular pattern, uncertainty, and noise. Regular pattern refers to periodical load profile. Uncertainty depends on some aperiodic and external factors such as weather, family activities, and individual preferences. Noise represents the residue that cannot be physically explained [11, 13, 37].

Most machine learning skills are competent in learning linear relationships and exploring regular patterns effectively. In contrast, these approaches with hand craft features cannot deal with uncertainty at household level that accounts for a great proportion. Different household holds different behavior and remarkable variations in time series delivering great stochasticity and nonlinear salience. Consequently, relying on shared model based on traditional model is almost doubtful. To address these problems, three categories of methods have been presented [37]:

(1) Clustering approaches [36, 57, 58] were designed to group households based on similar behavior at high level, which decreases uncertainty in each category and extracts more regular patterns to facilitate household load forecasting. However, how to segment subjects appropriately prohibits an acceptable result generally and this strategy is excessively sensitive to different dataset. Meanwhile, some researches [18–21] proposed aggregate load forecasting (ALF) to cancel uncertainty. ALF is actually considered as a larger granular level not specialized in household ones.

(2) Some spectral analyses such as Fourier transforms [23], wavelet analysis [22], and empirical mode decomposition were introduced in order to extract the regular pattern located in load profile. This strategy is not suitable for household load forecasting since regular information occupies relatively smaller proportion.

(3) In the domain of power delivery systems, sparse coding has been applied to the problem of energy disaggregation [59, 60]. Recently, sparse coding becomes preferred at household level that provides each house a profile description and an efficient approach to separate uncertainty, learning, and representing gross patterns of individual consumption [37, 61]. Yu et al. [61] analyzed and decomposed the dataset into fixed patterns that constitute ultimate format of encoding, which lacks flexibility to show various uncertainties. Shi et al. [37] used one-hot encoding to increase individual features in deep RNN to extract uncertainty achieving the state-of-

the-art performance. However, sparse encoding cannot handle massive users and express similarity among households effectively.

Through investigations, most of researches tried to adopt clustering, pattern separation, or sparse coding to learn uncertainties of load profile at household level. These strategies based on machine learning offer a restricted capability of modeling with hand craft features in time series analysis. In recent years, deep learning has appeared as a powerful tool in areas such as image processing and data analysis. With deeper architecture and sophisticated operations, deep neural networks provide superior ability of learning discriminative features and nonlinear relationships, which benefits extracting uncertainty at household load forecasting [29, 37].

1.2. Building Effective Network. Deep learning skills have been applied in load forecasting, and most of them rely on RNN or long short-term memory (LSTM) [27, 37, 62–66, 66]. LSTM is derived form RNN, both of which are successful in the target of sequence to sequence learning such as speech recognition and natural language processing in time series analysis. However, when managing long-term sequence, RNNs suffer from the problem of gradient disappearance severely, even though LSTM alleviates this case partly. Specifically, latest researches [9, 29] revealed that convolutional neural network (CNN) offers more advanced accuracy as a result of powerful capability of discriminative feature extraction. In addition, some mechanisms like residual connection cannot cause dramatical gradient disappearance even in deeper network. Consequently, related skills could be optimized to identify and learn both regular pattern and uncertainty in load profile at household level.

In this paper, we propose a multitask convolutional neural network with household profile encoding (MCNN-HPC). The novel encoding branch serves more effective description on household behavior especially focusing on uncertainty. In coordination with multiscale dilated convolutional neural network [9], our proposed model provided the state-of-the-art performance of VSTLF at household level. The key contributions are as follows:

(i) We propose a multitask neural network that consists of two branches. The baseline of one branch is built on multiscale dilated convolutions for load forecasting. The other branch based on deep convolutional autoencoder is responsible for household profile encoding.

(ii) A novel encoding strategy is designed to explore uncertainty in household behavior effectively. Compared with traditional technique in deep neural network, our proposed method has great predominance to express individual behavior feature and nonlinear correlation in time series analysis.

(iii) We present a novel mechanism of feature fusion between two branches, which is also interpreted as a superior feature selection process and leads to remarkable improvement in accuracy.

(iv) Our proposed network serves an end-to-end manner in training and inference process. Sufficient ablation studies were conducted to demonstrate effectiveness of innovations and great generalization in point and probabilistic load forecasting at household level.

The rest of the paper is structured as follows: Section 2 defines the problem and describes the details of our proposed model. Section 3 introduces the experiment setups. Section 4 exhibits and discusses the detailed results of comparison experiments. The conclusions are drawn in Section 5.

2. Methodology

2.1. Problem Formulation. Our research focuses on VSTLF at household level that pays more attention to load forecasting for the nearest point (the next 30 minutes) in very short term. In practice, we only employ the historical load data (X_t) for training and inference process of our proposed neural network. In electricity market, X_t is easily acquired from households via smart meters. Consequently, our task can be described to build nonlinear relationships between historical load sequences and predicted points as follows:

$$Y = f(X_t), \tag{1}$$

where $X_t = [x_1, x_2, \ldots, x_t]$ denotes the historical load sequence happening in $1, \ldots, t$ time, and $Y = [y_{t+1}, y_{t+2}, \ldots, y_{t+n}]$ represents the output of prediction. t and n manifest the length of the input and output sequence. When $n = 1$, the prediction target becomes the single-step forecasting VSTLF. If $n > 1$, it belongs to a multistep forecasting task. In point load forecasting, y_t becomes a scalar, while in probabilistic load forecasting y_t grows to a vector with length q, denoting q quantiles estimated at t.

2.2. Construction of MCNN-HPC

2.2.1. Backbone Network. Our proposed deep neural network consists of two branches corresponding to different tasks, respectively, in Figure 1. Forecasting branch is responsible for household load forecasting as the baseline of MCNN-HPC. Household profile branch provides more advanced encoding information to learn uncertainty of individual behavior based on historical load profile. Feature Fusion 1 to 3 as an innovative concatenation of different level network serves more excellent feature selection process than traditional manners. Both branches are fused at the end of the network with a fully connected layer providing an end-to-end manner for training and inference of load forecasting at household level.

2.2.2. Forecasting Branch. In forecasting branch, the baseline of the network includes multiple convolutional blocks with different dilated ratio kernels, which is able to extract

multiscale features reflecting various nonlinear relationships in sequence. This strategy has been demonstrated an advanced optimization applied in CNNs for load forecasting [9]. In practice, we set the input sequence a 48-dimensional vector and each point denotes half an hour. Consequently, the input vector refers to the load sequence of 24 hours before the predicted time. Forecasting branch consists of 8 convolutional blocks with dilated rates 1, 2, 4, and 8 convolutional kernels, respectively, and each block produces $8 \times 48 \times 1$ feature maps. In order to avoid gradient disappearance and enhance quality of training, forecasting branch increases lots of residual connections between blocks, illustrated in Figure 1.

2.2.3. Household Profile Branch.

Profile encoding branch is responsible for generating personalized code to learn and reflect uncertainty in daily life of each household. The input of household profile branch comes from a deep convolutional autoencoder (DCAE), illustrated in Figure 2. Depending on deeper networks and convolutions, DCAE serves more excellent capability of squeezing input sequence into latent-space representation that superiorly expresses inherent features and nonlinear relationships in time-series analysis [67, 68]. Our designed DCAE holds a symmetrical encoder-decoder structure with three convolutional blocks on both sides. 336-dimensional vector is devised as original and reconstructed input, where each point represents the load that has happened in every half hour on average 52 weeks in one year. Through maxpooling and upsampling operations, the output of middle layer (yellow color) is the specific encoding result for each household profile with 42 dimensions, which also offers the input of household profile branch in Figure 1. In practice, we use DCAE to generate 42-dimensional household feature vector for individuals based on historical load data reflecting discriminative uncertainty in behavior prominently. In addition, we design fully connected layers that constitute household profile branch. After two shared layers, three kinds of fully connected layers with different number of activation neurons are linked to forecasting branch for feature fusion.

2.2.4. Feature Fusion.

Feature selection is an essential process where features are automatically or manually selected and contribute most to prediction. In time series analysis, models based on machine learning try to present advanced supervised or unsupervised algorithms to explore more significant features to acquire potential nonlinear relationships. For example, Cai et al. [69] proposed a direct multistep model based on gated convolutional neural network (GCNN) for multistep load forecasting. GCNN module imports gated mechanism to select salient features in CNN achieving the state-of-the-art performance. However, this model also suffers from the problem of LSTM with limited feature expression and gradient disappearance.

In this paper, we propose a novel feature selection process shown in Figure 1, where outputs of household profile branch as learnable weights are fused into baseline of forecasting branch by multiplication operation. Three outputs of household profile branch are set to $1 \times 48 \times 1$, $8 \times 48 \times 1$, and $1 \times 8 \times 1$, respectively. As shown in Figure 1, in operations of feature Fusion 1 and 2, compared with traditional concatenation element-wise multiplication fusion makes sure of more effective feature selection process. The vector from household profile branch is filtered by Sigmoid activation and values are located within from 0 to 1. Moreover, in feature Fusion 3, as our proposed model focuses on single-step forecasting, we use the vector that consists of last points (red color) in each channel of feature maps to join in feature selection, which depends on an important assumption that the load that happened in the last half an hour has the closest relationship with the forecasting point. Relying on sophisticated studying in an end-to-end manner, the well-trained household profile branch provides proper weights for individuals to extract remarkable features, respectively, in order to understand their regular and uncertain pattern with the shared model. Therefore, the entire MCNN-HPC is able to explore more nonlinear relationships in consumption behavior of each house achieving more competent performance with great generalization for load forecasting at household level.

3. Implementation

3.1. Data Description.

The dataset selected in experiments is the Smart Metering Electricity Customer Behavior Trails (CBTs), which belongs to a smart metering project launched by the Commission for Energy Regulation in Ireland. The trails took place during 1st July 2009 and 31st December 2010 with over 5000 Irish residential consumers and small and medium enterprise (SMEs) participating. The full anonymized dataset is publicly available online and comprises three parts: (1) half-hourly sampled electricity consumption (kWh) from each participant; (2) questionnaires and corresponding answers from surveys; (3) customer type, tariff, and stimulus description, which specifies customer types, allocation of tariff scheme, and demand side management (DSM) stimuli [37]. In detail, there were 929 residential customers who did not join any demand program and enjoyed controlled stimulus and tariff. In other words, their consumption can realistically reflect behaviors filled with regular pattern and uncertainty.

3.2. Software and Hardware Platform.

All experiments were conducted on a cloud server with two NVIDIA P4 computing cards and the CPU with 8 cores. Deep neural models were implemented by the Keras framework with TensorFlow backend [70].

3.3. Program Implementation.

Our proposed model consists of two tasks: forecasting and household profile branches. At the beginning, an individual vector of historical load profile is encoded via a well-designed DCAE, and the output 42-dimensional feature vector is then delivered to household profile branch as input. With
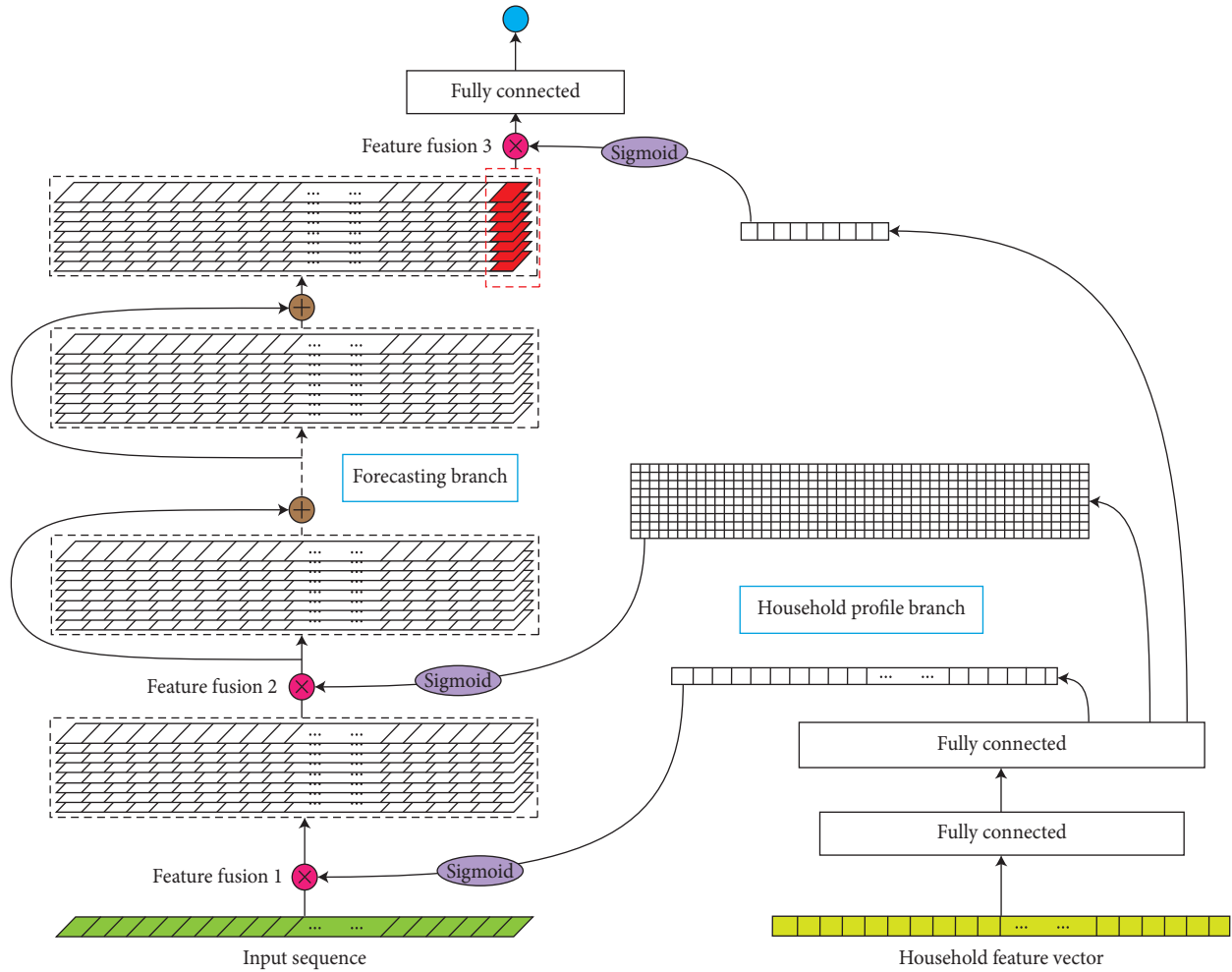
FIGURE 1: The architecture of MCNN-HPC. Our proposed network consists of the forecasting branch and household profile branch. Forecasting branch is responsible for load forecasting. Household profile branch describes specific behavior of each individual. Both of two branches are fused with an efficient approach for point or probabilistic forecasting.

effective feature fusion, both branches are integrated significantly with an end-to-end manner giving contributions to load forecasting at household level. The implementation process is divided into three stages: (1) data preprocessing; (2) household profile encoding; (3) forecasting. Details are described in Figure 3. We trained the proposed model for each customer with shared parameters for households. During the training, we used the learning rate decay and early stopping strategies based on the variation of validation loss to reduce computation cost and prevent overfitting.

3.4. Benchmarks and Setup. For the data preprocessing, as a result of noise interferences, we removed some redundant data and filled the missing ones by linear interpolation. For training process, the raw data from Irish dataset is manipulated into input through two branches, where household load profiles are captured from smart meters half hourly. At forecasting branch, for each household, the input sequence uses 24-hour load data before the forecasting time, a 48-dimensional vector.

Consequently, there are nearly 25,000 datasets which are divided into training set, validation set, and test set 80%, 10%, and 10%, respectively. 336-dimensional vector of load profile in one year is encoded into a 42-dimensional one based on DCAE, and then the feature vector is delivered to household profile branch as input for training and inference.

For the principle of fairness, in ablation studies, we kept the same configuration of compared neural models. The experiment setups and hyperparameters of convolutional neural networks are presented in Table 1.

4. Results and Discussion

4.1. Evaluation Metrics. For evaluation of our proposed model MCNN-HPC on smart metering load data from Irish load profile database, popular metrics are employed on point and probabilistic load forecasting at household level. Metrics of point forecasting include mean arctangent absolute percentage error (MAAPE) [71] and root mean squared error (RMSE).
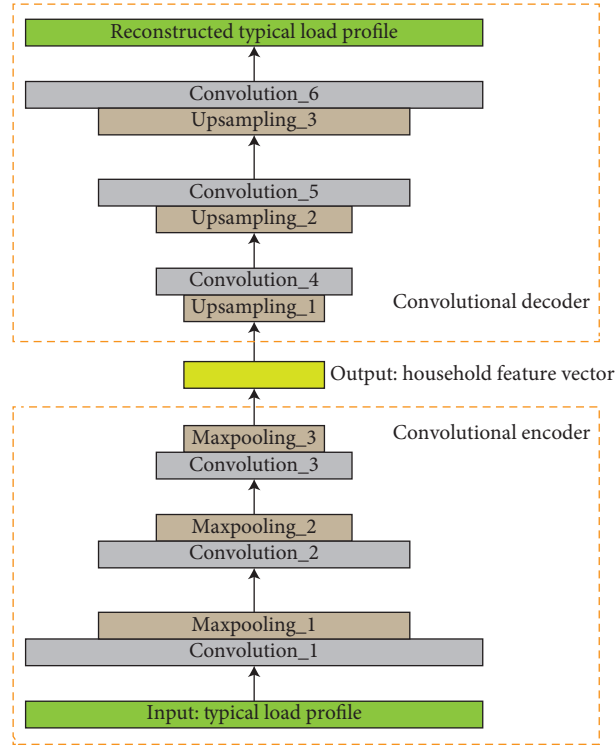
FIGURE 2: The network architecture of DCAE. The middle layer output is used as the input into household profile branch.

$$\text{MAAPE} = \frac{1}{N} \sum_{t=1}^{N} \text{AAPE}_t = \frac{1}{N} \sum_{t=1}^{N} \arctan\left(\left|\frac{y_t - \widehat{y}_t}{y_t}\right|\right), \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{N} (\widehat{y}_t - y_t)^2}{N}}, \quad (3)$$

where $\widehat{y}_t$ is the forecast value and y_t is the actual outcome value at time t. The mean absolute percentage error (MAPE) is one of the most widely used measures of forecast accuracy, due to its advantages of scale-independency and interpretability. However, MAPE has the significant disadvantage that it produces infinite or undefined values for zero or close-to-zero actual values. In order to address this issue in MAPE, MAAPE calculates the mean arctangent percentage error between the forecast and the eventual outcomes. MAAPE inherently preserves the philosophy of MAPE, overcoming the problem of division by zero by using bounded influences for outliers in a fundamental manner through considering the ratio as an angle instead of a slope [71].

For probabilistic forecasting evaluation, there are three commonly used attributes: reliability, sharpness, and resolution. Reliability refers to how close the predicted distribution is to the ground truth. Sharpness means how tightly the predicted distribution covers the actual curve. Resolution signifies how much the predicted interval varies over time. Measures like Kolmogorov–Smirnov, Cramer–von Mises, and Anderson–Darling statistics assess the unconditional coverage of a probabilistic forecasting rather than its sharpness or resolution. In this paper, the performance of

the probabilistic forecasting is evaluated by the average pinball score, which is a comprehensive measure metric considering not only reliability but sharpness and resolution. The quantile scores have the same equation with quantile loss, and pinball score is defined as follows:

$$\text{Pinscore} = \frac{1}{T_{\text{test}} \times Q} \sum_{t=1}^{T_{\text{test}}} \sum_{q=1}^{Q} (y_t - \widehat{y}_{t,q})(q - 1\{y_t < \widehat{y}_{t,q}\}),$$

$$(4)$$

where y_t is the truth at time t, $\widehat{y}_{t,q}$ denotes the forecast of quantile q at time t, Q refers to the defined number of quantiles, and T_{test} represents the number of samples in the test set. In addition, in order to make a proper evaluation on candidates, the prediction interval (PI) should be assessed. The Winkler score is another comprehensive measure that allows a joint assessment of the unconditional coverage and interval width. A central PI of time t with $100(1 - \alpha)\%$ confidence level is given as $[L_t, U_t]$, where L_t and U_t are the lower and the upper boundaries of the PI.

$$\text{Winkler} = \begin{cases} \delta_t, & L_t \leq y_t \leq U_t, \\[2mm] \delta_t + \dfrac{2(L_t - y_t)}{\alpha}, & y_t < L_t, \\[2mm] \delta_t + \dfrac{2(y_t - U_t)}{\alpha}, & y_t > U_t, \end{cases} \quad (5)$$

where the interval width δ_t is calculated by $\delta_t = U_t - L_t$. In this paper, we evaluate the PI coverage of 80% for Winkler
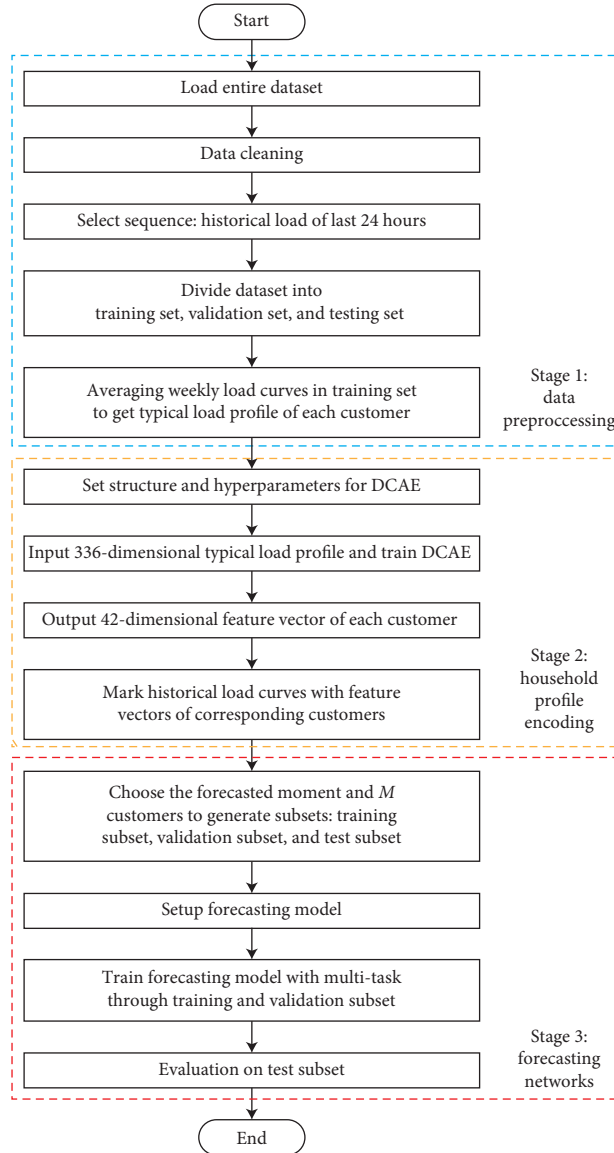
FIGURE 3: Flowchart for implementation of the proposed model.

score. Lower pinball score and Winkler score indicate a better performance.

4.2. Evaluation of Multitask Network. In order to evaluate the effectiveness of our proposed model with multitask, ablation studies were conducted on the contribution of household profile branch. In detail, we randomly selected 10 households to construct training set, validation set, and testing set, each of which contains nearly 25,000 48-dimensional sets. Table 2 illustrates results of this ablation study, which gives comparison performances on our proposed forecasting branch with (our proposed) or without (original) household profile branch demonstrating the effectiveness of multitask. Specifically, experiments were completed on 24 : 00, 20 : 00, 18 : 00, 16 : 00, 14 : 00, 12 : 00, 10 : 00, 8 : 00, and 4 : 00, respectively. As discussed in research [29], at different times the load profile reveals completely distinct scale, for example, in the evening higher

consumption and slight ones in morning or after midnight, which is necessary to be considered in study independently. Table 2 explicitly manifests that our proposed neural network with household profile branch has remarkable predominance from 12 : 00 to 24 : 00 on MAAPE and RMSE reflecting its great ability of learning uncertainties in different household. However, from 4 : 00 to 10 : 00 the performances of both models were close, and even sometimes the original one performed better. The main reason includes that the actual load in this span is relatively smaller and a little fluctuation in prediction could cause obvious errors in metrics. Moreover, lower load profile cannot be predicted easily, yet these errors produce minute influence on load forecasting. Between 16 : 00 and 24 : 00 load consumption behaviors of individuals are most active and filled with great uncertainty, which requires pressing demands on household load forecasting in DR significantly.

In addition, we increased additional households as training set to evaluate MCNN-HPC on the same 10 selected

TABLE 1: Shared parameters of neural networks and related selected strategies.

Parameters	CNN
Dilated rate	1, 2, 4, 8
Convolutional blocks	8
Kernels in block	24
Batch size	128
Input size	48, 336
Dropout rate	0.05
Loss function	MSE, quantile loss
Optimizing method	AMSGrad
Start learning rate	0.01
Learning rate decay	0.3
Training stop	Early stopping

households as testing set, which exploits the influence of growing dataset on the proposed model. As shown in Table 3, columns 20, 30, 50, and 100 refer to expanded training set and related percentages inspect the improvements of MAAPE and RMSE on 10 testing households. Table 3 gives results that present performance changes with the growth of training set. When 20 houses were introduced, the model served a significant improvement and 30 houses enhanced this trend, which demonstrates that expanded dataset benefits discriminative learning uncertainty of deep convolutional neural network. When the dataset was enlarged to 50 or 100 households, the performance of our model kept little improvement and became degraded gradually. It owes to the scale of tested convolutional neural network with limited trainable parameters. If blocks were increased leading to deeper network, more households could strengthen the capability of MCNN-HPC to explore regular pattern and uncertainty in individual behavior for load forecasting at household level.

4.3. Evaluation of Household Profile Branch. To verify the superiority of household profile branch, we compared our proposed network with the state-of-the-art model [37] using LSTMs to achieve predominant performance of load forecasting at household level. This model identifies different household with one-hot encoding that is concatenated by traditional feature fusion approach into load profile as additional channel of input sequence, which prompts to extract regular pattern and uncertainty. It should be noted that they adopt the method of concatenating representation vectors of different branches, which is quite different from ours. In this experiment, we randomly selected 20 houses that joined in this study. For a more detailed comparison, we chose 11 different times to evaluate both methods increasing 6 : 00 and 22 : 00. Figure 4 shows compared average performances between our proposed model and the state-of-the-art model with one-hot encoding, where MCNN-HPC had a remarkable predominance on MAAPE and RMSE. Specifically, over time when the load stays at low level, our proposed model outperformed one-hot encoding indicating more powerful capability of detecting discriminative nonlinear relationships in complicated cases. This ablation study proves more advanced mechanism of household profile

branch with advanced feature fusion process in load forecasting at household level.

In addition, we evaluated the effectiveness of feature fusion and paid more attention to structure of fusion in network. Four cases, including only Feature Fusion 1 mode, Feature Fusion 2 mode, Feature Fusion 3 mode, and our proposed strategy, were compared. Results are illustrated in Figure 1. The last one we evaluated integrates household profile encoding into each block by learnable weights, called full fusion. We randomly selected 30 households and benchmarked five cases compared to the network without household profile branch. Results are shown in Table 4, where Fusion 1, Fusion 2, and Fusion 3 provide relatively poor performances on MAAPE and RMSE. Full fusion strategy serves similar performances with our proposed model, and even at some time it performed better. However, full fusion model caused great computation cost as a result of learnable parameters explosion with increasing blocks and fully connected layers. Therefore, we preferred our proposed method of an alleviated approach to preserve the balance between effectiveness and efficiency.

4.4. Evaluation of Generalization of Our Proposed Model. To evaluate the generalization of our proposed model, we independently trained one model without household profile branch for each household. We adopted the 10 households selected in Section 3.2 and acquired well-trained 10 individual models, respectively. Then, the same 10 households were used to train and test our proposed MCNN-HPC for comparison with 10 models on average performances. Table 5 shows the results and demonstrates that our proposed model outperformed individual models on overall performances of MAAPE and RMSE reflecting the great generalization of MCNN-HPC. Meanwhile, the experiments gave more promising prospect for application in electricity market.

4.5. Evaluation of Our Proposed Model on Probabilistic Load Forecasting. Probabilistic load forecasting plays a crucial role in DR that can provide more significant information for consumer behavior analysis. In this section, 10 households selected in Section 3.2 were divided into training, validation, and testing sets to evaluate the performance of our proposed model on probabilistic load forecasting. We conducted an ablation study to compare MCNN-HPC with and without household profile branch to verify the effective encoding strategy in this area. Table 6 gives the results where our proposed items refer to the improvement ratio optimized by MCNN-HPC. It is found that our proposed model has a superior performance on Pinscore and Winkler80 at different time indicating the positive role of household profile branch on probabilistic load forecasting at household level.

In the same way, we tested 10 well-trained individual models for 10 households, respectively, the average performances of which were then compared to evaluate the generalization of our proposed model on probabilistic load forecasting. Table 7 shows the experimental results on

TABLE 2: Comparison performances of models with or without household profile branch.

Time	MAAPE		Improvement	Ratio (%)	RMSE		Improvement	Ratio (%)
	Original	Proposed			Original	Proposed		
24 : 00	0.3487	0.3291	0.0196	5.62	0.3892	0.3012	0.0880	22.63
20 : 00	0.3063	0.3203	−0.014	−4.57	0.5969	0.5903	0.0066	1.11
18 : 00	0.4782	0.4467	0.0315	6.59	0.8307	0.7949	0.0357	4.30
16 : 00	0.4499	0.4119	0.038	8.45	0.5767	0.5521	0.0246	4.27
14 : 00	0.4629	0.4605	0.0024	0.52	0.5047	0.5096	−0.0049	−0.97
12 : 00	0.4807	0.4758	0.0049	1.02	0.6657	0.6528	0.0129	1.94
10 : 00	0.4511	0.434	0.0171	3.79	0.5712	0.5776	−0.0064	−1.12
08 : 00	0.4437	0.4605	−0.0168	−3.79	0.3388	0.341	−0.0022	−0.65
04 : 00	0.3772	0.3494	0.0278	7.37	0.1058	0.1117	−0.0059	−5.58

TABLE 3: Performance improvement about certain 10 households of expanded training set with different scale relative to a training set with only 10 households.

| Time	MAAPE				RMSE			
	20 houses(%)	30 houses(%)	50 houses(%)	100 houses(%)	20 houses(%)	30 houses(%)	50 houses(%)	100 houses(%)
24 : 00	1.70	5.62	4.86	7.57	1.28	1.75	−1.32	−0.88
20 : 00	2.50	2.00	3.93	6.93	1.01	2.33	0.27	0.14
18 : 00	3.00	1.3	0.81	2.10	0.16	0.23	0.80	0.72
16 : 00	−0.34	1.29	0.27	−0.98	0.39	1.27	0.77	2.12
14 : 00	−0.11	−1.17	−0.87	−1.19	0.55	2.01	1.97	0.05
12 : 00	0.67	−0.88	0.74	−1.95	1.28	1.86	1.10	1.11
10 : 00	−0.92	1.22	2.65	4.68	1.48	0.04	−0.35	−1.51
08 : 00	3.24	3.60	3.04	4.06	4.25	4.07	3.98	3.85
04 : 00	0.03	−0.52	−5.21	−4.55	5.20	14.18	−5.91	−4.30
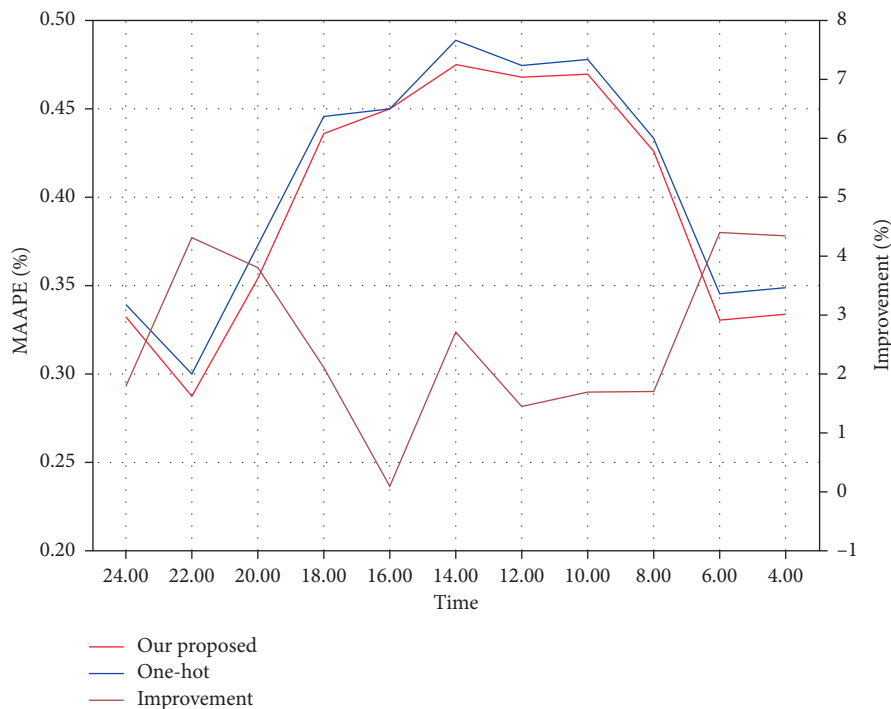
(a)

FIGURE 4: Continued.

(b)
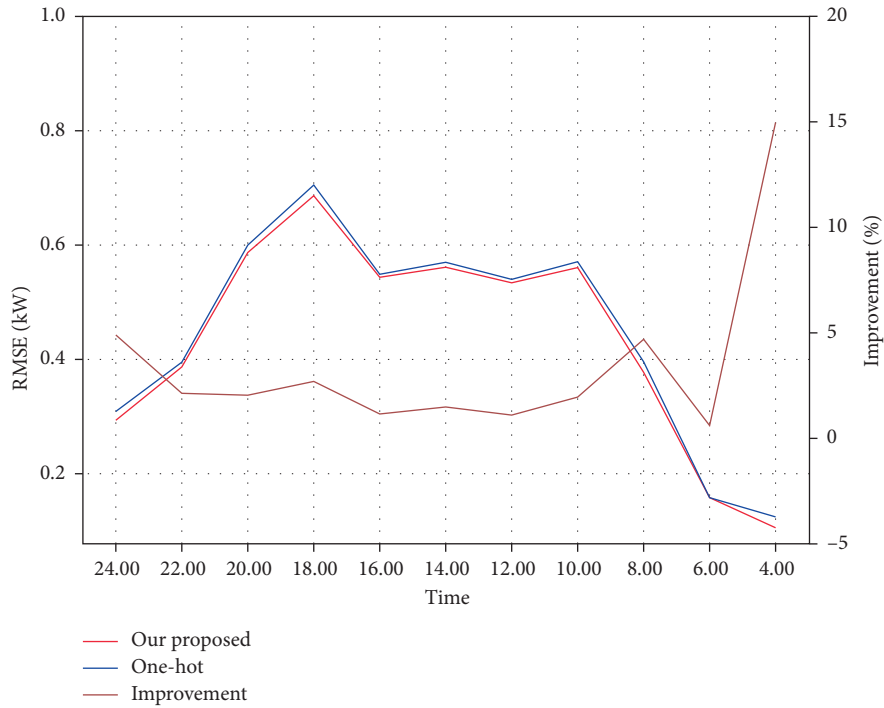
Figure 4: Comparison of our proposed model to the state-of-the-art method with one-hot encoding for 20 households. Red and blue curve denote performances in metrics of MAAPE and RMSE, respectively. Brown curve refers to improvement ratio our proposed model produced over the one with one-hot encoding.

Table 4: Performance improvement of different feature fusion strategies relative to the model without household profile branch.

Time	MAAPE					RMSE				
	Fusion 1 (%)	Fusion 2 (%)	Fusion 3 (%)	Full fusion (%)	Proposed (%)	Fusion 1 (%)	Fusion 2 (%)	Fusion 3 (%)	Full fusion (%)	Proposed (%)
24 : 00	0.42	−3.65	0.18	−1.96	−0.39	8.19	−0.46	6.76	7.00	11.27
20 : 00	1.42	2.59	1.73	2.67	2.53	2.35	0.29	2.35	2.38	1.94
18 : 00	2.82	−0.29	1.91	4.49	4.07	1.90	−0.47	2.80	3.29	3.89
16 : 00	0.48	2.50	1.53	1.58	1.32	2.33	0.15	2.54	1.54	1.98
14 : 00	−0.86	−0.73	0.73	−0.52	0.86	−0.05	0.29	0.18	0.82	−0.19
12 : 00	2.11	−2.60	0.96	0.06	0.34	1.39	−0.37	0.89	1.36	1.41
10 : 00	2.91	0.35	4.05	3.18	4.32	1.67	1.98	1.79	2.35	2.70
08 : 00	2.48	−1.38	3.77	4.44	4.83	2.86	−1.37	3.22	4.18	4.18
04 : 00	9.85	15.90	10.94	11.25	13.35	8.04	17.60	5.07	12.16	16.48

Table 5: Comparison of 10 individual models with our proposed model on average performances for generalization evaluation of point load forecasting.

Time	MAAPE		RMSE	
	Individual	Proposed(%)	Individual	Proposed(%)
24 : 00	0.3537	6.95	0.3127	3.68
20 : 00	0.3336	3.99	0.5992	1.49
18 : 00	0.4564	2.12	0.8036	1.08
16 : 00	0.4327	4.81	0.5734	3.71
14 : 00	0.4724	2.52	0.5109	0.25
12 : 00	0.4792	0.71	0.6497	−0.47
10 : 00	0.4297	−0.99	0.5739	−0.64
08 : 00	0.4642	0.8	0.3442	0.92
04 : 00	0.3793	7.88	0.1223	8.64

TABLE 6: Comparison of our proposed model with or without household profile branch.

Time	Pinscore		Winkler80	
	Original	Proposed(%)	Original	Proposed(%)
24:00	0.0766	5.49	3.5837	2.07
20:00	0.1285	3.53	4.3518	6.07
18:00	0.1834	35.59	5.9072	39.17
16:00	0.1148	4.7	3.8056	2.15
14:00	0.1208	2.61	3.4768	1.49
12:00	0.1314	4.06	4.113	3.62
10:00	0.1152	46.06	3.5079	15.2
08:00	0.0656	62.08	2.5948	22.82
04:00	0.0249	−12.58	1.8392	−6.55

TABLE 7: Comparison of 10 individual models with our proposed model on average performances for generalization evaluation of probabilistic load forecasting.

Time	Pinscore		Winkler80	
	Individual	Proposed(%)	Individual	Proposed(%)
24:00	0.0708	−2.23	3.4512	−1.69
20:00	0.1247	0.60	4.1012	0.33
18:00	0.1204	1.91	3.6731	2.17
16:00	0.1094	0.02	3.7660	1.12
14:00	0.1204	2.24	3.5536	3.62
12:00	0.1277	1.27	3.9713	0.18
10:00	0.0665	6.62	2.9948	0.67
08:00	0.0263	5.28	2.0321	1.45
04:00	0.0286	1.98	2.0254	3.25

Pinscore and Winkler80. In most times, MCNN-HPC served better accuracy over individual models on average, which provides great generalization of our proposed model in probabilistic load forecasting at household level.

5. Conclusion

This paper for the first time proposes a multitask deep neural network for load forecasting at household level. One of two branches is built on multiscale dilated convolutions for forecasting. The other branch that includes a deep convolutional autoencoder is responsible for extracting specific behavior of different household, which serves a novel mechanism of feature fusion between two branches, interpreted as a superior feature selection process leading to remarkable improvement in accuracy. We made sufficient ablation studies to verify performances of MCNN-HPC. All findings demonstrated the state-of-the-art achievement including the advancement of multitask design, the effectiveness of household profile encoding, and great generalization of our proposed model, especially in point and probabilistic load forecasting. In other words, MCNN-HPC is more competent in exploring regular pattern and uncertainty in time-series analysis. This paper focuses on providing attempting and learnings for deep learning skills for household load forecasting. Future works include designing more efficient household encoding strategy based on attention network. Moreover, more significant features like holiday or weather would prompt more advanced achievement of deep neural network remarkably.

Data Availability

The data used to support the findings of this study have been deposited in the Ireland CER repository; website is https://www.ucd.ie/issda/data/commissionforenergyregulationcer/

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

References

[1] Z. Li and Y. Tong, "Renewable energy basing on smart grid," in *Proceedings of the 2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM)*, pp. 1–4, IEEE, Chengdu City, China, September 2010.

[2] Modern Grid Initiative, *The Modern Grid Initiative: Modern Grid V2. 0 Powering Our 21st-Century Economy*, United States Department of Energy, National Energy Technology Laboratory, Washington, DC, USA, 2007.

[3] P. Siano, "Demand response and smart grids—a survey," *Renewable and Sustainable Energy Reviews*, vol. 30, pp. 461–478, 2014.

[4] C. W. Potter, A. Allison, and K. Westrick, "Building a smarter smart grid through better renewable energy information," in *Proceedings of the 2009 IEEE/PES Power Systems Conference and Exposition*, pp. 1–5, IEEE, Seattle, WA, USA, March 2009.

[5] A. Vos, "Effective business models for demand response under the smart grid paradigm," in *Proceedings of the 2009 IEEE/PES Power Systems Conference and Exposition*, p. 1, IEEE, Seattle, WA, USA, March 2009.

[6] K. Zhou and S. Yang, "Demand side management in China: the context of China's power industry reform," *Renewable and Sustainable Energy Reviews*, vol. 47, pp. 954–965, 2015.

[7] F. Saffre and R. Gedge, "Demand-side management for the smart grid," in *Proceedings of the 2010 IEEE/IFIP Network Operations and Management Symposium Workshops*, pp. 300–303, IEEE, Osaka, Japan, April 2010.

[8] Y. Wang, Q. Chen, H. Tao, and C. Kang, "Review of smart meter data analytics: applications, methodologies, and challenges," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3125–3148, 2018.

[9] Z. Deng, B. Wang, Y. Xu, T. Xu, C. Liu, and Z. Zhu, "Multiscale convolutional neural network with time-cognition for multi-step short-term load forecasting," *IEEE Access*, vol. 7, pp. 88058–88071, 2019.

[10] H. Tao and S. Fan, "Probabilistic electric load forecasting: a tutorial review," *International Journal of Forecasting*, vol. 32, no. 3, pp. 914–938, 2016.

[11] Y. Wang, Q. Xia, and C. Kang, "Secondary forecasting based on deviation analysis for short-term load forecasting," *IEEE*

Transactions on Power Systems, vol. 26, no. 2, pp. 500–507, 2010.

[12] W.-C. Hong, "Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm," *Energy*, vol. 36, no. 9, pp. 5568–5578, 2011.

[13] H. Tao and D. A. Dickey, *Electric Load Forecasting: Fundamentals and Best Practices*, SAS, Cary, NC, USA, 2012.

[14] F. L. Quilumba, W.-J. Lee, H. Huang, D. Y. Wang, and R. L. Szabados, "Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 911–918, 2014.

[15] H. Tao, Pu Wang, A. Pahwa, M. Gui, and S. M. Hsiang, "Cost of temperature history data uncertainties in short term electric load forecasting," in *Proceedings of the 2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems*, pp. 212–217, IEEE, Singapore, June 2010.

[16] H. Tao, P. Wang, and L. White, "Weather station selection for electric load forecasting," *International Journal of Forecasting*, vol. 31, no. 2, pp. 286–295, 2015.

[17] P. Wang, B. Liu, and T. Hong, "Electric load forecasting with recency effect: a big data approach," *International Journal of Forecasting*, vol. 32, no. 3, pp. 585–597, 2016.

[18] X. Sun, P. B. Luh, K. W. Cheung et al., "An efficient approach to short-term load forecasting at the distribution level," *IEEE Transactions on Power Systems*, vol. 31, no. 4, pp. 2526–2537, 2015.

[19] R. Li, C. Gu, F. Li, S. Gavin, and M. Dale, "Development of low voltage network templates—part II: peak load estimation by clusterwise regression," *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3045–3052, 2014.

[20] A. Espasa and I. Mayo-Burgos, "Forecasting aggregates and disaggregates with common features," *International Journal of Forecasting*, vol. 29, no. 4, pp. 718–732, 2013.

[21] J. Nowotarski, B. Liu, R. Weron, and T. Hong, "Improving short term load forecast accuracy via combining sister forecasts," *Energy*, vol. 98, pp. 40–49, 2016.

[22] Y. Chen, P. B. Luh, C. Guan et al., "Short-term load forecasting: similar day-based wavelet neural networks," *IEEE Transactions on Power Systems*, vol. 25, no. 1, pp. 322–330, 2009.

[23] R. Al-Otaibi, N. Jin, T. Wilcox, and P. Flach, "Feature construction and calibration for clustering daily load curves from smart-meter data," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 645–654, 2016.

[24] S. Dong, R. Li, R. Shi, and F. Li, "Analysis of the relationship between load profile and weather condition," in *Proceedings of the 2014 IEEE PES General Meeting—Conference & Exposition*, pp. 1–5, IEEE, Detroit, MI, USA, December 2014.

[25] M. Shepero, D. van der Meer, J. Munkhammar, and J. Widén, "Residential probabilistic load forecasting: a method using Gaussian process designed for electric load data," *Applied Energy*, vol. 218, pp. 159–172, 2018.

[26] A. Tascikaraoglu and B. M. Sanandaji, "Short-term residential electric load forecasting: a compressive spatio-temporal approach," *Energy and Buildings*, vol. 111, pp. 380–392, 2016.

[27] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, 2017.

[28] T. Kurniawan Wijaya, M. Vasirani, S. Humeau, and K. Aberer, "Cluster-based aggregate forecasting for residential electricity demand using smart meter data," in *Proceedings of the 2015 IEEE international conference on Big data (Big data)*, pp. 879–887, IEEE, Santa Clara, CA, USA, October 2015.

[29] Z. Deng, B. Wang, H. Guo, C. Chai, Y. Wang, and Z. Zhu, "Unified quantile regression deep neural network with time-cognition for probabilistic residential load forecasting," *Complexity*, vol. 2020, Article ID 9147545, 18 pages, 2020.

[30] C. Mohamed, "Clustering-based improvement of nonparametric functional time series forecasting: application to intraday household-level load curves," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 411–419, 2013.

[31] H. Stephen, J. Ward, D. V. Greetham, C. Singleton, and P. Grindrod, "A new error measure for forecasts of household-level, high resolution electrical energy consumption," *International Journal of Forecasting*, vol. 30, no. 2, pp. 246–256, 2014.

[32] A. Veit, C. Goebel, R. Tidke, C. Doblander, and H.-A. Jacobsen, "Household electricity demand forecasting: benchmarking state-of-the-art methods," in *Proceedings of the 5th International Conference on Future Energy Systems*, pp. 233-234, Cambridge, UK, June 2014.

[33] Y.-H. Hsiao, "Household electricity demand forecast based on context information and user daily schedule analysis from meter data," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 1, pp. 33–43, 2014.

[34] A. Marinescu, C. Harris, I. Dusparic, S. Clarke, and V. Cahill, "Residential electrical demand forecasting in very small scale: an evaluation of forecasting methods," in *Proceedings of the 2013 2nd International Workshop on Software Engineering Challenges for the Smart Grid (SE4SG)*, pp. 25–32, IEEE, San Francisco, CA, USA, May 2013.

[35] S. Humeau, T. K. Wijaya, M. Vasirani, and K. Aberer, "Electricity load forecasting for residential customers: exploiting aggregation and correlation between households," in *Proceedings of the 2013 Sustainable Internet and ICT for Sustainability (SustainIT)*, pp. 1–6, IEEE, Palermo, Italy, October 2013.

[36] B. Stephen, X. Tang, P. R. Harvey, S. Galloway, and K. I Jennett, "Incorporating practice theory in sub-profile models for short term aggregated residential load forecasting," *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1591–1598, 2015.

[37] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—a novel pooling deep RNN," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 5271–5280, 2017.

[38] G. Chitalia, M. Pipattanasomporn, V. Garg, and S. Rahman, "Robust short-term electrical load forecasting framework for commercial buildings using deep recurrent neural networks," *Applied Energy*, vol. 278, Article ID 115410, 2020.

[39] Z. Deng, C. Liu, and Z. Zhu, "Inter-hours rolling scheduling of behind-the-meter storage operating systems using electricity price forecasting based on deep convolutional neural network," *International Journal of Electrical Power & Energy Systems*, vol. 125, Article ID 106499, 2021.

[40] G. Sideratos, A. Ikonomopoulos, and N. D. Hatziargyriou, "A novel fuzzy-based ensemble model for load forecasting using hybrid deep neural networks," *Electric Power Systems Research*, vol. 178, Article ID 106025, 2020.

[41] L. Li, C. J. Meinrenken, and V. Modi, "Short-term apartment-level load forecasting using a modified neural network with selected auto-regressive features," *Applied Energy*, vol. 287, Article ID 116509, 2021.

[42] Y. Dong, X. Ma, and T. Fu, "Electrical load forecasting: a deep learning approach based on k-nearest neighbors," *Applied Soft Computing*, vol. 99, Article ID 106900, 2021.

[43] G. Dudek, P. Pełka, and S. Smyl, "A hybrid residual dilated LSTM and exponential smoothing model for midterm electric load forecasting," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2021.

[44] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Multi-sequence LSTM-RNN deep learning and metaheuristics for electric load forecasting," *Energies*, vol. 13, no. 2, p. 391, 2020.

[45] G. Dhiman and V. Kumar, "Spotted hyena optimizer: a novel bio-inspired based metaheuristic technique for engineering applications," *Advances in Engineering Software*, vol. 114, pp. 48–70, 2017.

[46] G. Dhiman and V. Kumar, "Emperor penguin optimizer: a bio-inspired algorithm for engineering problems," *Knowledge-Based Systems*, vol. 159, pp. 20–50, 2018.

[47] G. Dhiman and V. Kumar, "Seagull optimization algorithm: theory and its applications for large-scale industrial engineering problems," *Knowledge-Based Systems*, vol. 165, pp. 169–196, 2019.

[48] G. Dhiman and A. Kaur, "STOA: a bio-inspired based optimization algorithm for industrial engineering problems," *Engineering Applications of Artificial Intelligence*, vol. 82, pp. 148–174, 2019.

[49] S. Kaur, L. K. Awasthi, A. L. Sangal, and G. Dhiman, "Tunicate swarm algorithm: a new bio-inspired based metaheuristic paradigm for global optimization," *Engineering Applications of Artificial Intelligence*, vol. 90, Article ID 103541, 2020.

[50] M. Dehghani, Z. Montazeri, O. P. Malik, G. Dhiman, and V. Kumar, "BOSA: binary orientation search algorithm," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 9, pp. 5306–5310, 2019.

[51] G. Dhiman, M. Garg, A. Nagar, V. Kumar, and M. Dehghani, "A novel algorithm for global optimization: rat swarm optimizer," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–26, 2020.

[52] M. Dehghani, Z. Montazeri, Z. Montazeri et al., "MLO: multi leader optimizer," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 6, pp. 364–373, 2020.

[53] M. Dehghani, Z. Montazeri, Z. Montazeri, G. H. Guerrero, J. Guerrero, and G. Dhiman, "Darts game optimizer: a new optimization technique based on darts game," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 5, pp. 286–294, 2020.

[54] M. Dehghani, Z. Montazeri, G. Dhiman et al., "A spring search algorithm applied to engineering optimization problems," *Applied Sciences*, vol. 10, no. 18, p. 6173, 2020.

[55] G. Dhiman, D. Oliva, A. Kaur et al., "BEPO: a novel binary emperor penguin optimizer for automatic feature selection," *Knowledge-Based Systems*, vol. 211, Article ID 106560, 2021.

[56] G. Dhiman, "ESA: a hybrid bio-inspired metaheuristic optimization approach for engineering problems," *Engineering with Computers*, vol. 37, no. 1, pp. 323–353, 2019.

[57] P. Zhang, X. Wu, X. Wang, and S. Bi, "Short-term load forecasting based on big data technologies," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 3, pp. 59–67, 2015.

[58] Y. Wang, Q. Chen, M. Sun, C. Kang, and Q. Xia, "An ensemble forecasting method for the aggregated load with subprofiles," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3906–3908, 2018.

[59] M. Khodayar, J. Wang, and Z. Wang, "Energy disaggregation via deep temporal dictionary learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2019.

[60] A. Miyasawa, Y. Fujimoto, and Y. Hayashi, "Energy disaggregation based on smart metering data via semi-binary nonnegative matrix factorization," *Energy and Buildings*, vol. 183, pp. 547–558, 2019.

[61] C.-N. Yu, P. Mirowski, and T. K. Ho, "A sparse coding approach to household electricity demand forecasting in smart grids," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 738–748, 2016.

[62] Y. Wang, D. Gan, M. Sun, N. Zhang, Z. Lu, and C. Kang, "Probabilistic individual load forecasting using pinball loss guided LSTM," *Applied Energy*, vol. 235, pp. 10–20, 2019.

[63] R. Jiao, T. Zhang, Y. Jiang, and H. He, "Short-term non-residential load forecasting based on multiple sequences LSTM recurrent neural network," *IEEE Access*, vol. 6, pp. 59438–59448, 2018.

[64] W. He, "Load forecasting via deep neural networks," *Procedia Computer Science*, vol. 122, pp. 308–314, 2017.

[65] S. Wang, X. Wang, S. Wang, and D. Wang, "Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting," *International Journal of Electrical Power & Energy Systems*, vol. 109, pp. 470–479, 2019.

[66] B. Zhang, J.-L. Wu, and P.-C. Chang, "A multiple time series-based recurrent neural network for short-term load forecasting," *Soft Computing*, vol. 22, no. 12, pp. 4099–4112, 2018.

[67] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS One*, vol. 12, no. 7, 2017.

[68] O. Yildirim, R. S. Tan, and U. R. Acharya, "An efficient compression of ECG signals using deep convolutional autoencoders," *Cognitive Systems Research*, vol. 52, pp. 198–211, 2018.

[69] M. Cai, M. Pipattanasomporn, and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques," *Applied Energy*, vol. 236, pp. 1078–1088, 2019.

[70] M. Abadi, B. Paul, J. Chen et al., "Tensorflow: a system for large-scale machine learning," in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)*, pp. 265–283, Savannah, GA, USA, November 2016.

[71] S. Kim and H. Kim, "A new metric of absolute percentage error for intermittent demand forecasts," *International Journal of Forecasting*, vol. 32, no. 3, pp. 669–679, 2016.