IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 May 2024, accepted 16 May 2024, date of publication 28 May 2024, date of current version 6 June 2024.
Digital Object Identifier 10.1109/ACCESS.2024.3406439

RESEARCH ARTICLE

Enhancing Short-Term Electric Load Forecasting for Households Using Quantile LSTM and Clustering-Based Probabilistic Approach

ZAKI MASOOD, (Member, IEEE), RAHMA GANTASSI, (Member, IEEE), AND YONGHOON CHOI, (Senior Member, IEEE)
Department of Electrical Engineering, Chonnam National University, Gwangju 61186, South Korea
Corresponding author: Yonghoon Choi (yh.choi@jnu.ac.kr)

This work was supported in part by the Ministry of Education, Republic of Korea; and in part by the National Research Foundation of Korea under Grant NRF-2023S1A5C2A07096111.

ABSTRACT Electricity load forecasting is an essential part of power system planning and operation, and it is crucial to make accurate predictions. The smart grid paradigm and the new energy market necessitate better demand-side management (DSM) and more reliable end-user forecasts to system scale. This paper proposes a time-series clustering-based probabilistic electricity future prediction for short-term load forecasting (STLF), which makes forecasts more accurate and intelligent. The weather and data noise uncertainties are considered, with the load probabilistic interval as the model's output for individual and aggregated household energy consumption. This paper adapts the logarithm of the hyperbolic cosine (log-cosh) of the error value as quantile loss and time feature with long short-term memory (LSTM) to bridge the gap. This paper presents a framework for probabilistic electric load forecasting that incorporates clustering-based quantile-LSTM learning to improve the accuracy and robustness of short-term electricity prediction. The proposed model is primarily applied to energy demand information on 15-minute and 1-hour time horizons for day-ahead prediction tasks, which are the key concerns for electricity utilities. Various state-of-the-art regression learning techniques, i.e., quantile regression forest (QRF), quantile regression neural network (QRNN), quantile gradient boosting regression tree (QGBRT), and quantile LSTM (Q-LSTM), were utilized and evaluated. The findings indicate that the proposed cluster-based probabilistic forecasting approach outperforms the existing benchmark models in terms of prediction interval coverage probability (PICP) evaluation metric.

INDEX TERMS Deep regression learning, electricity load clustering, probabilistic forecasting, short-term load forecasting, smart grid.

I. INTRODUCTION

The The modern electrical system is a dynamic system that balances the production, transmission and consumption of electricity. This balance between supply and demand must be maintained at all times in order to ensure the reliability and stability of this system. In recent years, energy forecasting applications have been developed not only on the grid side of electricity networks but also on the grid end. These applications aim to balance between load and demand [1]. The growth in market competition and smart grid integration requirements has suggested that electricity price forecasting has become crucial for demand side management (DSM). Additionally, electricity load forecasting plays a significant role in addressing these challenges. The majority of early studies and current research in electricity load forecasting have been focused on probabilistic approaches [2], [3]. These approaches aim to estimate the entire probability distribution of time series data instead of only point forecasting. In recent years, probabilistic forecasting techniques have become

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

increasingly important in decision-making processes related to maintenance, planning, and future investments for both renewable [4] and non-renewable energy technologies [5]. A detailed discussion of a comprehensive, integrated solution incorporating probabilistic load forecasting has been presented [6]. This approach enables energy schedulers to gain insights into the role of critical variables in the entire distribution of energy consumption and explore policy-making. For instance, the increasing spread of intermittent renewable energies, particularly solar and wind, is leading to changes in the operation of the power grid. It allows monitoring of energy consumption and finding peak demand, reduces losses, mitigates risks, and ensures energy reliability for continuous operation [7], [8]. Moreover, building energy management system (BEMS) [9] and home energy management system (HEMS) [10] contributes to the effective use of electricity.

Over the last few years, various machine learning (ML) and deep learning techniques have been developed to estimate electricity load demand on short, medium, and long time scales. These forecasting methods can mainly be categorized into short-term load forecasting (STLF), medium-term load forecasting (MTLF), and long-term load forecasting (LTLF), which are based on statistical and artificial intelligence (AI) algorithms. On the other hand, deep learning techniques have been progressively developed and used for energy forecasting in the context of the smart grid. Among all, recurrent neural network (RNN), long short-term memory (LSTM) [11], convolutional Neural Networks (CNN) [12], and gated recurrent units (GRU) [13] are four recent popular models. It is also noteworthy that LSTM has been used to forecast building energy usage [14], weather forecast [15], and time series predictions [16]. For the STLF time-series problems, a detailed review of utilizing recurrent models was undertaken [17]. The outcomes demonstrated that the LSTM and GRU are superior to RNN at holding temporal dependency in time series problems.

On the other hand, a hybrid approach has been proposed in [18], which comprises a generalized extreme learning machine (ELM) to train an improved wavelet neural network (NN). Flexible computing frameworks and universal approximators are features of artificial neural networks (ANN) that enable it to address forecasting issues in a variety of sectors with great accuracy. According to previous research, such ANN technique performs more effectively in short-term forecasting than in long-term forecasting. Time is one of the most crucial variables for the demand response program. The time-series data conveys visual information about the customer's unpredictable load consumption nature, which we aim to perform cluster and forecast the load profiles. The authors of [19] determined that ANN is more effective at predicting short-term electricity consumption than long-term forecasting. On the other hand, authors in [20] review typical ANN techniques to anticipate short-term load demand and confirm that this technique has various advantages; the other is the accuracy of STLF.

II. LITERATURE REVIEW

In the field of regression learning, researchers have been investigating new ways with the primary focus to improve predictive models. To put it simply, statistical forecasting refers to the use of time-series analysis based on historical data to estimate something that may happen in the future. However, the analysis in [6] showed that the existing algorithms could perform well when they dealt with probabilistic forecasting instead of point forecasting. In addition, it has been shown that using the normality hypothesis can enhance the quality of probabilistic forecasts. For example, Wang et al. [21] models the conditional forecast residual using point load predictions and quantile regression with probabilistic forecasting. Specifically, their proposed model can be used for probabilistic-based power supply planning in the time-series problem. Zhang et al. [22] investigated probabilistic forecasting for the normal and the peak abnormal differential (PAD) load that utilizes probabilistic occurrence and magnitude forecasting of the PAD load, respectively. Moreover, a practical methodology to develop probabilistic forecasting has been proposed using quantile regression averaging (QRA) to a set of sister point forecasts [23]. Their methodology has significant practical value because it can leverage existing point load forecasting and does not depend on high-quality expert forecasts. The load curves, particularly at low to medium power consumption distribution levels, are complex and challenging to approximate using traditional methods. Therefore, deep learning is a promising solution for accurate load forecasting due to its impressive ability to handle nonlinear patterns and establish complex links between input and output. Furthermore, the unpredictability of the power consumption of the end-users of electricity is also due to their different priorities, affordability, varying social needs, and different lifestyles. The authors in [24], highlight the significance of clustering similar energy-profile customers together to make training databases for each cluster. It has now been suggested [25] that clustering improves the accuracy of load forecasting in the deep learning models for clients with highly unpredictable behavior.

A more extensive description can be found in a review [6], where new challenging and substantial areas of probabilistic load forecasting have been discussed in detail. Dahua [26], a hybrid model adopting temperature and load deviation to generate probabilistic forecasting. The authors proposed ANN probabilistic electricity forecasting with quantile optimization, considering the randomness of inputs and the output variation. In [27], the constrained weighted quantile loss (CWQLoss) for supervised regression has been proposed. The feature learning blocks are fully connected and convolutional layers and are trained end-to-end using stochastic gradient descent, which are the building blocks of the architecture. Numerical evaluation on a dataset reveals that this method outperforms others in prediction training time. Specifically, models like the quantile regression neural network (QRNN), quantile regression forest (QRF), quantile gradient boosting regression tree (QGBRT), and the quantile

LSTM (Q-LSTM) [28], [29], [30], [31], [32], [33] have been developed and are often utilized for probabilistic load forecasting. For example, [34] proposed an additive quantile regression model for a set of quantiles that forecasts uncertainties in electricity smart meter data using a boosting procedure. However, a number of questions regarding quantile regression for probabilistic load forecasting remain to be addressed.

Most early studies focus on QRNN and LSTM for probabilistic load forecasting. For instance, feature extraction by the minor absolute shrinkage and selection operator (LASSO) regression and QRNN model has been proposed for electricity consumption forecasting [28]. A shortcoming of QRNN is the computational time in the training model, which increases the training cost with large datasets. This problem has been addressed in [30], where batch training, early stopping, reducing training cost, and dropout are incorporated with improved QRNN (iQRNN). Another problem with such a QRNN tends to be overfitting. Specifically, over-fitting in electrical load forecasting leads to a model that performs well during training but exhibits low performance during validation. A case study approach solution in [35] highlights probabilistic forecasts with deep learning and compares the prediction intervals for GRU and LSTM. Their research primarily focuses on renewable sources for future power generation.

However, there is a research gap regarding their limited accuracy in probabilistic forecasting due to the incomplete capture of the probability distribution of the data.

This paper utilizes clustering-based probabilistic prediction on LSTM stacks with log-cosh quantile loss function to improve the forecasting accuracy of an individual and different household. Probabilistic electricity load forecasting is more complex than point forecasting because it forecasts the entire conditional distribution of future observations, not only the conditional mean. This approach integrates clustering and probabilistic electric forecasting using LSTM and log-cosh loss function in the training and testing process. Furthermore, exploratory data analysis and feature importance are examined in individual households. Specifically, we consider STLF as a typical and challenging time-series problem and present a comprehensive evaluation of the theoretical framework and existing solutions. The main contributions of this paper are summarized as follows:

1) Propose a framework for probabilistic electric forecasting, which utilizes clustering-based quantile-LSTM learning to find an adequate solution to electricity forecasting. The log-cosh loss function has been used as a model optimization parameter in training and testing.
2) In our study, we utilized two publicly available real-world datasets: [36] and [37]. The proposed framework is then implemented in a statistical test to see any substantial difference in performance error. The implementation of four cutting-edge, QRF, QRNN, QGBRT, Q-LSTM, and the proposed framework, is analyzed and compared to find the best solution to solve the problem.
3) The implementation of state-of-the-art memory cells, such as RNN, LSTM, and encoder-decoder architecture, was analyzed and compared while clustering. In particular, evaluate cluster data and later trains separate regressors, which results in more relevant data groups fitted to probabilistic forecasting.

Our work presented in this paper will promote probabilistic electricity load forecasting for smart and sustainable energy networks.

The rest of this paper is organized as follows. The following Section III describes problem formulation and proposes clustering-based LSTM learning along with log-cosh quantile optimization for probabilistic electric forecasting. The experiment setup and evaluation metrics are described in Section IV. Simulation results and the advantage of the proposed method are presented in Section V, followed by the conclusion in Section VI.

III. PROPOSED METHODOLOGY

The proposed framework for electricity forecasting is depicted in Fig. 1, comprising of four key steps: data preprocessing, data clustering, training and testing, and probabilistic forecasting. The initial stage involves extracting useful information from the household's electricity consumption dataset, which requires exploratory analysis to identify any missing information in the dataset. We resampled data from one-minute and 15-minute intervals to one-hour intervals to reduce training time and test the model's accuracy. The details on data preprocessing will be given in the following Section III. The second step, clustering, is a critical aspect in the analysis and mining of multivariate time series data for prediction applications, as highlighted in previous studies [38]. The predictive model should be able to forecast demand using minimum available historical data [39]. Therefore, clustering was utilized to capture short-term hidden data patterns in the original dataset. In addition, utilities, in particular, frequently have very little information about their customers other than aggregate BEMS level loads, with no knowledge of interior details about household appliance groups [40]. Therefore, time-series based individual and different household data are trained based on the electricity demand patterns for each cluster group 1 to K in steps three and four. Finally, prediction accuracy is evaluated at a 15-minute and one-hour time horizon.

A. ENCODER-DECODER LSTM

The Encoder-Decoder LSTM model is a RNN version that can deal with the sequence of data to forecast sequence-to-sequence (Seq2Seq) prediction problems. The Encoder-Decoder network models are more commonly used to time-series forecasting, language process and text summarizing, where Seq2Seq prediction is important [41]. The

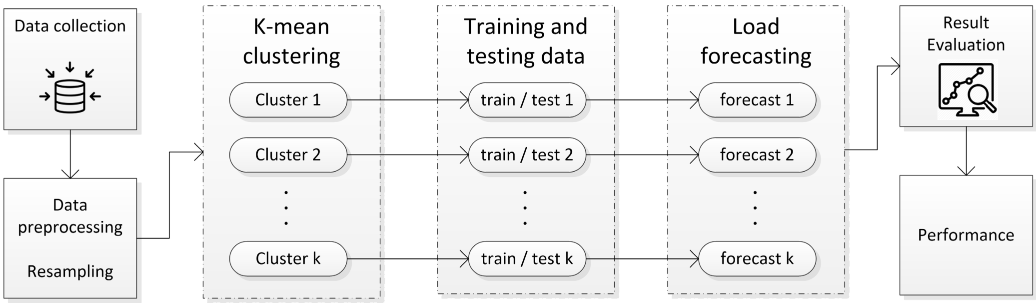


FIGURE 1. Overview of the proposed approach for probabilistic forecasting of household electricity consumption.

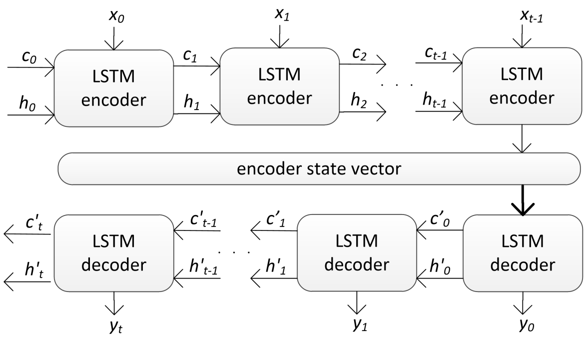


FIGURE 2. An architecture of encoder-decoder LSTM model.

architecture of an Encoder-Decoder LSTM [42] is shown in Fig. 2, where each rectangular block holds an LSTM Cell. The network architecture mainly consists three parts: LSTM encoder, encoder state vector and LSTM decoder. One can easily see that the encoder part is stacked with LSTM cells, and each of the cells allows a single element from the sequence as input. Conclusively, h_t is the encoder vector, the final hidden state produced by the encoder.

On the other hand, the decoder generates the output at each LSTM cell. The vector forwards the information for all input elements to the decoder for the predictions. The mathematical equations for the Encoder-Decoder model are represented in (1) and (2), where $x_t = \{x_1, x_2, \ldots, x_T\}$ and $y_t = \{y_1, y_2, \ldots, y_{T'}\}$ are the input time sequence and targeted output sequence of the forecasting model, respectively. The decoder is a stack of LSTM cells, where each cell predicts an output. The vector h_t is the initial state for the decoder part. To determine the output, at each LSTM cell, the state of the previous layer and output from the last LSTM cell supplies as input. Thus, the last hidden state, it computes the output of the next hidden state. A symbol y_t represents the final output at a time step t. The importance of the Seq2Seq model lies in that it can outline sequences of diverse lengths to each other. In other words, the size of the input and output sequence is different. The next decoder generates an output sequence, at each step taking at time t, the previous state, and a weighted combination of all the encoder outputs (i.e., encoder state vector).

$$h_t = \text{LSTM}_{encoder}\left(x_t, h_{t-1}\right), \tag{1}$$

$$h'_t = \text{LSTM}_{decoder}\left(y_t, h'_{t-1}\right). \tag{2}$$

B. QUANTILE LSTM BASED FORECASTING FRAMEWORK

In this section, we illustrate probabilistic electricity load forecasting with a quantile loss function. The framework of the proposed probabilistic forecast process has been summarized in Fig. 3, which consists of four major components. These components include a stack of LSTM cells, time feature one-hot coding, concatenation, and fully connected quantile output. In the first step, preprocessed cluster data is fed into a stack of LSTM cells as time series training and testing groups. Since, the available data recorded from different households and from different smart meters in case of a individual household electric power consumption (IHEPC), which is discussed in detail in Section IV. Therefore, preparing the training data according to the LSTM requirements is critical for building and training the model. The x_{t-m} represents the input data at time t, number of time period ahead m. Next, one-hot encoding is adopted to represent numerical data as a discrete element of time feature input $\left(F_{hourly}, F_{weekday}, F_{holiday}, \text{and } F_{weekday}\right)$, weather and smart meter feature importance in case of IHEPC. The final step is the fully connected output which performs a quantile prediction interval based on the learning features. Eq. (3) shows quantile loss function, which penalizes the loss according to the spercentile also called the pinball loss function [43].

$$L_{t,q}\left(y_t, \hat{y}_t^q\right) = max \begin{cases} q\left(y_t - \hat{y}_t^q\right), & y_t \geq \hat{y}_t^q, \\ \left(\hat{y}_t^q - y_t\right)\left(1-q\right), & \hat{y}_t^q > y_t, \end{cases} \tag{3}$$

where y_t is the true value and $\hat{y}_t^q$ represents the quantile forecast. Whereas q is the regression loss, which ranges between 0 and 1. As a result, predicting a percentile within a range of certain upper and lower bound percentages. Thus, if the percentile is smaller, greater, or equal to the probabilistic prediction $\hat{y}_t^q$, penalize loss.

In the process of LSTM mapping inputs to outputs, optimizers determine the value of the weight parameters that minimize the error as already been discussed in detail in Section III. Specifically, optimizers have a significant impact on the model's accuracy. One primary problem with the gradient-descent method is that the loss function is required to be twice differentiable. However, the original quantile loss in (3) is the non-differentiable loss function, which makes the gradient-based optimization approach extrinsic.

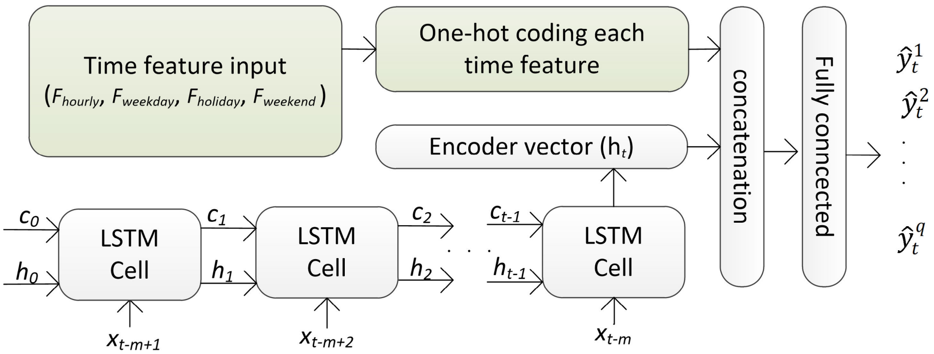


FIGURE 3. Architecture of the proposed quantile-based LSTM network model for the probabilistic electric forecasting.

Here, we adopt the log-cosh as a loss function, which is the logarithm of the hyperbolic cosine of the error value. The log-cosh loss function is a sort of estimator that picks relative solutions to the median rather than the mean. Furthermore, the log-cosh loss helps to find a better balance between reconstruction and latent space optimization, which improves accuracy performance [44]. Specifically, such estimators are more tolerant of outliers in the dataset, which is one of the main reasons the log-cosh loss function is preferred over others. Logcosh loss function for neural network aims to combine the benefit of not overweighting outliers with MSE of continuous derivative near the mean. The log-cosh loss function is defined in (4) and (5), where a is the parameter that regulates smoothness.

$$f\left(t;a\right)=\frac{1}{a}log\left(cosh\left(at\right)\right)=\frac{1}{a}log\left(\frac{e^{at}e^{-at}}{2}\right), \tag{4}$$

$$L\left(t;a\right)=\frac{1}{a}\sum_{t=1}^{T}log\left(cosh\left(a\left(y-y_t^i\right)\right)\right), \tag{5}$$

This formula is similar to Huber loss [45], so it has all of the advantages of Huber loss. The most important benefit of this method is that it performs well compared to the Huber loss function. For instance, robust to outliers and having a variable gradient. Thus, for our proposed probabilistic model, the log cosh function becomes

$$L_{t,q}(y_t,\hat{y}_t^q)=\begin{cases}\dfrac{log(cosh(qa|y_t-\hat{y}_t^q|))}{a}, & y_t\geq\hat{y}_t^q,\\ \dfrac{log(cosh(qa|\hat{y}_t^q-y_t|)(1-q))}{a}, & \hat{y}_t^q>y_t,\end{cases} \tag{6}$$

where y_t is the true value and $\hat{y}_t^q$ is the predicted value at quantile at time t. Let us take the advantage of log-cosh function, which also combines the benefit of MAE and MSE by being approximately to $12(y_t - \hat{y}_t^q)$ when the error is small, and to $|y_t - \hat{y}_t^q| - log2$ when differences are large. It has been established that hyperparameter $'a'$ regulates the smoothness and tuning while learning to allow switching between different loss functions [46]. Therefore, during the training phase the learning network adopted the log-cosh loss function.

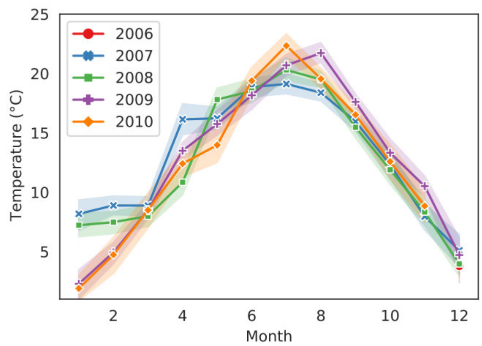


FIGURE 4. Average yearly temperature data (in degrees Celsius) from 01 January 2006 to 31 December 2010.

By varying the value of q, one can easily obtain a total of Q quantiles of the electric load over time T. As a result, the average quantile loss represents the quantile forecasting model's overall performance as

$$\overline{L}_{t,q}=\frac{1}{T.Q}\sum_{t=1}^{T}\sum_{q=1}^{Q}L_{t,q}\left(y_t,\hat{y}_t^q\right). \tag{7}$$

IV. EXPERIMENT SETUP AND DATASETS

This section describes the experimentation and clustering process, which includes the exploratory dataset analysis and later implementation in detail. In this paper for the experiment, we use two publicly available electricity consumption datasets one is from France, and the other is from Portugal. These data sets are useful because they are based on real-world datasets, which reduce the risk and avoid assumptions. These datasets are available at the UCI machine learning repository [49]. The first is the IHEPC dataset [36] of a household contains 2075259 measurements at a sampling rate of one-minute from Sceaux, France. The second is a UCI electricity load diagram [37] dataset from 2011 to 2014 consists of the electricity consumption of 370 clients with a sampling period of 15 minutes. Furthermore, MinMax is used to normalize the dataset, which ranges between [−1,1]. In preparing the data for the learning model, we substitute missing values with the available power measurement of neighboring consumption data. For example, to avoid the effect of missing values, the authors in [50],

TABLE 1. The attribute and description of IHEPC [47] and weather [48] Datasets.

No.	Attributes	Description	Min	Max
1	Date	Date in format dd/mm/yyyy	16/12/2006	26/11/2010
2	Time	time in format hh:mm:ss	17:24:00	21:02:00
3	GAP	global active power one-minute sampling rate over a period in (kW)	0.076	11.122
4	GRP	global reactive power one-minute sampling rate over a period in (kW)	0.000	1.390
5	Voltage	voltage one-minute sampling rate over a period in (V)	223.200	254.500
6	GI	global intensity one-minute-averaged current intensity in (A)	0.200	48.500
7	SM_1	sub-metering 1 from dishwasher, oven and microwave in (kWh)	0.000	88.000
8	SM_2	sub-metering 2 from washing-machine, tumble-drier,	0.000	80.000
		refrigerator and light in (kWh)	0.000	80.000
9	SM_3	sub-metering 3 from electric water-heater and air-conditioner in (kWh)	0.000	31.000
10	Rain	rain prediction in (mm)	0.000	44.800
11	t_{max}	maximum temperature in a single day in °C	-3.900	35.600
12	t_{min}	minimum temperature in a single day in °C	-8.900	22.600

[51], and [52], replaced these invalid measurements with observations at the same time as the previous day, and the average values of energy consumption. We employed linear interpolation to fill in missing values in a time series by estimating them based on neighboring data points with the same hour and minutes day of the week. Therefore, a linear relationship between values and calculates an interpolated value for the missing point and does not significantly impact the training of the learning model.

A. REAL-WORLD DATASET FOR THE ELECTRICITY LOAD FORECASTING

The IHEPC electricity load was measured for the 47 months between 16 December 2006 to 26 November 2010. The application employs data obtained at a sampling rate of one-minute from different smart metering SM_1, SM_2 and SM_3. The original data is the multivariate time series with nine attributes and its description is presented in Table 1 (first nine rows with maximum and minimum values). This is crucial to prevent high maximum demand charges. Furthermore, the complexity has been further enhanced by providing additional data of [53], which is at 7km distance from the Paris, which includes daily weather attributes Rain, t_{max} and t_{min} between 16 December 2006 to 26 November 2010. The description of weather data is also represented in Table 1 (last three rows with maximum and minimum values). For the visualisation, Fig. 4 represents the maximum and minimum temperature effects on the household electricity daily consumption from 2006 to 2010. It shows that the household need for heating when the temperature drops. As a result, there is a rise in the electricity demand. Additionally, when the temperature rises, cooling demands lead to an increase in electricity demand. Thus, the weather impact and the load demand in a region are likewise closely correlated and statistically significant. On the other hand, in the case of different households [37] there are 140256 number of sample points with electricity consumption measured in kW.

All datasets were preprocessed, which included outlier and missing value detection and sub-replacement. To some extent, the use of the daily energy consumption approach is rational. However, energy consumption behaviors differ significantly among days due to the cyclicality of productive forces as shown in Figs. 5. Consequently, electricity power consumption characteristics on relatively weekly scales cannot be extracted from typical daily load curves. As stated in Section III, the inclusion of time features such as weekly and weekends in the dataset is crucial. For the purpose of comparison, the representative profile for the household each year can be summarized using two boxplots: one for weekdays and another for weekends, as shown in Fig. 5. Weekends impact of energy used and the patterns are typically relatively different from weekdays. The line plot and box plot confirms the distribution of electricity load on weekdays and weekends of the IHEPC dataset. Across all years, weekend consumption is higher than weekday consumption. Next, the clustering technique is applied to the average electric power consumption on both datasets and feature analysis of the IHEPC dataset.

B. ELECTRICITY LOAD CLUSTERING

In this section, we conducted an analysis of data aggregations and observed separability within the aggregated data. Many clustering techniques have been used in the electricity industry [54], which can separate groups of customers and smart meter data, making it easier for prediction models to extract patterns in data. In the case of both IHEPC and different customer data, we adapt the K-means method [55], the cluster heads are chosen based primarily on Euclidean distances and the residuals. Specifically, the center node collects the identifier, residual energy, and location of all data and saves it in a list [56]. The clustering algorithm is applied to similar combinations to cluster household data. Initially, K was unknown; therefore, the silhouette score found an interpret consistency within data clusters to determine the optimal number of groups. The distance between the resulting clusters can be investigated using silhouette analysis. Eq. (8) represents the silhouette score that confirms the data fitness as the optimal number for the clustering changes.

$$S_i = \frac{b_i - a_i}{max(a_i, b_i)}, \tag{8}$$

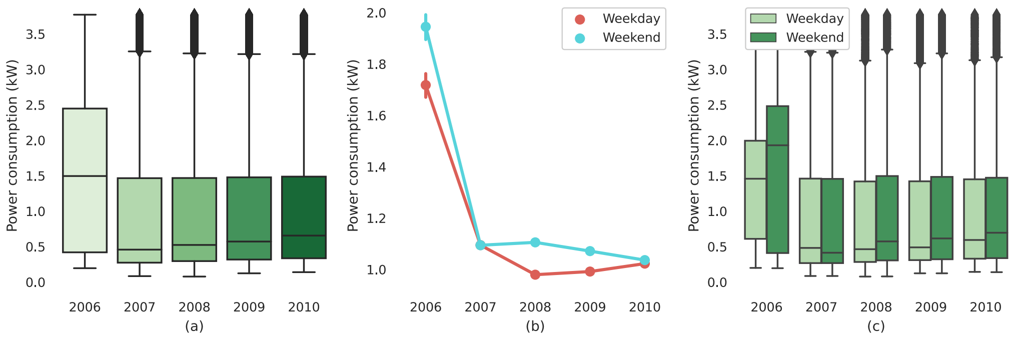


FIGURE 5. Average electricity usage comparison in weekdays and weekends from 16 December 2006 to 26 November 2010. Weekday loads have a more elevated phase of divergence than weekend loads.

where a_i and b_i are the average distances of a data point from all other points in the corresponding and nearby clusters, respectively. The silhouette score reveals how close each point in one set is to points in neighboring groups, thus allowing for the selection of the number of clusters [57]. Similar to this, silhouette analysis and the clustering results are evaluated in order to cluster households on their energy patterns and appliance of electricity use. The clustering steps are summarized as

- **Step-1:** choose the random value of K for the number of clusters.
- **Step-2:** random selection of number of centroids.
- **Step-3:** find Euclidean distance among data points and the centroids.
- **Step-4:** according to the distance assign each closer data point to the nearest cluster.
- **Step-5:** estimate the mean within each cluster and update the new centroid of each cluster.
- **Step-6:** repeat 3, reallocate datapoints to the new closest centroid.
- **Step-7:** repeat 2 to 7 until get the lowest sum of variance (variance of each cluster).

The illustration of our clustering results for the IHEPC and different household data can be seen in Fig. 6 and Fig. 7, respectively. Fig. 6 shows different cluster patterns of IHEPC data, the t-SNE representation of clustering results from different smart metering data, and the best cluster number is K = 4. Therefore, household data is divided into four groups, and centroids load profiles for different patterns from all sub-metering data. On the other hand, to confirm the clustering results for different households, Fig. 7 shows the t-SNE representation of the cluster data, where the loads in clusters 2, 3, and 5 are close to other loads in the same cluster. The few confused and intersecting seeds indicate that the loads in clusters 2 and 5 have regular load patterns and can be classified accurately. However, because the loads in cluster 3 are dispersed, forecasting the loads in cluster 3 is difficult. From the different pattern combinations, it can be seen the optimal number of clusters is K = 6. Next, the proposed clustering framework forecasts the load data for each cluster based on the train and test subsets.

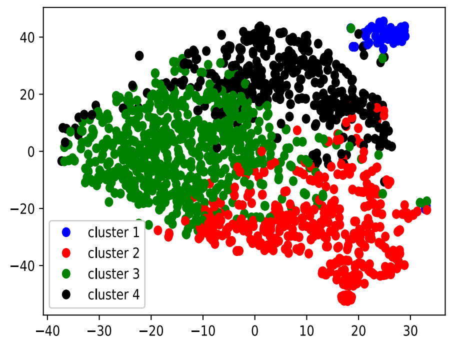


FIGURE 6. t-SNE of IHEPC with a number of clusters K = 4.

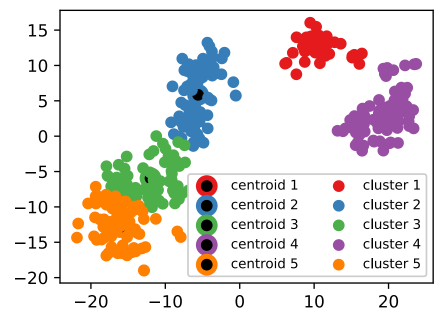


FIGURE 7. t-SNE of the different customers with a number of clusters K = 6.

In regression analysis, the mean squared error (MSE), mean absolute error (MAE), root mean squared error (RMSE), and mean absolute percentage error (MAPE) metrics are used to assess the model's performance [58], [59]. Therefore, cluster numbers are verified to reflect the variety of the load profiles adequately using the average MSE, MAE, RMSE, and MAPE results with testing datasets. The clustering framework is measured against the state-of-the-art deep learning schemes to confirm the effectiveness of forecasting as shown in Table 2. It shows that the average MSE, MAE, RMSE, and MAPE values are high compared to the cluster-based training of LSTM and its cell variants.

The selection of hyperparameters has a significant impact on the performance of an algorithm. After a different number of trails, fine-tuned parameter settings are represented in Table 3. The results have been obtained over both datasets

TABLE 2. Point forecasting performance comparison of IHEPC data with and without clustering on 1-hour time horizon.

Datasets	Forecasting Models	MSE	MAE	RMSE	MAPE%
IHEPC	RNN	0.453	0.434	0.673	55.66
	LSTM	0.417	0.408	0.646	51.24
	Seq2Seq LSTM	0.398	0.387	0.630	42.91
	LSTM (clustering)	0.322	0.331	0.567	38.44
different customer	RNN	0.0056	0.0057	0.075	13.72
	LSTM	0.0040	0.0043	0.064	12.64
	Seq2Seq LSTM	0.0034	0.0038	0.059	11.86
	LSTM (clustering)	0.0029	0.0033	0.054	7.75

TABLE 3. Paramters settings for the training model.

Parameters	Value
Forecasting model	LSTM and its variant cells
Training dataset	70% of the total input
Testing dataset	30% of the total input
Validation split	25% of the train dataset
Minmax normalization	−1 to 1
Number of units/sequence	16, 32, 128
Batch size	20 to 512
Regularization	dropout 0.2 each layer
Early Stopping	Yes
Number of maximal epochs / tree	150 / 500
Optimization algorithm	Adam
Loss function	log-cosh quantile

and tested during the testing period on a 15-minute and 1-hour time horizon. As an example, Table 2 illustrates the forecasting results for a 1-hour time horizon, comparing the outcomes with and without clustering. For instance, the RMSE results of the IHEPC dataset are 15.75%, 12.23%, and 10.0% reduced compared to RNN, LSTM, and Seq2Seq LSTM forecasting models without clustering, respectively. For the case of different customers, the results of Table 2 show that using substantial customers separately has a positive impact on forecasting errors. The following section shows a numerical evaluation for the proposed quantile-LSTM-based probabilistic forecasting on these clusters from both datasets.

V. NUMERICAL EVALUATION

In this section, numerical evaluations are provided to validate the performance of the proposed method with other state-of-the-art learning models. The parameter settings for the proposed method and comparative learning models can be seen in Table 3. Additionally, the table illustrates the parameter settings for PICP and the allocation proportion of these settings during multiple estimator trials conducted while the model was running over 15-minute and 1-hour horizons. The tests are run on a single NVidia GPU to accelerate computations. We use Tensorflow as the backend, Gluonts and the Keras deep learning application programming interface (API) to forecast the models in the Python environment [60], [61], [62]. The maximum number of train epochs used for all models in the experiment is set to 150, whereas the number of trees is set to 500 for tree methods. We incorporate early stopping, which prevents the model from overfitting. Each cluster datum is divided into 70% and 30% training and testing subsets, respectively. The allocation proportion of the parameter settings for the model performed during the model running for several estimator trials. We observed both the training and validation loss decrease during the training process. Our goal is to find a balance that allows the model to learn patterns from the training data while also providing a potent evaluation of the testing data. A further 25% of the training data is used to validate the experiment. As discussed log-cosh function in Section III, it is used to find quantile scores with LSTM learning on the performance of obtained data with the validation dataset. Thus, the forecasting performances are evaluated using the overall log-cosh loss function with 50% and 90% prediction intervals.

The feature importance score refers to techniques that assign input features according to how well they are able to predict a given target variable [63]. Since the feature importance and quality have an impact on the performance of the proposed model. Therefore, Fig. 8 is driven by extracting and selecting features, which includes component and correlation analysis of IHEPC data of Table 1. Specifically, represents all the input variables and the input variable's global feature importance score. The temperature values used to input the model are calculated as the mean of temperature values from the original dataset. Moreover, the lag selection is used to find the most relevant input variables in a dataset. The 96-hour lag selection is used to predict future behavior that shows a powerful positive autocorrelation in GAP. One can easily observe that variables GAP-1 is 0.34, SM_3 is 0.24, and GRP is 0.08, which significantly impact the power consumption patterns. After feature selection, the model tends to become more interpretable. This differentiation allows our model to adapt to the specific dynamics of short-term variations such as hourly, daily, and weekly patterns, providing a context-aware forecasting capability. Ignoring such temporal dynamic features results in failing to capture the nuances of changing patterns and predictions.

Nevertheless, t_{max} is another critical factor influencing the electrical behavior of residential households. For example, if t_{max} drops, the air-conditioning system will be activated to meet comfort demands, which increase SM_3 electricity consumption. It is important to note that the time scale represents a consistent customer behavioral pattern in the residential load profile. For this purpose, the feature importance score

TABLE 4. Performance comparison of PICP associated to all scenarios with time horizon for the day ahead forecasting.

Datasets	Forecasting Models	15-minute horizon		1-hour horizon	
		PICP-50%	PICP-90%	PICP-50%	PICP-90%
IHEPC	QRF	49.66	88.38	48.45	86.17
	QRNN	48.64	89.65	50.32	86.88
	QGBRT	51.54	90.87	49.76	87.35
	Q-LSTM	52.53	91.90	51.14	91.65
	Proposed	54.55	94.26	53.44	92.74
Different Customer	QRF	49.66	88.17	47.45	87.64
	QRNN	50.22	90.42	47.97	88.25
	QGBRT	51.36	91.39	49.38	90.37
	Q-LSTM	50.91	92.77	50.37	91.47
	Proposed	52.55	93.13	51.64	92.85

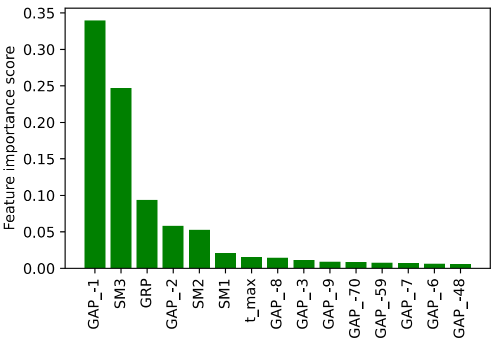


FIGURE 8. The feature score of top selected features.

is only represented for GAP, GRP, t_{max}, and smart meters (SM_1, SM_2 and SM_3) variables in Fig. 8. We use the widely accepted aspects to emphasize the notion of the proposed method. Consequently, electricity consumption series over the time intervals and the corresponding hours over the previous 96 hours, smart metering data, temperatures over the last 96 hours, the hour of the day, the day of the week, and the distinction between weekdays and weekends have been considered. As a result, these feature contributions data significantly reduce the dimension of input data and speed up calculation. Next, according to the cluster results of IHEPC in Section IV, the load and temperature data are used to train the model.

The assessment was carried out using the prediction interval coverage probability (PICP) function indicator with prediction intervals of 50%, and 90%, which is discussed in detail. The PICP in (9) has been adopted to construct prediction intervals to quantify the forecast process. The PICP evaluates the prediction interval accuracy, which shows the percentage of targets within the upper and lower bounds. The greater the PICP value, the more targets fall within the prediction interval.

$$\text{PICP} = \frac{1}{N}\sum_{t=1}^{N}\mu_t, \tag{9}$$

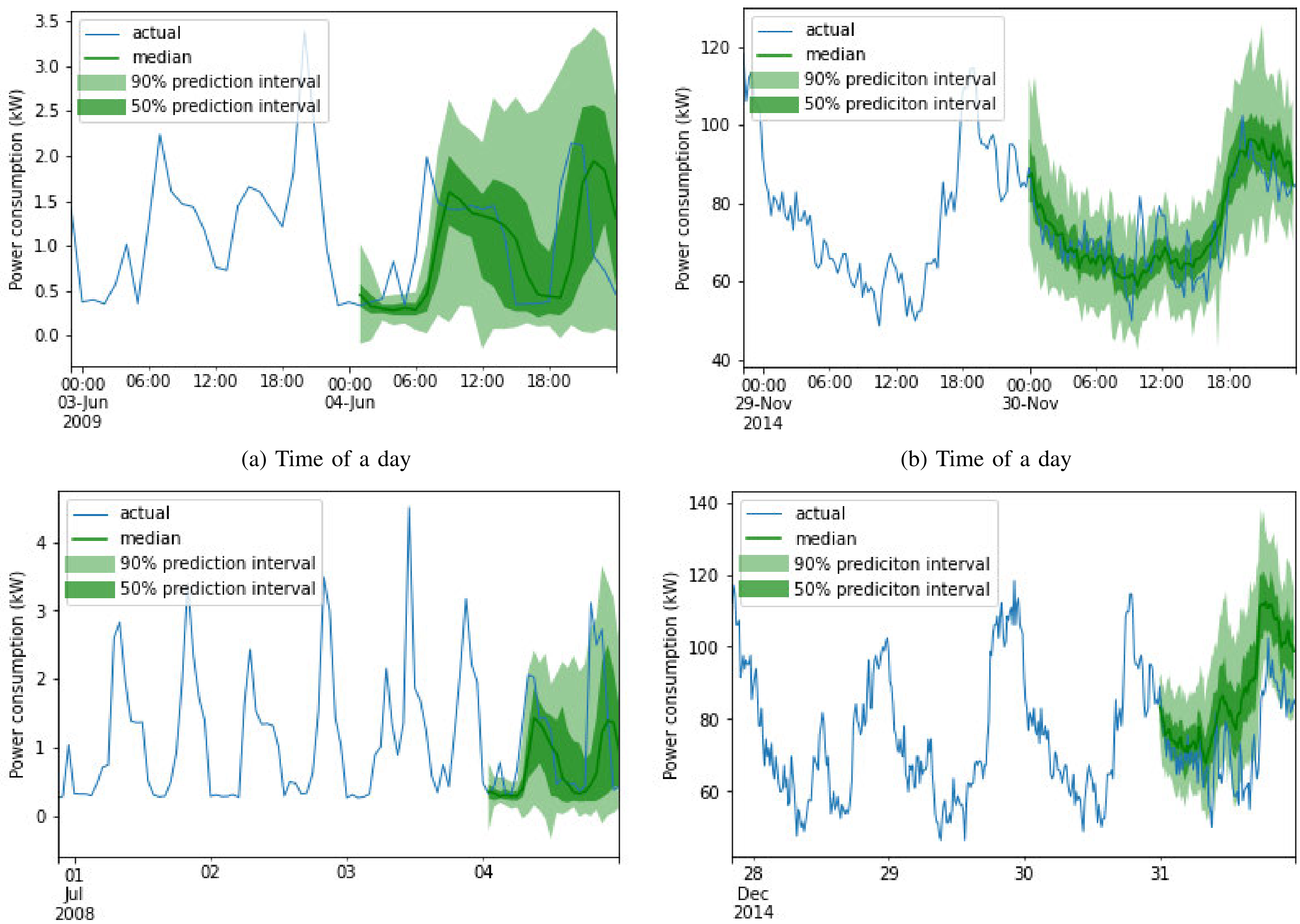
where μ_t has only two possible values with L_t and U_t as lower and upper bounds, respectively and can be written as

$$\mu_i = \begin{cases} 1, & \text{if } y_t \in [L_t, U_t], \\ 0, & \text{if } y_t \notin [L_t, U_t]. \end{cases} \tag{10}$$

Next, the PICP of the proposed method are compared in Table 4 and the probabilistic forecasting results are illustrated in Fig. 9. According to Table Table 4, the prediction of the proposed method gives trustworthy outcomes for the electricity forecast because the prediction interval range measured by the PICP is close to the nominal coverage probability. It shows the PICP performance comparison by probabilistic forecasting on cluster data with different test and training scales. The results were statistically significant compared to QRF, QRNN, QGBRT and Q-LSTM. Decision trees are used as the primary learners in the QRF ensemble with randomization, where the number of trees is set to 150. Next is the QGBRT, where 500 trees are used to make the QGBRT decisive enough because of the gradient boosting. As can be observed, all models perform better except the QRF and QGBRT, although efforts have been made to modify their hyperparameters. The PICP rate of Q-LSTM is interesting, which has a significant difference in the IHEPC case of 15-minute and 1-hour time horizons compared to QRF, QRNN and QGBRT. The results demonstrate that the proposed method better quantifies the uncertainties using a limited number of training samples.

In most cases, the proposed log-cosh quantile method outperforms Q-LSTM [29]. As shown in Table 4, PICP indicators of the proposed models outperform those of other models across the entire cluster set of randomly selected households from different customer and IHEPC datasets. For instance, the PICP values estimated from four comparative models are 86.17%, 86.88%, 87.35%, and 91.65%, whereas that from the proposed is 92.74%, which increased on the 1-hour horizon. Statistically, the proposed model performed the best, achieving average rates of 6.57%, 5.86%, 5.39% and 1.09% higher than the QRF, QRNN, QGBRT, and Q-LSTM, respectively. Furthermore, the results are further validated on a 15-minute horizon, which shows the increase is 5.88%, 4.61%, 3.39% and 2.36% higher than the QRF, QRNN, QGBRT, and Q-LSTM, respectively. The proposed model has a higher potential to provide a meritorious electricity probabilistic prediction and significantly reduce the uncertainty inherent in forecasting.

In the case of IHEPC, the proposed method validates using data from the 47 months of cluster results with overall



(a) Time of a day (b) Time of a day

(c) Day of a month (d) Day of a month

FIGURE 9. **An example of a probabilistic electric forecast shows four randomly selected scenarios for the household-associated probability distribution. (a) actual load and prediction interval of the IHEPC dataset at varying hours of the day (b) forecast of random selection from different customer datasets at varying hours of the day (c) forecast of IHEPC data for different days of a month (d) different customer data at different days of a month.**

PICP to test the prediction interval accuracy. The experiment considers time features, load characteristics, temperature, and the day's 15-minute and 1-hour time horizons. Fig. 9a and 9c shows the forecasting for the IHEPC electricity data with actual load and prediction intervals. The results shows the average power consumption value versus the probability density associated with the power consumption value. It shows the 50% and 90% quantile, predictive means, and actual load observations for three days from for the visualization. The findings demonstrate that the observations mostly be covered by 90%, with only a few abrupt changes being contradicted. Another example shows a dataset of the one-day forecast findings in Fig. 9c, where five days prediction interval from July 3 to July 5, 2008. In the case of different households, it shows the result from a random selection of two households, where the proposed method achieves the best performance on both 15-minute and 1-hour time.

Fig. 9b depicts a day-ahead probabilistic distribution prediction of a randomly selected household consumption from the different customer datasets between the days of December 2014. The actual load (blue line) of the randomly selected household data has been shown in Fig. 9b, together with quantile forecasts 50% and 90% prediction intervals based on the proposed method. Fig. 9d shows the day-ahead probabilistic load forecasting results. It shows the graph for the month of July 2008, where the blue line represents the actual loads, while the green shadowed area represents the load forecast prediction intervals. It is clear that most of the time, the prediction intervals generated by the proposed method cover the actual value with a prediction interval of 90%. Most of the mean probabilistic average load forecasts were relatively accurate in predicting the actual loads, which improved over IHEPC case outcomes but still not as well as the different household cases. Overall, our results demonstrate a strong effect of the proposed model to tackle the uncertainty of the STLF task based on the clustering-based probabilistic forecasting obtained by sampling the trained neural networks. Therefore, grouping customers into categories based on comparable patterns helps to reduce prediction uncertainty and boost forecast accuracy simultaneously.

VI. CONCLUSION

In conclusion, the findings of this study suggest that the utilization of the log-cosh loss function in conjunction with clustering-based probabilistic electric forecasting using quantile LSTM lead to improved accuracy and performance in energy load forecasting, highlighting the potential for further development and implementation in practical applications. The results of our investigation indicate that the proposed cluster-based probabilistic forecasting approach utilizing log-cosh quantile LSTM outperforms state-of-the-art tree-based methods including QRF, QRNN, QGBRT, and Q-LSTM across both 15-minute and 1-hour time horizons. The probability density distributions are obtained by developing a quantile LSTM model with a log-cosh loss function, effectively preventing overfitting and improving residential electric forecasting accuracy. Overall, the results show the effectiveness of the proposed clustering-based probabilistic forecasting model for reducing the uncertainty of short-term load forecasting. The results of this study suggest that incorporating probabilistic prediction techniques, such as the proposed cluster-based probabilistic forecasting with quantile LSTM, can lead to improved accuracy and better quantification of uncertainties in electricity forecasting.

REFERENCES

[1] S. S. Ali and B. J. Choi, "State-of-the-art artificial intelligence techniques for distributed smart grids: A review," *Electronics*, vol. 9, no. 6, p. 1030, Jun. 2020.

[2] S. Smyl and N. G. Hua, "Machine learning methods for GEFCom2017 probabilistic load forecasting," *Int. J. Forecasting*, vol. 35, no. 4, pp. 1424–1431, Oct. 2019.

[3] Z. Cao, C. Wan, Z. Zhang, F. Li, and Y. Song, "Hybrid ensemble deep learning for deterministic and probabilistic low-voltage load forecasting," *IEEE Trans. Power Syst.*, vol. 35, no. 3, pp. 1881–1897, May 2020.

[4] H. Wang, N. Zhang, E. Du, J. Yan, S. Han, and Y. Liu, "A comprehensive review for wind, solar, and electrical load forecasting methods," *Global Energy Interconnection*, vol. 5, no. 1, pp. 9–30, Feb. 2022.

[5] Y. Wang, N. Zhang, Y. Tan, T. Hong, D. S. Kirschen, and C. Kang, "Combining probabilistic load forecasts," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3664–3674, Jul. 2019.

[6] T. Hong and S. Fan, "Probabilistic electric load forecasting: A tutorial review," *Int. J. Forecasting*, vol. 32, no. 3, pp. 914–938, Jul. 2016.

[7] S. Sepasi, E. Reihani, A. M. Howlader, L. R. Roose, and M. M. Matsuura, "Very short term load forecasting of a distribution system with high PV penetration," *Renew. Energy*, vol. 106, pp. 142–148, Jun. 2017.

[8] Y. He, Y. Qin, X. Lei, and N. Feng, "A study on short-term power load probability density forecasting considering wind power effects," *Int. J. Electr. Power Energy Syst.*, vol. 113, pp. 502–514, Dec. 2019.

[9] U. Zafar, S. Bayhan, and A. Sanfilippo, "Home energy management system concepts, configurations, and technologies for the smart grid," *IEEE Access*, vol. 8, pp. 119271–119286, 2020.

[10] D. Mariano-Hernández, L. Hernández-Callejo, A. Zorita-Lamadrid, O. Duque-Pérez, and F. Santos García, "A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis," *J. Building Eng.*, vol. 33, Jan. 2021, Art. no. 101692.

[11] G. Van Houdt, C. Mosquera, and G. Nápoles, "A review on the long short-term memory model," *Artif. Intell. Rev.*, vol. 53, no. 8, pp. 5929–5955, 2020.

[12] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. Int. Conf. Eng. Technol. (ICET)*, Aug. 2017, pp. 1–6.

[13] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2017, pp. 1597–1600.

[14] N. Somu, G. Raman M R, and K. Ramamritham, "A deep learning framework for building energy consumption forecast," *Renew. Sustain. Energy Rev.*, vol. 137, Mar. 2021, Art. no. 110591.

[15] Z. Karevan and J. A. K. Suykens, "Transductive LSTM for time-series prediction: An application to weather forecasting," *Neural Netw.*, vol. 125, pp. 1–9, May 2020.

[16] Y. Hua, Z. Zhao, R. Li, X. Chen, Z. Liu, and H. Zhang, "Deep learning with long short-term memory for time series prediction," *IEEE Commun. Mag.*, vol. 57, no. 6, pp. 114–119, Jun. 2019.

[17] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—A novel pooling deep RNN," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Sep. 2018.

[18] M. Rafiei, T. Niknam, J. Aghaei, M. Shafie-Khah, and J. P. S. Catalão, "Probabilistic load forecasting using an improved wavelet neural network trained by generalized extreme learning machine," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6961–6971, Nov. 2018.

[19] S. Singh, S. Hussain, and M. A. Bazaz, "Short term load forecasting using artificial neural network," in *Proc. 4th Int. Conf. Image Inf. Process. (ICIIP)*, Dec. 2017, pp. 1–5.

[20] A. Baliyan, K. Gaurav, and S. K. Mishra, "A review of short term load forecasting using artificial neural network models," *Proc. Comput. Sci.*, vol. 48, pp. 121–125, Jan. 2015.

[21] Y. Wang, Q. Chen, N. Zhang, and Y. Wang, "Conditional residual modeling for probabilistic load forecasting," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 7327–7330, Nov. 2018.

[22] J.-F. Toubeau, J. Bottieau, F. Vallée, and Z. De Grève, "Deep learning-based multivariate probabilistic forecasting for short-term scheduling in power markets," *IEEE Trans. Power Syst.*, vol. 34, no. 2, pp. 1203–1215, Mar. 2019.

[23] B. Liu, J. Nowotarski, T. Hong, and R. Weron, "Probabilistic load forecasting via quantile regression averaging on sister forecasts," *IEEE Trans. Smart Grid*, vol. 8, no. 2, pp. 730–737, Mar. 2017.

[24] K. Aurangzeb, M. Alhussein, K. Javaid, and S. I. Haider, "A pyramid-CNN based deep learning model for power load forecasting of similar-profile energy customers based on clustering," *IEEE Access*, vol. 9, pp. 14992–15003, 2021.

[25] K. Aurangzeb, "Anomalies and major cluster-based grouping of electricity users for improving the forecasting performance of deep learning models," *Frontiers Energy Res.*, vol. 11, Nov. 2023, Art. no. 1284076.

[26] D. Gan, Y. Wang, S. Yang, and C. Kang, "Embedding based quantile regression neural network for probabilistic load forecasting," *J. Mod. Power Syst. Clean Energy*, vol. 6, no. 2, pp. 244–254, Mar. 2018.

[27] M. Lopez-Martin, A. Sanchez-Esguevillas, L. Hernandez-Callejo, J. I. Arribas, and B. Carro, "Additive ensemble neural network with constrained weighted quantile loss for probabilistic electric-load forecasting," *Sensors*, vol. 21, no. 9, p. 2979, Apr. 2021.

[28] Y. He, Y. Qin, S. Wang, X. Wang, and C. Wang, "Electricity consumption probability density forecasting method based on LASSO-quantile regression neural network," *Appl. Energy*, vols. 233–234, pp. 565–575, Jan. 2019.

[29] D. Gan, Y. Wang, N. Zhang, and W. Zhu, "Enhancing short-term probabilistic residential load forecasting with quantile long–short-term memory," *J. Eng.*, vol. 2017, no. 14, pp. 2622–2627, Jan. 2017.

[30] W. Zhang, H. Quan, and D. Srinivasan, "An improved quantile regression neural network for probabilistic load forecasting," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 4425–4434, Jul. 2019.

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

[32] Y. Wang, Q. Chen, and C. Kang, "Probabilistic residential load forecasting," in *Smart Meter Data Analytics*. Cham, Switzerland: Springer, 2020, pp. 247–269.

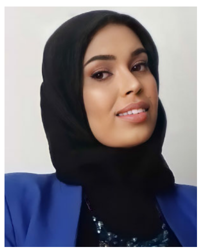
[33] W. Zhang, H. Quan, and D. Srinivasan, "Parallel and reliable probabilistic load forecasting via quantile regression forest and quantile determination," *Energy*, vol. 160, pp. 810–819, Oct. 2018.

[34] S. Ben Taieb, R. Huser, R. J. Hyndman, and M. G. Genton, "Forecasting uncertainty in electricity smart meter data by boosting additive quantile regression," *IEEE Trans. Smart Grid*, vol. 7, no. 5, pp. 2448–2455, Sep. 2016.

[35] O. F. Eikeland, F. D. Hovem, T. E. Olsen, M. Chiesa, and F. M. Bianchi, "Probabilistic forecasts of wind power generation in regions with complex topography using deep learning methods: An Arctic case," *Energy Convers. Manag., X*, vol. 15, Aug. 2022, Art. no. 100239.

[36] D. Dheeru and T. E. Karra, ''Individual household electric power consumption data set,'' *UCI Mach. Learn. Repository*, 2018.
[37] D. Dheeru and T. E. Karra, ''Electricity load diagrams 2011–2014 data set,'' *UCI Mach. Learn. Repository*, 2011.
[38] T. Räsänen and M. Kolehmainen, ''Feature-based clustering for electricity use time series data,'' in *Proc. Int. Conf. Adapt. Natural Comput. Algorithm*. Springer, 2009, pp. 401–412.
[39] J. Bedi and D. Toshniwal, ''Deep learning framework to forecast electricity demand,'' *Appl. Energy*, vol. 238, pp. 1312–1326, Mar. 2019.
[40] E. Barbour and M. González, ''Enhancing household-level load forecasts using daily load profile clustering,'' in *Proc. 5th Conf. Syst. Built Environ.*, Nov. 2018, pp. 107–115.
[41] S. Du, T. Li, Y. Yang, and S.-J. Horng, ''Multivariate time series forecasting via attention-based encoder–decoder framework,'' *Neurocomputing*, vol. 388, pp. 269–279, May 2020.
[42] Z. Masood, R. Gantassi, and Y. Choi, ''A multi-step time-series clustering-based Seq2Seq LSTM learning for a single household electricity load forecasting,'' *Energies*, vol. 15, no. 7, p. 2623, Apr. 2022.
[43] X. Huang, L. Shi, and J. A. K. Suykens, ''Support vector machine classifier with pinball loss,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 984–997, May 2014.
[44] P. Chen, G. Chen, and S. Zhang, ''Log hyperbolic cosine loss improves variational auto-encoder,'' Tech. Rep., 2018.
[45] G. P. Meyer, ''An alternative probabilistic interpretation of the Huber loss,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5257–5265.
[46] O. Karal, ''Maximum likelihood optimal and robust support vector regression with lncosh loss function,'' *Neural Netw.*, vol. 94, pp. 1–12, Oct. 2017.
[47] SCarnot, ''Extraction and analysis of historical French weather data,'' Tech. Rep., 2019.
[48] Météo-France. *Paris-Montsouris Weather Data*. Accessed: Jul. 15, 2022. [Online]. Available: https://www.infoclimat.fr/climatologie-mensuelle/07156/fevrier/1905/paris-montsouris.html
[49] D. Dheeru and T. E. Karra, ''Uci machine learning repository,'' *UCI Datasets*, 2018.
[50] N. Khan, I. U. Haq, F. U. M. Ullah, S. U. Khan, and M. Y. Lee, ''CL-Net: ConvLSTM-based hybrid architecture for batteries' state of health and power consumption forecasting,'' *Mathematics*, vol. 9, no. 24, p. 3326, Dec. 2021.
[51] M. A. Alghamdi, A. S. A.-M. Al-Ghamdi, and M. Ragab, ''Predicting energy consumption using stacked LSTM snapshot ensemble,'' *Big Data Mining Anal.*, vol. 7, no. 2, pp. 247–270, Jun. 2024.
[52] M. Abdel-Basset, H. Hawash, K. Sallam, S. S. Askar, and M. Abouhawwash, ''STLF-Net: Two-stream deep network for short-term load forecasting in residential buildings,'' *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 7, pp. 4296–4311, Jul. 2022.
[53] V. Dehalwar, A. Kalam, M. L. Kolhe, and A. Zayegh, ''Electricity load forecasting for urban area using weather forecast information,'' in *Proc. IEEE Int. Conf. Power Renew. Energy (ICPRE)*, Oct. 2016, pp. 355–359.
[54] B. Nepal, M. Yamaha, A. Yokoe, and T. Yamaji, ''Electricity load forecasting using clustering and ARIMA model for energy management in buildings,'' *Jpn. Architectural Rev.*, vol. 3, no. 1, pp. 62–76, Jan. 2020.
[55] A. Likas, N. Vlassis, and J. J. Verbeek, ''The global k-means clustering algorithm,'' *Pattern Recognit.*, vol. 36, no. 2, pp. 451–461, Feb. 2003.
[56] B. Ben Gouissem, R. Gantassi, and S. Hasnaoui, ''Energy efficient grid based k-means clustering algorithm for large scale wireless sensor networks,'' *Int. J. Commun. Syst.*, vol. 35, no. 14, p. e5255, Sep. 2022.
[57] S. Yilmaz, J. Chambers, and M. K. Patel, ''Comparison of clustering approaches for domestic electricity load profile characterisation–Implications for demand side management,'' *Energy*, vol. 180, pp. 665–677, Aug. 2019.
[58] A. De Myttenaere, B. Golden, B. Le Grand, and F. Rossi, ''Using the mean absolute percentage error for regression models,'' in *Proceedings*, 2015, p. 113.
[59] T. Chai and R. R. Draxler, ''Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature,'' *Geosci. Model Develop.*, vol. 7, no. 3, pp. 1247–1250, Jun. 2014.
[60] M. Abadi et al., ''TensorFlow: A system for large-scale machine learning,'' in *Proc. 12th USENIX Symp. Oper. Syst. Design Implement.*, 2016, pp. 265–283.
[61] A. Alexandrov, K. Benidis, M. Bohlke-Schneider, V. Flunkert, J. Gasthaus, T. Januschowski, D. C. Maddix, S. Rangapuram, D. Salinas, J. Schulz, L. Stella, A. Caner Türkmen, and Y. Wang, ''GluonTS: Probabilistic time series models in Python,'' 2019, *arXiv:1906.05264*.
[62] A. Gulli and S. Pal, *Deep Learning With Keras*. Packt Publishing Ltd, 2017.
[63] G. K. Rajbahadur, S. Wang, G. A. Oliva, Y. Kamei, and A. E. Hassan, ''The impact of feature importance methods on the interpretation of defect classifiers,'' *IEEE Trans. Softw. Eng.*, vol. 48, no. 7, pp. 2245–2261, Jul. 2022.

ZAKI MASOOD (Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Bahria University, Pakistan, in 2010 and 2015, respectively, and the Ph.D. degree in electrical engineering from Chonnam National University (CNU), South Korea, in 2021. He is currently a Postdoctoral Researcher with the Energy ICT Laboratory, CNU. He was a Senior Lecturer with the Department of Electrical Engineering, Bahria University. His research interests include wireless energy harvesting, the Internet of Things (IoT) technology, and artificial intelligence (AI) energy for smart grids.



RAHMA GANTASSI (Member, IEEE) received the Ph.D. degree in computer science from the University of Tunis El Manar (UTM), Tunisia, in 2021. She is currently a Postdoctoral Researcher Fellow with the Energy ICT Laboratory, Chonnam National University (CNU), South Korea. She has published several academic articles in international refereed journals and conferences. Her current research interests include the Internet of Things (IoT), optimization of wireless sensor networks (WSN), and optimizing quality of service (QoS) in WSN. She has also served as a local organizing and program committee member for various international conferences and as a reviewer for several refereed journals.

YONGHOON CHOI (Senior Member, IEEE) received the B.S. degree in electronics engineering from Sungkyunkwan University, Seoul, South Korea, in 1999, and the M.S. and Ph.D. degrees in information and communications engineering from Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2003 and 2010, respectively. He was a Postdoctoral Visiting Scholar with the Department of Electrical Engineering, Stanford University, Stanford, CA, USA, from 2010 to 2013. Since 2014, he has been with the Department of Electrical Engineering, Chonnam National University, South Korea, where he is currently an Associate Professor. His research interests include the design, analysis, optimization of wired/wireless communication systems, smart grid communications and networking, energy trading, network economics, and IT convergence networks.