Probability density function forecasting of residential electric vehicles charging profile

Jamali Jahromi Ali, Mohammadi Mohammad, Afrasiabi Shahabodin, Afrasiabi Mousa, Aghaei Jamshid

DOI: 10.1016/j.apenergy.2022.119616

Probability Density Function Forecasting of Residential Electric Vehicles Charging Profile

Ali Jamali Jahromi[1], Mohammad Mohammadi[1], Shahabodin Afrasiabi[2], Mousa Afrasiabi[3], Jamshid Aghaei[3]

[1]*Department of Power and Control Engineering, School of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran*

[2]*Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, Canada*

[3]*School of Energy Systems Lappeenranta-Lahti University of Technology (LUT), Lappeenranta, Finland*

E-mail Address: jamalijahromiali@gmail.com, m.mohammadi@shirazu.ac.ir, sh.afrasiabi@usask.ca, musa.afrasiabi@lut.fi

Abstract— **This paper presents the main principle of the probability density function forecasting approach in residential electric vehicle (REV) charging profile. To this end, an end-to-end deep learning structure is designed and integrated with kernel density estimation (KDE). The designed deep network consists of a convolutional neural network (CNN), a deep attention mechanism, gated recurrent unit (GRU), and autoregressive (AR) blocks to directly learn spatial-temporal features as well as long-term behavior. Then, the outputs are fed into a KDE to extract full-statistical information in terms of the probability density function (PDF).**

***Index Terms*—Deep learning, residential electric vehicle, probabilistic forecasting, kernel density estimator**

NOMENCLATURE

Parameters	
Z_n	Output vector of convolutional layer
u_t	Update gate in gated recurrent unit block
W, K	dense layer kernel weights
b	corresponding biases
r_t	Reset gate in gated recurrent unit block
Z_t	Hidden states in gated recurrent unit block
$\hat{Z}_t$	Candidate hidden states in gated recurrent unit block
c_i	Context vector in attention mechanism
α_{ik}	Weights in attention mechanism (computed by the SoftMax function)
e_{ik}	Output of score function in attention mechanism
W_a, U_a, V_a	Trainable attention weights
C^{AR}, d^{AR}	Coefficients of the autoregressive model
Z_t^{AR}	Output of autoregressive component
Z_t^F	Final prediction by the nonlinear structure of deep learning
$\hat{Z}_t^{final}$	Represents the final prediction of the model at time stamp t
h	Bandlimit value of kernel density estimator
δ	Hyperparameter of Huber loss function
Functions and Variables	
$\odot$	Hadamard product
$\otimes$	Convolution operation
$g^{ReLU}(\bullet)$	Rectified linear unit activation function
L_δ	Huber loss function
$\hat{f}(y)$	Kernel density estimator
$K(\chi)$	Gaussian kernel function
H	Historical data
n_i	Number of historical data
F	Observation data
n_o	Number of observation data
n^{th}	Number of convolutional layer filters sweeps through the input dataset
Symbols and Acronyms	
min	Minute
G2V, V2G	Grid to vehicle and Vehicle to grid
Y_a, Y_p	Actual and predicted values
$\mathrm{CDF}(Y_p), \mathrm{CDF}(Y_a)$	Predictive and actual cumulative distribution function

I. INTRODUCTION

A. *Motivation and Outlook*

Nowadays, the energy crisis and greenhouse gas emissions have become one of the most important issues for human society [1]. Because of this, the international community is looking for solutions to replace renewable energy with fossil fuels [2]. Transportation industry uses fossil fuels significantly, where the use of electric vehicles (EVs) in transportation fleet helps to reduce fossil fuel consumption and greenhouse gas emissions.

Electric vehicles play a prominent role in the transportation and electric grids due to increasing the diversity of fuel choice for the owners, environmentally-friendly impacts, and capability to perform as mobile energy storage devices [3]. Hence, in recent years, the establishment of EVs usage has been rapidly increased from 2010 to 2020 from 11 thousand to 30 million [4]. Furthermore, predict that the number of EVs will surpass 125 million vehicles worldwide [5].

Nowadays, EV owners prefer to supply their required electrical energy by residential electrical sources instead of charging stations and parking lots [6, 7]. However, uncontrolled and unprincipled residential electrical vehicle (REV) charging may violate the operating limitations of power distribution systems and cause problems such as increased load during peak hours, voltage fluctuations, and increased mains losses [8, 9]. Hence, an optimal decision-making process is crucial for the short/long term scheduling for REVs. Recently, the aggregator agents have been formed to provide the optimal setpoint in the short/long term of REVs and perform as the intermediate entities to benefit from the regulation of the REVs charging behaviour. REV aggregators are also beneficial for the owners and upstream grids through minimization of the individual EV cost and supporting the network by vehicle-to-grid (V2G) capability [10].

In this context, the nonstationary and volatility of the REVs is a huge challenge. This extremely challenging task motivates us to design an accurate forecasting engine to provide necessary information of the charging consumption in the look-ahead times considering spatial-temporal-long patterns. In addition to deterministic forecasting, a probabilistic forecast in forms of probabilistic density function (PDF) can also assist REVs aggregator in planning to charge/discharge vehicles. Thus, probabilistic REVs charging profile forecasting is important to offer complete facts approximately future charging profile intake to lessen operation costs and enhance the reliability of the power grid [11]. The advantage of PDF is in providing comprehensive statistical information in future time periods. In this paper, we consider an indirect approach to PDF extraction. In the first step, a prediction of the charge of home electric vehicles is made by a deep neural network designed. In the second step, using the results obtained from the first step, we predict the PDF in each time period.

B. Brief Survey on Previous Works

Generally, residential electric vehicles charge prediction will be categorized into deterministic and probabilistic forecasting. The majority of courses have targeted on deterministic forecasting [12]. Probabilistic short-term forecasting method can generally divide into three major groups: i) Prediction intervals (PIs) strategies try to replace a group of values with a fixed of PIs for destiny cognizance of a random variable. In PIs extraction methods, a set of probabilistic values are extracted based on an error base function to form PIs by a predictive model [13]. In general, traditional PIs extraction methods, including delta, Bayesian and mean variance, suffer from high computational costs [14]. In these techniques, a self-belief degree is predefined, and a way to pick out the proper confidence level is missing. ii) Another type of probabilistic prediction is the quintile prediction. Quantile prediction is the same as PI prediction except those quantiles are estimated instead of PIs. For instance, in [15] the charge of electric vehicles is predicted in forms of quantile using convolutional neural networks. iii) Another type of probabilistic forecasting is a probabilistic density function (PDF) that includes all comprehensive statistical information. PDF prediction generates a ways greater record in assessment with distinctive representations including PIs and quantiles, which can be derived from the PDFs or CDFs. To the best of our knowledge, unlike the primary classes, the charge prediction of REVs in PDF format has been less studied. However, similar time series predictions, such as the residential load, have been predicted in terms of kernel density estimator (KDE) [16].

On the other hand, time series predictions are divided into four categories: i) persistence, ii) physical, iii) statistical, and iv) data-driven methods.

The physical-based prediction models attempt to define a mathematical relationship between the prediction values and the observation values. The physical models are usually used in weather prediction such as numerical weather prediction (NWP) [17]. However, physical models are time-consuming due to the high computational rate, and are not a practical solution for short-term forecasting [18], in particular dynamic time series such as EV charging profile.

The statistical models have described the outputs based on the probabilistic variables, descriptive variables, online measurement, and historical data. The majority of the statistical models are from the autoregressive (AR) family of models. AR-based models are auto-corrective and linear models, which can model long and regular patterns including daily/weekly/yearly periodic patterns. In general, statistical models use a linear approximation for prediction. The autoregressive moving average (ARMA) has been presented in [19] for the large-scale EV fleet charging profile prediction. In [19], instead of actual historical data, predefined probability density functions (PDFs) have been used to generate historical data. In the highly nonlinear and non-stationary time such as actual EV charging profile, statistical models as the main model, are suffered from disability to capture the nonlinear behaviour of EV charging profiles. Moreover, in the small-scale EV charging profile, statistical models are only able to describe the regular pattern and cannot directly model a time series with a large share of uncertainty [16], such as the REV charging profile.

Data-driven-based short-term forecasting methods have been spotlighted in recent years due to their capability of handling nonlinear and complex models. Data-driven models are only dependent on historical data and can be categorized into shallow and deep-based structures. For instance, support vector machine (SVM) [20] and k-nearest neighbor (kNN) [21] have presented to predict the EVs charging profile. However, the shallow-based method cannot properly characterize the features of a non-linear behavior of REVs charging profile. Therefore, these techniques can be combined with supplementary extraction/selection blocks to enhance the performance. For instance, a combination of empirical mode decomposition and SVM has been developed in [22] for large-scale demand forecasting. The wavelet transform+random forest in [23] and principal component analysis+Fourier transform+neural network in [24] have been presented for the renewable power generation and solar irradiation forecasting, respectively. However, there are several disadvantages: i) additional blocks increase the computational cost, ii) combination with feature selection/extraction methods are not general solutions [25] for the short-term forecasting, and iii) extract the regular pattern.

To the best of our knowledge, the nonlinear properties and uncertainties of the REVs charging profile cannot be realized based on the physical models or capture by the linear/statistical models and shallow-based models. Moreover, the feature selection/extraction techniques extract the regular pattern and neglect the uncertain behaviour. Therefore, REVs charging profile forecasting based on the combination methods cannot be considered a suitable option for the small-scale REVs demand prediction.

This paper aims to explore the possibility of using advanced deep learning for direct learning uncertainty. Deep learning is a branch of machine learning methods based on deep architectures that combine several layers of processing in a neural network and the possibility of learning relationships.

The drawbacks and limitations of the previous prediction models would be tackled by using a data-driven deep learning approach, which is usually implemented through the NN architecture. The main advantage of deep models is their capability to learn features from given inputs automatically (i.e., no additional feature extraction is needed). Due to the capability in reproducing complicated and nonlinear features from the raw time series data, deep learning is able to increase learning capability and making the precision and reliability in the time series prediction problems such as REVs charging profile forecasting. Deep neural networks (DNN) would be branched into varies principal categories, including auto-encoder (AE), deep Boltzmann machine (DBM), convolutional neural network (CNN), and recurrent neural network (RNN) [26].

AE and DBM are presented for estimation state of charge (SOC) in EVs battery [27] and forecasting load level of the power system in [28]. Although AE and DBM make the learning ability from raw data throughout a dimensional decrease procedure, the principal defects of AR and DBM are the inability to understand long sequences in both spatial-temporal features. Recurrent neural network (RNN) models [29] have gained popularity in a wide range of data related to the past. In particular, the two types of RNNs, long-short-term memory (LSTM) [30] and gated recurrent unit (GRU) [31], have enhanced performance in data with learning time-variant behaviour. The main problem with RNNs is their poor performance in extracting the spatial features. To extract fully spatial features, CNN has been considered as a potential candidate [32], while CNN is not able to fully understand the temporal features of the volatile time series such as small-scale electric vehicle charging profiles. The deep neural network-based EVs charging profile forecasting has also been studied in the previous literature. For instance, in [33], the LSTM model is used to predict the charge demand of EVs based on raw data. Furthermore, in [34], the CNN model is used to predict the charge required by EVs charging stations.

TABLE I

SUMMARIZED DESCRIPTION, PROS, AND CONS OF PREVIOUS SHORT-TERM FORECASTING MODELS ON REV CHARGING PROFILE FORECASTING

	Persistence	Physical	Statistical	Data-driven
Description	Charging of electric vehicles in the future is predicted as in previous time intervals.	The charging time series of electric vehicles is modelled based on mathematical equations.	Linearization based on the distance between the actual amount of data in the present and the past.	A Model is constructed based on the historical data to characterize the time series information in look-ahead times.
Pros	Good performance in very short time series forecasts.	Good performance in long time series.	Easy modelling based on electric vehicle charging time series pattern	Ability to receive long-term nonlinear time series.
Cons	Inability to detect and predict short-term and long-term time series.	Improper performance due to the very high volume of calculations and the lack of detection of uncertainty in the series of charging of EVs.	Linearization of the problem and ignoring the uncertainties of the problem.	Shallow-based: cannot be applicable as an accurate forecasting engine in the REVs charging profile due to the disability in the handling the time series with a large scale of the uncertainty. Existing deep-based: inability to fully characterize the spatial-temporal features and uncertain behaviour of REVs

Thus, this paper aims to develop a deep structure to understand the spatial-temporal features as well as model the long patterns only based on raw data (without additional feature engineering block).

C. Contribution

This paper attempts to develop a DNN structure basis on the historical data (actual data of REVs) to accurately forecast the REVs connected to the EV aggregators. Then, according to the results obtained from the deterministic forecast, we will extract the PDF for each time intervals. For this purpose, a deep structure with the ability to extract different types of time series features has been designed in this paper. The proposed method can directly realize the regular pattern as well as uncertain pattern (both has an equal share in the charging profile of Revs) of the REVs charging profile behaviour. In the proposed approach, the designed long-short-term learning network consists of CNN, GRU, AR, attention mechanism, and fully-connected neural (FCN) layers. CNN learns the spatial features (REVs at different locations with small-scale charging profiles) in the learning process and GRU improves the capability of understanding the temporal features. The attention mechanism enhances the ability of the designed network by assigning higher learning weights to the features with higher importance. To extract the long-term dependencies including REVs runtime behaviour, AR can extract the limited range of the charging profile of REVs. Finally, KDE is used to extract PDF at any time interval. Hence, the novelties can be summarized as:

- A deep-based structure is combined with AR to fully understand the residential load charging profile features i.e., temporal, spatial, and daily/weekly/yearly behavior of REVs, directly from the raw data
- To the best of our knowledge, it is the first time that probabilistic REVs charging profile prediction is analyzed based on actual data.
- A Combined model is designed to realize the uncertain behavior of the REVs charging profile without any cancelation of uncertainty by clustering/classification and pre-process the EVs profile or separating the regular pattern based on raw data.

D. Organization

The rest of the paper is organized as follows: The background of the proposed probabilistic deep structure is introduced in Section I. Section II describes the problem statement. Section III describes the proposed deep probabilistic forecasting structure. Section IV describes the numerical results and discussion of the proposed probabilistic end-to-end long-short-term learning network before concluding in Section V.

II. PROBLEM STATEMENT

The charging profile of REVs has fully observed time series signals. Let $(H, F) = (h_1, f_1), \ldots, (h_n, f_n)$ be data, where $H \in \mathbb{R}^I\ \forall\, i = 1, \ldots, n_i$, is n_i historical data, and $F \in \mathbb{R}^O\ \forall\, o = 1, \ldots, n_o$ is the n_o observation data. The important goal inside the data-driven short-time charging profile forecasting is the construction of the community to venture a hard and fast of $\acute{F}$ with the minimal difference with F. In this paper, a complicated deep structure, specifically the long-spatial-temporal learning network (LSTLNet) is proposed to expect the charging profile throughout a short-time period length. Next, the output of the LSTLNet approach is given to a PDF estimator to provide a probabilistic forecast in PDF for each time interval. Real-world applications often contain a mixture of short-term and long-term new release patterns. The daily charging profile of a set of REVs is illustrated in Fig 2. Accordingly, there are two essential patterns within each day charging profile, which are periodically repeated all through a day/week/month. The everyday patterns show the charging conduct of the exceptional EVs in a day. For instance, based totally on Fig 1, some REVs charging is begun within the early hours of a day. Then, the charging has increased in middle hours, and eventually reduces in past due hours. The 2nd form of pattern inside the charging behavior of REVs involves long-term patterns inclusive of workday/weekends, seasonal outcomes, and so forth. The forecasting method for the charging profile of REVs should recognize each type of charging behaviors. On the other hand, the PDF estimator must also be able to extract PDFs with high reliability and accuracy. To this end, the kernel density estimator (KDE) is used to estimate PDFs for each interval in this paper.

The REVs behavior analysis shows that the share of the uncertainty pattern in charging REVs is nearly the same as the ordinary sample, therefore, classical forecasting methods aren't realistic tools on this regard [30, 31].

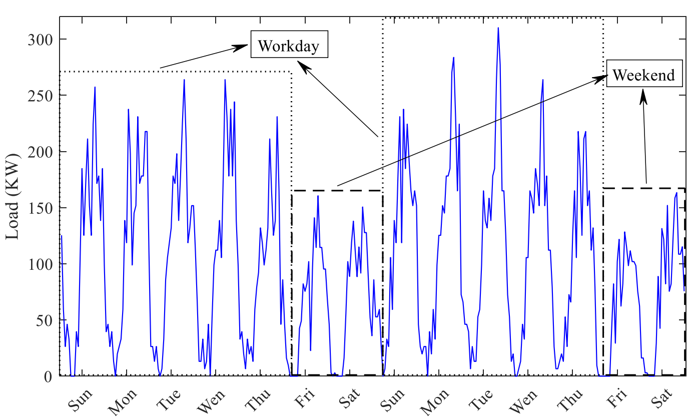

Fig. 1. The hourly charging required for 2 weeks

Fig 2 compares the average charging profile and realistic charging profile of 348 REVs. As may be visible from this Fig, the differences between the 2 plots show a sizable percent of the uncertainty sample in the REVs profile, which complicates the charging profile forecasting problem.

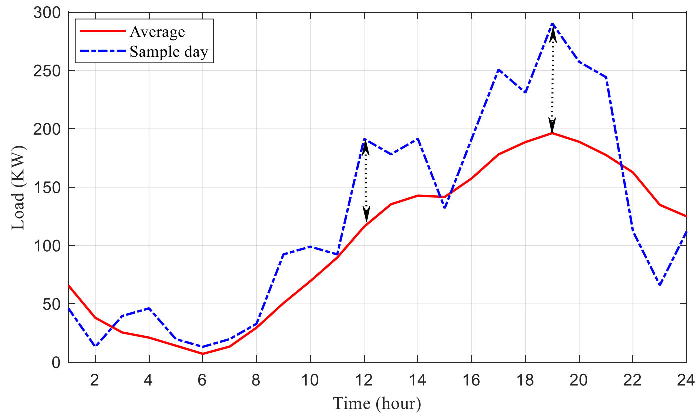

Fig. 2. Average year and sample day REVs load

III. PROPOSED DEEP PROBABILISTIC FORECASTING STRUCTURE

The proposed probabilistic forecasting approach includes four main blocks, i.e., convolutional neural network (CNN), attention mechanism, gated recurrent unit (GRU), autoregressive (AR), and KDE. The input data is fed into the CNN block to attention mechanism block to GRU block and summation with the AR block and at the end, the result of the proposed approach is fed into PDF Estimator block and give a PDF for each interval in the next 24-hours.

E. Convolutional Layer

The first block of the network is a convolutional layer. The convolutional layer can extract fully spatial capabilities, and learn short-time period styles as well as local dependencies [32]. The convolutional layers consist of numerous filters, that are acknowledged with the aid of width w and height h, where h is the same as the input dataset. The output of the n^{th} filter sweeps through the input dataset, and the outputs are:

$$Z_n = g^{ReLU}(W_n \otimes H + b_n) \qquad (1)$$

where $\otimes$ shows the convolution operation, while the output vector is shown by Z_n. Furthermore, $g^{ReLU}(\bullet)$ shows rectified linear unit (ReLU) activation function. The ReLU activation function is used to take away the negative values, and consequently, reduce the ability of the version to suit or train from the information nicely and described as:

$$g^{ReLU}(x) = \begin{cases} 0\,, & if\ x \le 0 \\ x\,, & otherwise \end{cases} \qquad (2)$$

F. GRU Block

The convolution block output is given to the recurrent units. As noted before, gated recurrent units (GRUs) are incorporated into the designed network to enhance getting to know capacity via shooting temporal capabilities. GRU is a time-efficient version of LSTM, which performs extra correctly than LSTM in 1D-time series forecasting [33]. GRU includes important gates such as the update and the reset gates. The update gate plays primarily based on:

$$u_t = g^{ReLU}(W_u h_t + K_u\, Z_{t-1} + b_u) \qquad (3)$$

Where W, K and b are parameter matrices and vector. The reset gate resets the redundant information with:

$$r_t = g^{ReLU}(W_r h_t + K_r Z_{t-1} + b_r) \quad (4)$$

Furthermore, the hidden states and candidate hidden states of recurrent units at time t is calculated as follows:

$$Z_t = (1 - u_t) \odot Z_{t-1} + u_t \odot \hat{Z}_t \quad (5)$$

$$\hat{Z}_t = g(W_z h_t + K_z(r_t \odot Z_{t-1} + b_z)) \quad (6)$$

in which $\odot$ is the Hadamard product and h_t is the input of this deposit at time t.

G. Attention Mechanism

One of the fundamental typically omitted issues within the time series forecasting is the interference of the alternative features and emphasis on the capabilities with higher significance weights. To this end, an attention mechanism is a device that may be incorporated into the deep models. The attention mechanism presents the potential to adaptively regulate the learning weights and focus at the function maps with higher significance weight. The word "attention" refers to the allocate extra attention to the features with higher significance weight throughout the training method. Furthermore, the attention mechanism mitigates the affect of the capabilities with decrease significance weights and, even may cast off the point characteristic maps. Thus, attention mechanisms can enhance the training process by way of allocating more attention to the capabilities with higher importance weight and mitigate the facts redundancy and overfitting with the aid of eliminating and allocate less attention to the features with lower importance capabilities. The attention mechanism is used to assigns unique significance to the one of a kind parts of the input collection at each window position of the input variables.

The context vector c_i for the output O_t is generated using the attention weights:

$$c_i = \sum_{k=1}^{n} \alpha_{ik} h_k \quad (6)$$

The weights α_{ik} are computed by the SoftMax function given as follows:

$$\alpha_{ik} = \frac{exp\ (e_{ik})}{\sum_{k=1}^{n} exp\ (e_{ik})} \quad (7)$$

$$e_{ik} = V_a tanh(U_a S(t-1) + W_a h_k) \quad (8)$$

where e_{ik} shows the output of the score function. W_a, U_a, and V_a are trainable weights, that are called *attention weights.*

H. Autoregressive Component

The scale of the outputs is related to the dimensions of the inputs. The neural networks, irrespective of whether of being deep networks or now not, have not considered this problem. In the time series like REVs charging profile, the dimensions of the inputs (quantity of REVs, kind of REV, Type of REVs battery, and charging demand over a selected duration) is continuously changing non-periodically. Therefore, ignoring the scale of the inputs can cause insufficiency of the short-time period forecasting method. To address this difficulty, in the proposed LSTLNet, an AR block has been added to the designed community. AR block improves the accuracy of the proposed network through capturing linear components and thinking about the size of the input. Denoting the forecast result of the AR component as Z_t^{AR}, and the coefficients of the AR model as C^{AR} and d^{AR}, where the size of C^{AR} over the input variables, the AR model is:

$$Z_{t,i}^{AR} = \sum_{k=1}^{n} C_k^{AR} \tilde{Z}_{t-k,i} + d^{AR} \quad (9)$$

The final prediction of the LSTLNet model after the prediction is calculated by the nonlinear structure of deep learning and the linear prediction by AR component is calculated as follows:

$$\hat{Z}_t^{final} = Z_t^F + Z_t^{AR} \quad (10)$$

where $\hat{Z}_t^{final}$ represents the final prediction of the model at time stamp t.

I. Kernel Density Estimator

After making a deterministic forecast by the proposed LSTLNet method, we will extract the PDF for each time interval. In this paper, our proposed method for extracting PDF is Kernel Density Estimator (KDE). KDE is particularly based on a set of observations and random variables from an unknown distribution feature to estimate its density characteristic. The KDE approach neither relies upon at the previous information of the data distribution, nor attaches any hypothesis to the statistics distribution, so it is an estimate that handiest the usage of the sample statistics itself. Suppose Y_n $(n = 1, 2, \ldots, m)$ is a pattern taken from a continuous distribution. The KDE of the density function $f(y)$ at any point y is described as:

$$\hat{f}(y) = \frac{1}{mh} \sum_{n=1}^{m} K\left(\frac{Y_n - y}{h}\right) \quad (11)$$

where, m is the number of input variables in each interval, $K(\cdot)$ is a kernel function, and h is the bandlimit value. A kernel distribution is described via a smoothing characteristic and a bandlimit value, which manipulate the smoothness of the ensuing density curve. Here, the Gaussian kernel function [35] is used, as presented by:

$$K(\chi) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\chi^2} \quad (12)$$

J. Loss Function

In the training process of the neural network-based time series forecasting, a square error-based loss function is usually utilized. However, the square error-based loss function might lead to large gradient values or even to overfitting. Therefore, it is possible that the training algorithm (Adam optimization) cannot converge to the optimal values during the training process. To address this issue, this paper uses the Huber loss function [36] is combined loss function based on mean absolute error (MAE) and mean square error (MSE), as:

$$L_\delta = \begin{cases} \frac{1}{2}\left(y - f(x)\right)^2, & if\ |y - f(x)| \le \delta \\ \delta|y - f(x)| - \frac{1}{2}\delta^2, & otherwise \end{cases} \quad (13)$$

where L_δ is the loss function, and y and $f(x)$ show the actual and predicted values and hyperparameter δ.

The Huber loss function involves an additional variable, δ, which is considered as a hyperparameter. To find optimal value for the hyperparameter, this paper uses Cross-Validation [37] as the hyperparameter optimization algorithm. By finding the optimal value for δ, the robustness of the designed deep network enhances against outliers and abrupt changes in the non-stationary and complex times series such as REV charging profile.

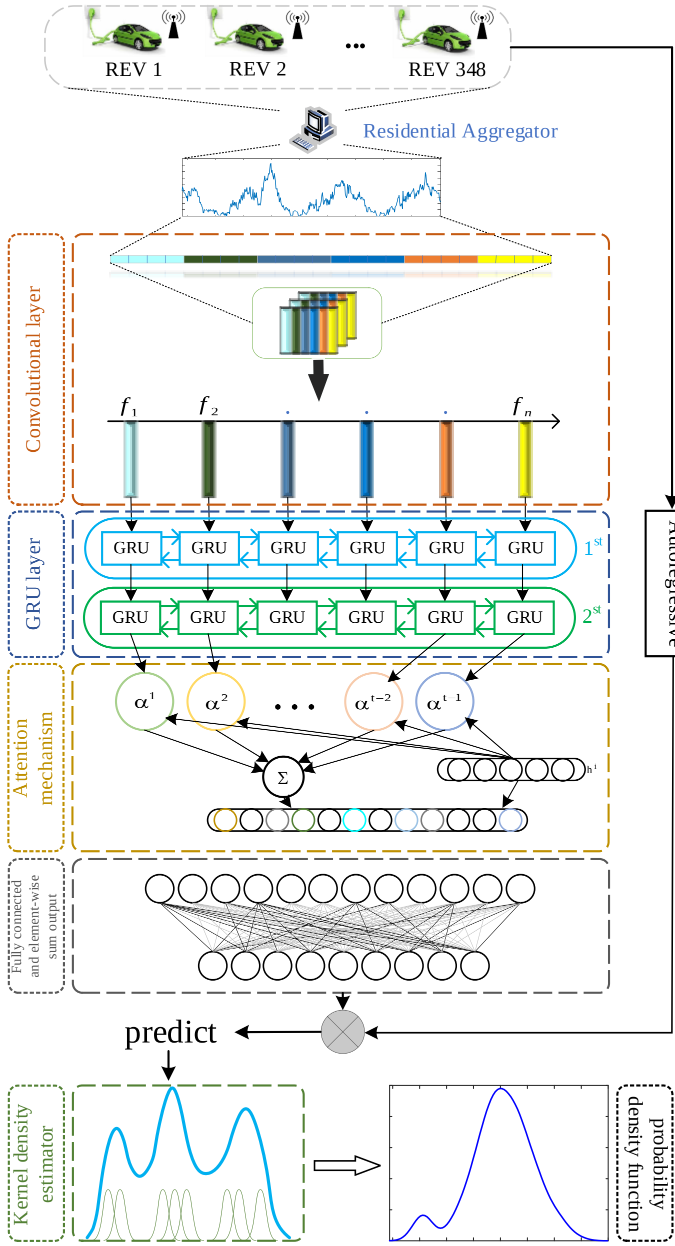

Fig 3. An overview of the PDF estimator structure

K. Overall Structure

In this paper, the proposed LSTLNet model is a flexible architecture used to accurately predict the charge required by electric vehicles. Then KDE is used to estimate the PDF at each time interval. The overall framework of proposed probabilistic REVs forecasting is shown in Fig 3. In the following, the process of the proposed network is given in detail:

1- The actual charging profile of REVs is organized as 2D-input:

$$y_a^{N_i} = \begin{bmatrix} y_a^1 & \cdots & y_a^{N\times 31} \\ \vdots & \ddots & \vdots \\ y_a^N & \cdots & y_a^{N\times 32} \end{bmatrix} \quad (14)$$

The input data is shaped as 4D-tensor with size of (TS, 1, N, 31, 1), where the value of N is the number of samples in a week ($N = 2016$ for time resolution 10min, $N = 336$ for time resolution 30-min, and $N = 168$ for time resolution 1-hr).

2- The input data is fed into the convolutional layers in form of $N \times 1$ different windows, each window composed of 512 filters. The output size of convolution layers is (TS, 1, N, 31, 1).

3-The output of convolutional layers (considering attention mechanism) is flattened by *flatten* techniques and output size is (TS, 1, 100).

4- The output of the previous layer is given to two GRU blocks to extract temporal features of timeseries. The output of these layers has the dimensions (TS, 1, 512), (TS, 1, 128) respectively.

5- In parallel with the nonlinear prediction structure, an AR linear structure makes a linear prediction of the required charge by considering the EVs battery model and their number.

6- The output of both linear and nonlinear structures enters as input to the two fully connected layers. In the first layer, the outputs are generated with dimensions $TS \times 2$ and the second layer produces the prediction output with dimensions $TS \times 1$.

7-Finally, the values obtained from the previous step are given to a KDE structure to estimate the PDF at each time interval.

It is worthwhile to note that the dropout technique is used to prevent overfitting. Besides, the Huber loss function might prevent gradient explosion and overfitting.

IV. SIMULATION AND RESULTS

To test the performance of the designed probabilistic deep-based approach, actual data is adopted from [38] within 10-min time resolution, 30-min time resolution, 60-min time resolution. The actual data involves the charging profile of 348 different REVs in 200 different houses, where some house owners charged more than one EV. The data is recorded from 1st January 2010 to 30th December 2010. The REVs are located in the Midwest region of the United States of America.

In the forecasting process, about 60% is devoted to the training process, approximately 20% is devoted to the validation, and the remaining 20% has considered for the testing process. The results are implemented in TensorFlow in a computer with Intel Core i7-6500U CPU@ 3.1 GHz, 48-GB RAM.

To evaluate the probabilistic forecasting performance, two different accuracy criteria have been selected including mean absolute percentage error (MAPE), root mean square error (RMSE):

$$MAPE = \left(\frac{1}{n}\sum_1^n \left|\frac{Y_a - Y_p}{Y_a}\right|\right) \times 100 \quad (15)$$

$$RMSE = \sqrt{\frac{\sum_1^n (Y_a - Y_p)^2}{n}} \quad (16)$$

where Y_a and Y_p show the actual and predicted values, and n is the number of samples. The continuous ranked probability score (CRPS) examines the calibration and sharpness of the forecasted PDF concurrently, as [52]:

$$CRPS = \frac{1}{N}\sum_{t=1}^N \int_{-\infty}^{+\infty} \left(CDF(Y_p) - CDF(Y_a)\right)^2 dy \quad (17)$$

where $\mathrm{CDF}(Y_p)$ and $\mathrm{CDF}(Y_a)$ show the predictive and actual cumulative distribution function [31]. Cross-entropy (CE) is some other metric this is applied in this paper for probabilistic REVs charge forecasting evaluate, which is described as:

$$\mathrm{CE} = -\sum_{t=1}^{N} \mathrm{P}(\hat{\mathrm{Y}}|\mathrm{X}) \log\left(\mathrm{P}(\hat{Y}|\mathrm{X})\right) = -\sum_{t=1}^{N} \frac{1}{\mathrm{N}} \log\left(\mathrm{P}(\hat{Y}|\mathrm{X})\right) \quad (18)$$

CE is extra sensitive to uncommon events than CRPS. If the measured charge values are very specific to the mean value of charge estimation, with CRPS, probabilistic REVs charge forecasting approach consequences are superb. However, with CE, if this perturbation is out of the distribution, CE suggests that the probabilistic REVs charge forecasting technique has endless error.

For the sake of comparison, three different shallow-based structures including artificial neural network (ANN), k-nearest neighbour (kNN), and support vector machine (SVM) have been considered. Besides, three strong deep structures, including long short-term memory (LSTM), GRU, and two-dimensional convolutional neural network (2D-CNN) are considered to verify the superiority of the proposed network in comparison with deep networks. The detailed information on these methods is given:

1- ANN has 4 hidden layers, 64 hidden neurons in the first layer, 128 and 86 hidden neurons in the next two parallel layers and one hidden neuron in the last one, 20% dropout in each layer, 70% train data, 30% validation data, ReLU activation function, Adam optimizer.
2- SVM is implemented based on radial basis function (RBF) kernel and cross-validation.
3- KNN performs with one nearest neighbor based on Euclidean distance between two variables and Bayes' discussion rule.
4- LSTM has 4 hidden layers with 20% dropout in each layer, 70% train data, 30% validation data, 500 epochs, 256 units, ReLU activation function and Adam optimizer.
5- GRU is implemented with ReLU activation function, 70% train data, 30% validation data, 500 epochs, 256 units and consists of three hidden layers.
6- 2D-CNN is utilized with 128 filters, 500 epochs, 70% train data, 30% validation data, and a maximum pooling size of four, with a ReLU activation function.

A. *Case I: 10-min time resolution*

In this subsection, the results of the designed deep structure are examined on the dataset within 10-min time resolution. Predicted PDF for diverse hours of a day and the related actual values acquired through the usage of the designed LSTLNet are proven in Fig.4. Figs 4(b), 4(a), 4(c) and 4(d) suggests individual PDFs for off-peak (01:00), mid-peak (08:20) and peak (11:50 and 18:30) hours, respectively. The real value of REVs charging profile also are shown in Fig.4, that lets in you to verify the closeness of the PDFs anticipated with the aid of manner of the proposed LSTLNet. In addition, the sharpness of the extracted PDF is clearly shown in Fig.4. To explicitly reveal the performance, the prediction durations with look-ahead time up to 144-time intervals received with the aid of the proposed LSTLNet and real observations in a random day (November 19, 2010), are depicted in Fig.5, Wherein the confidence covers the variety of 10%-80%. Fig.5 suggests that the proposed LSTLNet can cover the commentary in built PIs.

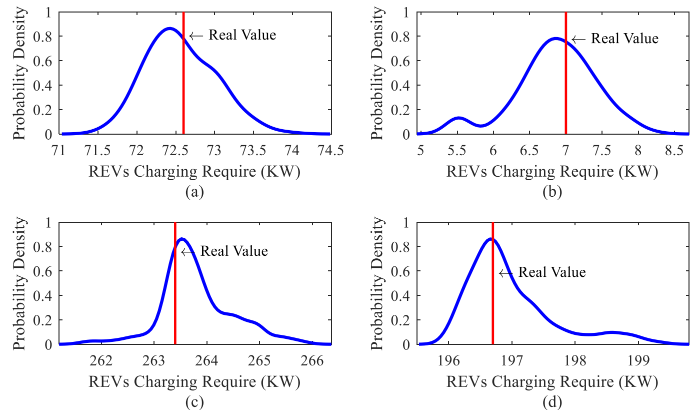

Fig. 4. Predicted PDFs with 10-min time resolution (a) 1:00 (b) 8:20 (c) 11:50 (d) 18:30

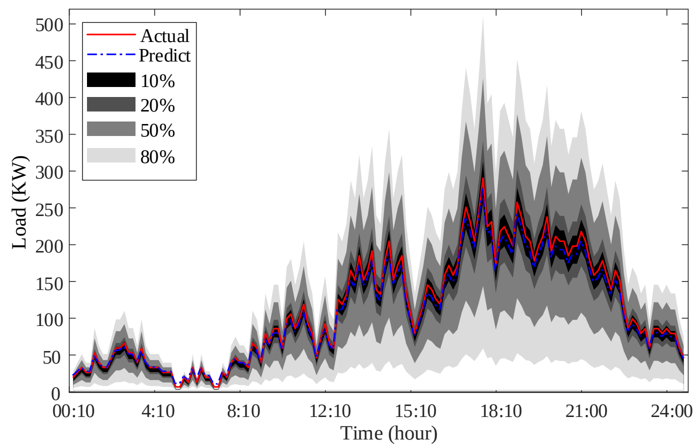

Fig. 5. Predicted intervals with 10-min time resolution in a sample day

Table I compares the results obtained from the proposed method with 6 different methods. The superiority of the proposed method is obvious based on Table I. According to the RMSE accuracy criterion, the proposed method shows about 28.623%, 34.948%, and 43.219% improvement in forecasting accuracy compared to 2D-CNN, GRU, and LSTM as the deep methods. Furthermore, Table I shows that the proposed deep network performs significantly more accurately than shallow-based methods, approximately 50.191%, 53.073%, and 54.932% compared to KNN, SVM, and ANN methods, respectively, in terms of NRMSE. It is worthwhile to note that the accuracy criteria are obtained based on the whole testing dataset.

Table I accuracy criteria in 10-min time resolution

Forecasting method	**Median**		**CRPS**	**CE**
	MAPE%	**RMSE**		
The proposed approach+KDE	1.1875	0.4551	0.01281	3.6059
2D-CNN+KDE	1.4769	0.6376	0.01750	5.7434
GRU+KDE	1.6835	0.6996	0.01922	6.8462
LSTM+KDE	1.7661	0.8015	0.02252	8.4861
kNN+KDE	2.0663	0.9137	0.02612	9.4196
SVM+KDE	2.1723	0.9698	0.02702	9.5892
ANN+KDE	2.2429	1.0098	0.03395	10.0136

B. *Case II: 30-min time resolution*

In this case, the results of some probabilistic time-series forecasting methods on the REV charging profile within 30-min time resolution have been discussed. Firstly, a comparison between the PDF forecasted and actual data is illustrated in Fig 6.

Accordingly, the results are almost the same as the actual data in a sample day, therefore, the high level of the accuracy of the proposed network is clear based on Fig 6. To illustrate the accuracy of the proposed method in detecting uncertainties and its ability to predict the nonlinear charge profile of REVs, Fig 7, compares the predicted values over 48 time periods with PIs in the range of 10% to 80%.

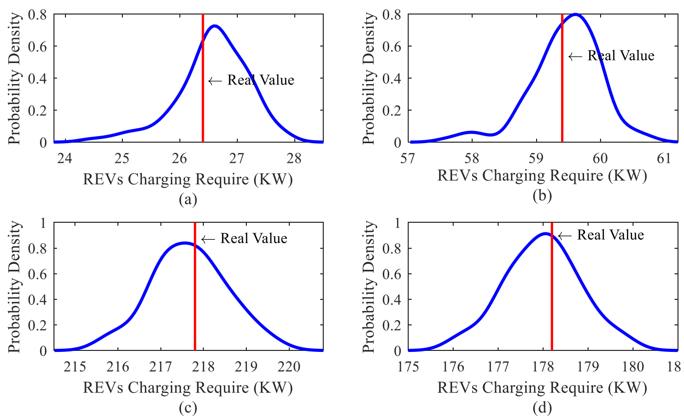

Fig. 6. Predicted PDFs with 30-min time resolution (a) 00:00 (b) 7:30 (c) 11:30 (d) 17:00

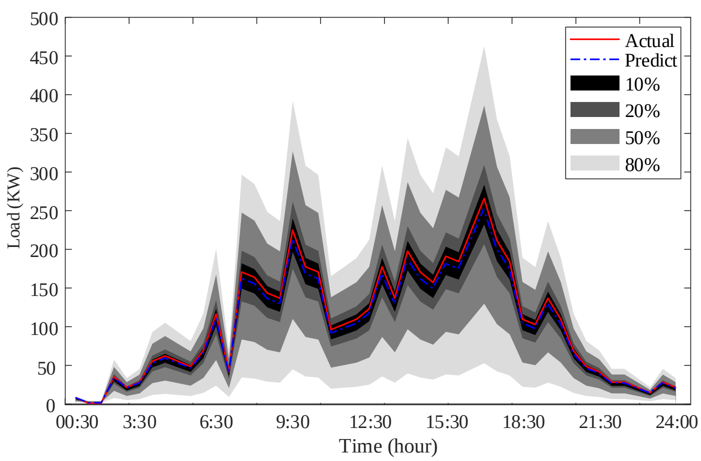

Fig. 7. Predicted intervals with 30-min time resolution in a sample day

In addition, the results of the designed probabilistic LSTLNet are compared in Table II based on four different criteria. The probabilistic proposed method is significantly superior in comparison with previously presented deep and shallow-based networks. For instance, in terms of CE, the improved accuracy of the proposed methods is 65.473%, 64.32%, 63.177%, 52.907%, 40.664%, and 35.988%, respectively.

Table III. accuracy criteria in 30-min time resolution

Forecasting method	Median		CRPS	CE
	MAPE%	RMSE		
The proposed approach+KDE	1.1180	0.4549	0.01640	3.4329
2D-CNN+KDE	1.3762	0.5529	0.01458	5.3629
GRU+KDE	1.5630	0.6285	0.01613	5.7855
LSTM+KDE	1.7104	0.6882	0.01768	7.2897
kNN+KDE	1.9670	0.8961	0.02583	9.3227
SVM+KDE	2.1389	0.9522	0.02751	9.6213
ANN+KDE	2.1846	0.9867	0.03237	9.9428

C. *Case III: 60-min time resolution*

The actual data is also recorded within 60-min time resolution. Thus, this case discusses the results of the proposed deep method on the dataset within 60-min time resolution. The highly accurate performance of the designed network is clear based on Fig 8. Figure 8 compares the actual and predicted PDF of REVs charging profile. Fig 9 shows the PIs, the placement of the predicted values in the middle of which well illustrates the accuracy of the method used to predict the REVs charge. Furthermore, the accuracy criteria of the several data-driven-based REV charging profile forecasting methods are depicted in Table III. Based on CRPS, the results show improved accuracy approximately 23.293%, 32.977%, 38.748% in comparison with 2D-CNN, GRU, and LSTM. In comparison with shallow-based networks, the proposed method+KDE improves the accuracy of KNN to about 46.889%, SVM to 57.959%, and ANN to 64.314%.

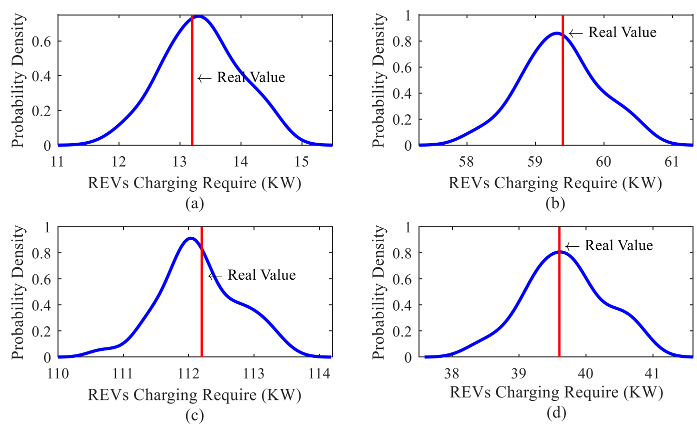

Fig. 8. Predicted PDFs with 60-min time resolution (a) 1:00 (b) 6:00 (c) 18:00 (d) 23:00

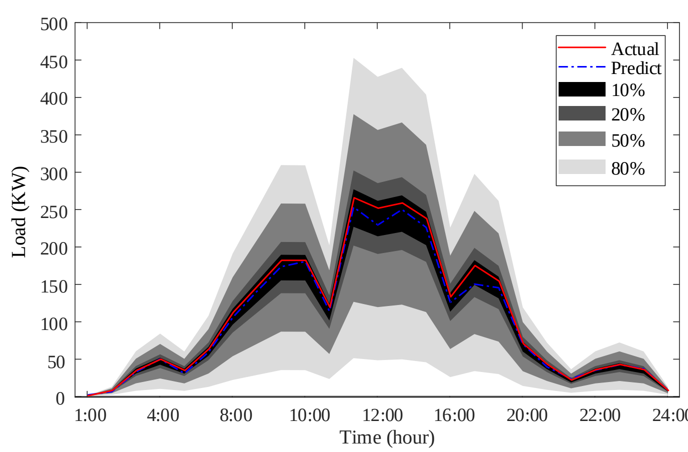

Fig. 9. Predicted intervals with 60-min time resolution in a sample day

Table III. accuracy criteria in 60-min time resolution

Forecasting method	Median		CRPS	CE
	MAPE%	RMSE		
The proposed approach+KDE	1.0536	0.2842	0.01067	3.2371
2D-CNN+KDE	1.2974	0.3827	0.01391	5.5587
GRU+KDE	1.4773	0.4285	0.01592	5.7809
LSTM+KDE	1.6004	0.4743	0.01742	6.7712
kNN+KDE	1.8259	0.8127	0.02009	8.6704
SVM+KDE	1.9837	0.8861	0.02538	9.0018
ANN+KDE	2.0643	0.9604	0.02990	9.6720

V. CONCLUSION AND FUTURE WORK

Nowadays, planning for optimal energy consumption is considered a very important matter. A significant percentage of energy is spent on the transportation industry. On the other hand, in the last decade, people have significantly turned to the use of electric vehicles, especially residential electric vehicle. Therefore, forecasting the required charge of residential electric vehicles as a significant load in the power system, allows the network operator to plan for charging and discharging residential electric vehicles and take steps towards optimal energy consumption. While it seems that presenting PDFs in look-ahead times is very important for probabilistic analysis and thus the performance of power systems. This paper provides an overall overview of the probabilistic forecasting approach based on the combination of the new deep structure, namely LSTLNet and KDE to provide full-statistical information of the REV charging information for the look-ahead times. In the proposed network, by using convolution layers, attention mechanism, and GRU layers, the proposed approach can understand complex, non-stationary, volatile, and uncertain behaviour of REVs directly from raw data. The proposed end-to-end deep network is also integrated with AR to take into account the effects of the number of EVs, the type of EVs, and the type of EV battery. Numerical analysis on the actual dataset of 348 REVs in the east of the USA shows that the proposed network is approximately at least 30% more accurate than the known deep networks and about 60% more accurate than the known shallow methods presented in previous works.

In our future work, in addition to designing PDF forecasting and estimation methods with more appropriate accuracy, we plan to design a stochastic framework for the operation and planning of the EV aggregators based on full statistics information in terms of predicted PDFs.

REFERENCES

[1] G. Aydin, I. Karakurt, and K. Aydiner, "Analysis and Mitigation Opportunities of Methane Emissions from the Energy Sector," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects,* vol. 34, no. 11, pp. 967-982, 2012/04/06 2012.

[2] A. Azadeh and S. Tarverdian, "Integration of genetic algorithm, computer simulation and design of experiments for forecasting electrical energy consumption," *Energy Policy,* vol. 35, no. 10, pp. 5229-5241, 2007/10/01/ 2007.

[3] M. Moschella, M. A. A. Murad, E. Crisostomi, and F. Milano, "Decentralized Charging of Plug-In Electric Vehicles and Impact on Transmission System Dynamics," *IEEE Transactions on Smart Grid,* vol. 12, no. 2, pp. 1772-1781, 2021.

[4] J. Yano, T. Muroi, and S.-i. Sakai, "Rare earth element recovery potentials from end-of-life hybrid electric vehicle components in 2010–2030," *Journal of Material Cycles and Waste Management,* vol. 18, no. 4, pp. 655-664, 2016/09/01 2016.

[5] Electric Vehicles. International Energy Agency (IEA). Paris [Online], April). Available: http:// www.iea.org/fuels-and-technologies/electric-vehicles.

[6] M. N. Akhter, S. Mekhilef, H. Mokhlis, and N. M. Shah, "Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques," *IET Renewable Power Generation*, vol. 13, no. 7, pp. 1009-1023Available: https://digital-library.theiet.org/content/journals/10.1049/iet-rpg.2018.5649

[7] J. Bailey, A. Miele, and J. Axsen, "Is awareness of public charging associated with consumer interest in plug-in electric vehicles?," *Transportation Research Part D: Transport and Environment,* vol. 36, pp. 1-9, 2015/05/01/ 2015.

[8] M. S. ElNozahy and M. M. A. Salama, "A Comprehensive Study of the Impacts of PHEVs on Residential Distribution Networks," *IEEE Transactions on Sustainable Energy,* vol. 5, no. 1, pp. 332-342, 2014.

[9] E. Sortomme, M. M. Hindi, S. D. J. MacPherson, and S. S. Venkata, "Coordinated Charging of Plug-In Hybrid Electric Vehicles to Minimize Distribution System Losses," *IEEE Transactions on Smart Grid,* vol. 2, no. 1, pp. 198-205, 2011.

[10] M. Alipour, B. Mohammadi-Ivatloo, M. Moradi-Dalvand, and K. Zare, "Stochastic scheduling of aggregators of plug-in electric vehicles for participation in energy and ancillary service markets," *Energy,* vol. 118, pp. 1168-1179, 2017/01/01/ 2017.

[11] Y. Wang, N. Zhang, Y. Tan, T. Hong, D. S. Kirschen, and C. Kang, "Combining Probabilistic Load Forecasts," *IEEE Transactions on Smart Grid,* vol. 10, no. 4, pp. 3664-3674, 2019.

[12] M. H. Amini, O. Karabasoglu, M. D. Ilić, K. G. Boroojeni, and S. S. Iyengar, "ARIMA-based demand forecasting method considering probabilistic model of electric vehicles' parking lots," in *2015 IEEE Power & Energy Society General Meeting*, 2015, pp. 1-5.

[13] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Lower upper bound estimation method for construction of neural network-based prediction intervals," *IEEE transactions on neural networks,* vol. 22, no. 3, pp. 337-346, 2010.

[14] L. Buzna *et al.*, "An ensemble methodology for hierarchical probabilistic electric vehicle load forecasting at regular charging stations," *Applied Energy,* vol. 283, p. 116337, 2021/02/01/ 2021.

[15] S. Peng, G. Zheng, and J. Zou, "Prediction of Probability Density of Electric Vehicle Load Based on Deep Learning QRDCC Model," in *2019 IEEE 3rd Conference on Energy Internet and Energy System Integration (EI2)*, 2019, pp. 1225-1229.

[16] M. Afrasiabi, M. Mohammadi, M. Rastegar, L. Stankovic, S. Afrasiabi, and M. Khazaei, "Deep-Based Conditional Probability Density Function Forecasting of Residential Loads," *IEEE Transactions on Smart Grid,* vol. 11, no. 4, pp. 3646-3657, 2020.

[17] M. P. Almeida, O. Perpinan, and L. Narvarte, "PV power forecast using a nonparametric PV model," *Solar Energy,* vol. 115, pp. 354-368, 2015.

[18] J. C. López, M. J. Rider, and Q. Wu, "Parsimonious Short-Term Load Forecasting for Optimal Operation Planning of Electrical Distribution Systems," *IEEE Transactions on Power Systems,* vol. 34, no. 2, pp. 1427-1437, 2019.

[19] M. H. Amini, A. Kargarian, and O. Karabasoglu, "ARIMA-based decoupled time series forecasting of electric vehicle charging demand for stochastic power system operation," *Electric Power Systems Research,* vol. 140, pp. 378-390, 2016/11/01/ 2016.

[20] E. S. Xydas, C. E. Marmaras, L. M. Cipcigan, A. S. Hassan, and N. Jenkins, "Forecasting Electric Vehicle charging demand using Support Vector Machines," in *2013 48th International Universities' Power Engineering Conference (UPEC)*, 2013, pp. 1-6.

[21] M. Majidpour, C. Qiu, P. Chu, R. Gadh, and H. R. Pota, "A novel forecasting algorithm for electric vehicle charging stations," in *2014 International Conference on Connected Vehicles and Expo (ICCVE)*, 2014, pp. 1035-1040.

[22] Q. Liu, Y. Shen, L. Wu, J. Li, L. Zhuang, and S. Wang, "A hybrid FCW-EMD and KF-BA-SVM based model for short-term load forecasting," *CSEE Journal of Power and Energy Systems,* vol. 4, no. 2, pp. 226-237, 2018.

[23] M. Zolfaghari and M. R. Golabi, "Modeling and predicting the electricity production in hydropower using conjunction of wavelet transform, long short-term memory and random forest models," *Renewable Energy,* vol. 170, pp. 1367-1381, 2021/06/01/ 2021.

[24] H. Lan, C. Zhang, Y.-Y. Hong, Y. He, and S. Wen, "Day-ahead spatiotemporal solar irradiation forecasting using frequency-based hybrid principal component analysis and neural network," *Applied Energy,* vol. 247, pp. 389-402, 2019/08/01/ 2019.

[25] S. Afrasiabi, M. Afrasiabi, B. Parang, and M. Mohammadi, "Integration of accelerated deep neural network into power transformer differential protection," *IEEE Transactions on Industrial Informatics,* vol. 16, no. 2, pp. 865-876, 2019.

[26] M. Afrasiabi, M. Mohammadi, M. Rastegar, and A. Kargarian, "Multi-agent microgrid energy management based on deep learning forecaster," *Energy,* vol. 186, p. 115873, 2019.

[27] M. Fasahat and M. Manthouri, "State of charge estimation of lithium-ion batteries using hybrid autoencoder and Long Short Term Memory neural networks," *Journal of Power Sources,* vol. 469, p. 228375, 2020/09/01/ 2020.

[28] S. Ryu, J. Noh, and H. Kim, "Deep Neural Network Based Demand Side Short Term Load Forecasting," *Energies,* vol. 10, no. 1, 2017.

[29] J. L. Elman, "Finding structure in time," *Cognitive science,* vol. 14, no. 2, pp. 179-211, 1990.

[30] J. Li *et al.*, "A Novel Hybrid Short-Term Load Forecasting Method of Smart Grid Using MLR and LSTM Neural Network," *IEEE Transactions on Industrial Informatics,* vol. 17, no. 4, pp. 2443-2452, 2021.

[31] S. Afrasiabi, M. Afrasiabi, B. Parang, and M. Mohammadi, "Designing a composite deep learning based differential protection scheme of power transformers," *Applied Soft Computing,* vol. 87, p. 105975, 2020/02/01/ 2020.

[32] S. M. J. Jalali, S. Ahmadian, A. Khosravi, M. Shafie-khah, S. Nahavandi, and J. P. S. Catalao, "A Novel Evolutionary-based Deep Convolutional Neural Network Model for Intelligent Load Forecasting," *IEEE Transactions on Industrial Informatics,* pp. 1-1, 2021.
[33] J. Zhu *et al.*, "Electric Vehicle Charging Load Forecasting: A Comparative Study of Deep Learning Approaches," *Energies,* vol. 12, no. 14, 2019.
[34] J. Zhu, Z. Yang, Y. Guo, J. Zhang, and H. Yang, "Short-Term Load Forecasting for Electric Vehicle Charging Stations Based on Deep Learning Approaches," *Applied Sciences,* vol. 9, no. 9, 2019.
[35] J. Che and J. Wang, "Short-term load forecasting using a kernel-based support vector regression combination model," *Applied Energy,* vol. 132, pp. 602-609, 2014/11/01/ 2014.
[36] D. Li, M. Han, and J. Wang, "Chaotic Time Series Prediction Based on a Novel Robust Echo State Network," *IEEE Transactions on Neural Networks and Learning Systems,* vol. 23, no. 5, pp. 787-799, 2012.
[37] C. Chen, C. Yan, N. Zhao, B. Guo, and G. Liu, "A robust algorithm of support vector regression with a trimmed Huber loss function in the primal," *Soft Computing,* vol. 21, no. 18, pp. 5235-5243, 2017.
[38] M. Muratori, "Impact of uncoordinated plug-in electric vehicle charging on residential power demand," *Nature Energy,* vol. 3, no. 3, pp. 193-201, 2018/03/01 2018.