

Article

Electric Vehicle Charging Load Forecasting: A Comparative Study of Deep Learning Approaches

Juncheng Zhu ¹, Zhile Yang ^{2,3,*}, Monjur Mourshed ³, Yuanjun Guo ², Yimin Zhou ², Yan Chang ⁴, Yanjie Wei ² and Shengzhong Feng ²

¹ School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China

² Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

³ School of Engineering, Cardiff University, Cardiff CF24 3AA, UK

⁴ School of Software Engineering, University of Science and Technology of China, Hefei 230026, China

* Correspondence: zl.yang@siat.ac.cn

Received: 19 June 2019; Accepted: 5 July 2019; Published: 13 July 2019



Abstract: Load forecasting is one of the major challenges of power system operation and is crucial to the effective scheduling for economic dispatch at multiple time scales. Numerous load forecasting methods have been proposed for household and commercial demand, as well as for loads at various nodes in a power grid. However, compared with conventional loads, the uncoordinated charging of the large penetration of plug-in electric vehicles is different in terms of periodicity and fluctuation, which renders current load forecasting techniques ineffective. Deep learning methods, empowered by unprecedented learning ability from extensive data, provide novel approaches for solving challenging forecasting tasks. This research proposes a comparative study of deep learning approaches to forecast the super-short-term stochastic charging load of plug-in electric vehicles. Several popular and novel deep-learning based methods have been utilized in establishing the forecasting models using minute-level real-world data of a plug-in electric vehicle charging station to compare the forecasting performance. Numerical results of twelve cases on various time steps show that deep learning methods obtain high accuracy in super-short-term plug-in electric load forecasting. Among the various deep learning approaches, the long-short-term memory method performs the best by reducing over 30% forecasting error compared with the conventional artificial neural network model.

Keywords: load forecasting; LSTM; electric vehicles; deep learning

1. Introduction

Uninterrupted supply of electricity is crucial to the functioning of the modern civilization. Today's electricity grid is highly complex and increasingly vulnerable to the potential disruptions. Load forecasting has, therefore, been a key measure in power system planning, scheduling and operation. With the increasing penetration of variable renewable energy resources, accurate forecasting of both generation and demand profiles are important for effective and economic dispatching of power contributors. Load forecasting can be categorized into short-, medium- and long-term depending on the time span or resolution [1]. The short-term load forecasting is useful for utility optimal operations and scheduling while the long-term load forecasting is delivered in the system planning stage. Moreover, minute-level load forecasting is super-short-term load forecasting [2] and has been utilized in real-time power quality and security monitoring.

The electrification of the transportation sector is seen as an effective means to reduce greenhouse gas emissions from the burning of fossil fuel. Other environmental concerns such as urban air quality and related health impacts have also prompted policy makers and stakeholders to opt for the popularization of electric vehicles (EVs) [3] in replacing traditional internal combustion engine (ICE)

based vehicles. EVs can be considered as zero emissions vehicle during its operation when electricity from renewable sources are used to charge them. However, the rapid development of the EV industry is introducing new challenges to the existing power system structure owing to their large battery capacity [4] and highly stochastic individual charging behavior.

Model based load forecasting techniques include statistical models using recursive and traditional mathematical tools [5–8], and artificial intelligence models involving various state-of-the-art machine learning approaches [9,10]. Traditional forecasting methods are generally straightforward and utilizing explainable presentations in the model composition, whereas artificial intelligence methods produce grey or black-box models in generating the forecasting results. Due to the strong adaptive learning and generalization ability, artificial neural network (ANN) has become successful in delivering load forecasting tasks [11]. However, increasing resolution and dimensionality of the emerging dataset challenge canonical ANN approaches. Deep learning methods have been on the spotlight and seen remarkable success in image semantic segmentation and feature classification [12–14], natural language processing [15] and various computational extensive science and engineering fields [16]. The load forecasting problem is a featured time-series problem which is of strong similarity with natural language processing, where the deep learning method would have the potential to effectively contribute.

In the distribution grid level, EV charging load exerts strong pressures due to its highly periodical and fluctuate characteristics. The EV demand curve would see significant peak-valley differences and large spikes in featured time slots particular in a super-short-term time scale. A more precise forecasting of such novel load type is potential to significantly contribute to the power system operator both from economic and security issues, where more powerful tools are in need. The key contributions of the paper are as below:

(1) New model: a novel exploratory super-short-term multi-step load forecasting model is proposed particularly for modeling the super-short-term EV charging load. The minute-level extra-short-term model plays a crucial role in the operation and maintenance of the EV charging aggregators as well as in the power flow analysis and control [17]. The impact of multiple time step of the historical data on the forecasting accuracy would be evaluated.

(2) New method: a new EV load forecasting deep learning framework is established and applied in the modeling task, where six conventional and deep learning based methods are comprehensively and comparatively studied in the evaluations on various index criteria.

(3) New scenario: a brand new scenario with real world historical data of a whole year on plug-in electric vehicle charging stations in Shenzhen has been adopted in validating the model effectiveness. According to our knowledge, this is the first study that utilizes real world super-short-term EV charging data rather than simulation data to forecast the multi-time step extra-short-term charging load profiles.

The rest of the paper is organized as follows: Section 2 elaborates the background of the load forecasting and briefly reviews load forecasting methods; Section 3 demonstrates the principle of several featured deep learning method and the corresponding load forecasting framework; Section 4 gives a detailed discussion and analysis of the charging data we used; the experimental results of comparative studies are shown in Section 5; in the end, Section 6 concludes the paper.

2. Literature Study

Based on the characteristics of load-based time dimension expansion, the initial research on load forecasting problem is based on the time series prediction method proposed by Box et al. [5] in 1976. The method has low input requirements for the load forecasting model, which only considers the time series input of historical data and does not consider other multi-faceted influencing factors that affect the load. The literature [18] provides a load peak model that takes external factors into account such as weather and humidity. Based on the Box–Jenkins method, Hagan et al. [19] proposed an autoregressive moving average model (ARMA) model prediction method, and Juberias et al. [20] established the autoregressive integral moving average model (ARIMA) model to achieve load forecasting. In order to

further improve the accuracy of load forecasting, many hybrid methods of predictive models have been proposed. Jie et al. [21] combine the seasonal exponential adjustment method with the regression methods. Pai and Hong [22] applied support vector machine (SVM) to load forecasting and used the simulated annealing algorithm to select the kernel parameters. Guo et al. [23] utilized time-indexed autoregressive with exogenous terms (ARX) models with two-stage weighted least squares regression for modeling hourly cooling load.

Due to the different types of load and the complexity of influencing factors, the selection of input features and the method in constructing load forecasting models become important. Many intelligent prediction methods have been proposed utilizing more relevant information. The authors in [24] used a random forest approach to build a load forecasting model and the inputs are refined by expert feature selection using fuzzy rules. Feng and Xu [25] proposed an appropriate combinational approach for short-term gas load forecasting based on genetic algorithm-optimized neural network. Kouhi and Keynia [26] proposed a cascade neural network method and used a two-stage feature selection method for selecting the best input. Mahmoud et al. [27] utilized a tuning fuzzy system and ANN method for modeling the medium voltage load. Existing approaches have proved the applicability for the time series based mathematical models and computational intelligence models in solving the load forecasting problem. However, new participants such as renewable generation and EVs have seen more complicated characteristics and higher uncertainties, which challenge the conventional approaches.

In 2006, the deep learning concept was firstly proposed by Hinton et al. [28]. The deep learning methods have stronger nonlinear learning ability, robustness and generalization than the traditional methods, in particular for the large scale data resources. The model trained by deep learning methods could be applied to large scale and intractable scenarios, where appropriate adjustments of only limited hyper-parameters could achieve desired effects. Featured deep learning based network include convolutional neural network (CNN), recurrent neural network (RNN), Deep Boltzmann Machines [29], Stacked AutoEncoder [30], etc. Long short-term memory is an improved RNN, which was first introduced by Hochreiter et al. [31] and aimed to relief the gradient vanish problem of original RNN. In 2016, Marino et al. [32] used long-short-term memory (LSTM) for building energy load forecasting and discussed the effect of neuron nodes numbers on the forecasting error. In 2017, Kong et al. [33] applied LSTM to the residential load forecasting, and LSTM showed the best performance compared to other counterparts. Zheng et al. [34] presented a hybrid LSTM model for short-term load forecasting and obtained comparatively good results compared with other counterparts. In 2018, Bouktif et al. [35] proposed an optimal LSTM model for electric load forecasting using feature selection and genetic algorithm. This method trained several linear and nonlinear machine learning algorithms and selected the best performance algorithm as the baseline. These case studies have effectively verified the feasibility and superiority of deep learning in particular LSTM methods in the field of load forecasting.

In recent years, plug-in electric vehicles (PEVs) have emerged worldwide. The large power capacity of their battery provides an unprecedented challenge to the existing power system. Accurate and efficient load forecasting for PEV charging is critical for the maintenance and operation of charging stations [36,37]. Mu et al. [38] presented a Spatial-Temporal model to evaluate the impact of large scale deployment of PEVs on urban distribution networks. Qian et al. [39] proposed a methodology for modeling and analyzing the load in a distribution system considering PEV battery charging, and adopted Monte Carlo simulation in scenario generations. Alizadeh et al. [40] proposed a stochastic model based on queuing theory for PEV charging load. Luo et al. [41] proposed a Monte Carlo simulation based model to forecast the charging load of the PEVs in China. Lu et al. [42] utilized a random forest to forecast the 15 min level EV charging data. However, these traditional methods are difficult for quantifying the external factors that affect the charging load of PEVs, and it is impossible to establish a deterministic model. In our previous study [43], the deep learning method is used for hourly level PEV load forecasting and obtained well performance. However, the minute level super-short-term forecasting is more challenging. In this paper, the super-short-term PEV charging load model is established using minute level historical data for training, validation and test. Moreover,

the performance of multiple deep learning methods in solving the super-short-term PEV forecasting problem is comprehensively evaluated.

3. The Deep Learning Based PEV Charging Load Forecasting Framework

In 2015, LeCun et al. [44] systematically reviewed the featured methods and the applications of deep learning in the literature, where CNN and RNN have been among the most powerful tools in solving image based and time sequential data problems, respectively. Specifically, the recurrent nets have shone light on sequential data such as text and speech. The load forecasting data are sequential data that is similar to the text and speech, and RNN is the sequence-based model, which is potential to demonstrate better performance than traditional methods in solving time series problems.

3.1. RNN Model

For load forecasting problem, when the load value at time $t-1$ as the input of RNN, the model can output the load value of time t , RNN models can better capture the characteristics of input data by using recurrent structure shown in Figure 1.

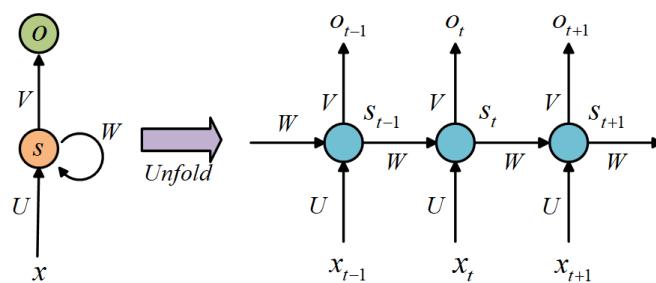


Figure 1. Basic RNN structure.

For time t :

$$S_t = \phi(Us_t + WS_{t-1} + b_1), \quad (1)$$

$$o_t = \phi(VS_t + b_2), \quad (2)$$

$$\hat{y}_t = \phi(o_t), \quad (3)$$

where x_t , S_t and o_t denote input, hidden and output unit at t , respectively. Network connection weights are denoted by V , W , U . Moreover, b and \hat{y} represent the bias and predicted output value, and ϕ denotes activation function. With the increase amount and dimension of data, RNN has to remember a lot of information before the time t , which leads to the vanishing gradient problem. In light of this, the LSTM method is proposed for solving this problem.

3.2. LSTM Model

The LSTM structure maintains the key logic of original recurrent network scheme. The major difference between LSTM and RNN is that the LSTM method adds a “processor” to the algorithm to determine whether the input information is useful or not. The processor item is named ‘cell’ to include the all features of LSTM modules. As shown in Figure 2, three gates are designed in the cell named input gate (i_t), forget gate (f_t) and output gate (o_t), respectively, for maintaining and updating valuable information of the data before time t . The model training method for LSTM is the well adopted back-propagation through time (BPTT) [45]. It has been proved that LSTM is an effective method to solve the problem of long-range dependencies, and it has universal applicability in various learning and prediction problems. The cell states and parameters’ updating scheme is shown as follows:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f), \quad (4)$$

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \quad (5)$$

$$\tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c), \quad (6)$$

$$C_t = i_t * \tilde{C}_t + f_t * C_{t-1}, \quad (7)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \quad (8)$$

$$h_t = o_t * \tanh(C_t), \quad (9)$$

where i_t , f_t , and o_t are input, forget and output gate, respectively. h_{t-1} is the output at $t - 1$ time slot, and x_t is the input at current moment, and C_{t-1} is the memory from previous block. The forget gate (f_t) reads the information in h_{t-1} and x_t , then outputs a value between 0 and 1 for the cell state C_{t-1} , and 1 denotes “completely reserved” and 0 represents “completely discarded”. The input gate i_t decides how much new information is to be added into the cell state, and the first stage is the *sigmod* layer that decides which information will be updated. The second stage is that the *tanh* layer generates a vector which is the new candidate value \tilde{C}_t . The memory of current block C_t is generated by the item’s accumulation of previous block and input gate. Finally, the output gate (o_t) outputs a value that determines the cell state, where W and b are weight and bias. σ and *tanh* are activation functions shown as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}}, \quad (10)$$

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (11)$$

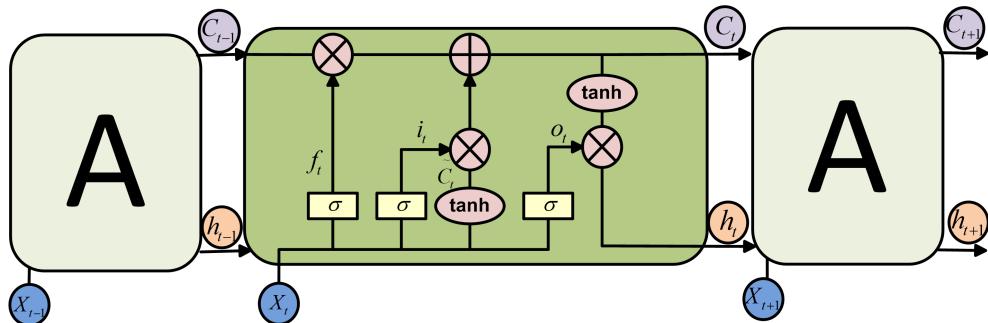


Figure 2. The structure of LSTM.

3.3. The LSTM Based PEV Charging Load Forecasting Framework

The framework for PEV charging load forecasting using LSTM is shown in Figure 3. In this case, the model training and validation data are the real world PEV charging load with one minute time intervals. The original data are the accumulated sampling output of the overall charging actions, e.g., from the beginning of each charging process to the end of charging time for each charging post. Therefore, the framework starts from data pre-processing for the input, and the preprocessed input is the PEV charging load power per minute. Then, the whole prepared data set is divided into training, test and validation set. The training set is used to train the model, while the validation set is used to tune the hyper-parameters in order to get the best performance forecasting model, and the test set is used to verify the validity of the model.

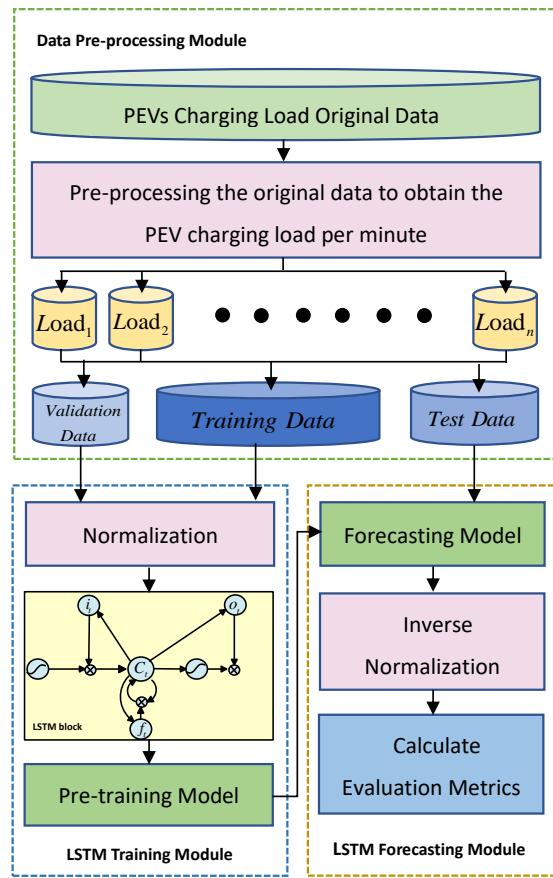


Figure 3. The LSTM based forecasting framework.

The data set should be normalized before being fed into the model, through which the calculations in the training is simplified and the network convergence is accelerated. The normalization formula is as follows:

$$y = \frac{x - x_{min}}{x_{max} - x_{min}}. \quad (12)$$

The next step is to set a time step for look-back, and Figure 4 shows how the time step works. If $TimeStep = 1$, the model uses the information from the previous moment (x_{n-1}) as the input, and the output is the current time information (x_n). If $TimeStep = 3$, the model would utilize n previous moment information, e.g., $x_{n-3}, x_{n-2}, \dots, x_{n-1}$, as the inputs, and the output is the current time information (x_n). Proper hyper-parameters are preset for the LSTM net and the training set data are adopted to train the model using the BPTT method. Finally, test set data are used to evaluate the performance of the model.

After the LSTM block, there is a dense layer that maps the outputs of LSTM block to a single value. The PEV charging load forecast of a one minute time interval is obtained by the inverse normalization of the output data from LSTM block. In this paper, several other deep learning based models are also considered in the comparative studies, and they share the same forecasting framework. Due to the space limitation, the other methods are not detailed in this section. All of the framework modules are implemented in a desktop workstation with 3.0 GHz Intel i7 and 64 GB memory, and the GPU is Geforce Nvidia GTX-1080Ti with all codes running in Keras library [46] with Tensorflow [47] backend.

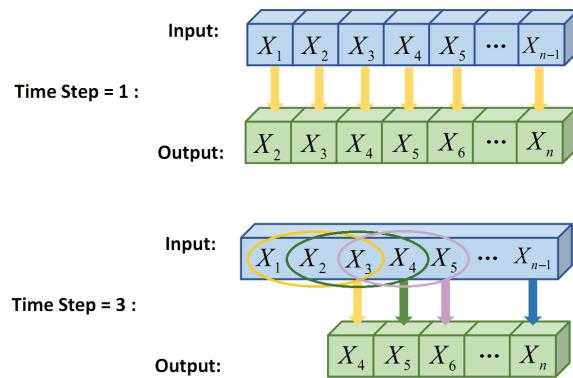


Figure 4. Time step working mechanism.

4. Data Analysis

In dealing with the load forecasting and energy prediction tasks, data-driven approaches have achieved good results compared to other analytical models [48]. The deep learning model adopted in this paper is also a data-driven method, in which the data analysis is vital to train the model for the given task. In this section, the characteristics of the dataset will be comprehensively analyzed.

4.1. Data Statistical System

The dataset used in this paper is collected from a large scale PEV charging station in Shenzhen with photovoltaic panels on the roof and considerable energy storage equipped. The charging station has 64 parking spaces for pure-electrified buses, 12 charging spaces for cars, and 24 built-up charging piles. Our dataset is the charging load collected from the 24 charging piles. The charging station's power distribution mode and data statistics system diagram are shown in Figure 5. The charging station has two power distribution modes: solar carport power generation and battery energy storage. The charging data are stored in the charging pile and transmitted to the data center through the wireless router. The data center is established in the cloud and managed by the centralized platform. The real-time data would be downloaded and delivered into the data processing step that is implemented in the local work station.

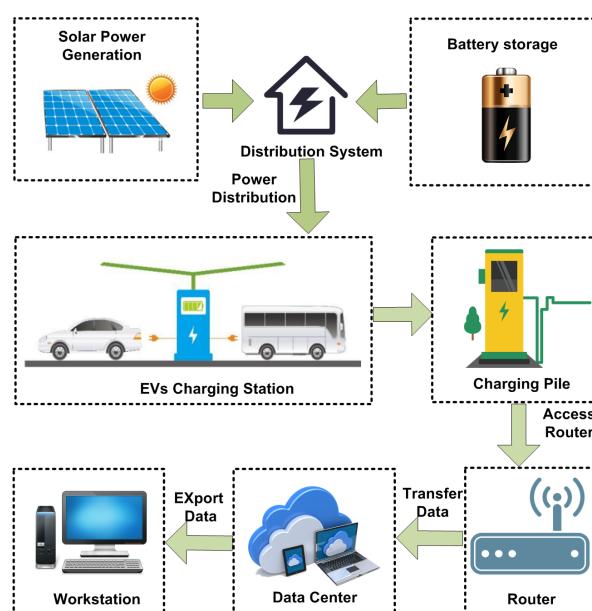


Figure 5. Power distribution mode and data statistics system.

4.2. Data Pre-Processing and Feature Analysis

The original data contain three data types: charging start time, charging end time and total charging amount. The data ranges from 31 March 2017 to 17 July 2018. Due to the fact that many cars choose to charge at night, the charge statistics span covers two days long. In order to obtain the one year data from 1 July 2017 to 30 June 2018, we choose the data from 3 June 2017 to 1 July 2018. Then, the outlier data in the dataset are found and processed with the method elaborated in the following paragraphs.

A fetched load value is firstly judged in Equation (13), and determined by Equation (14) once identified as an outlier or missing data. In this paper, the outlier data are majorly tackled by interpolation methods. The detailed threshold handling method is shown as below:

when:

$$\max\left[\left|\frac{y_{dt} - y_{dt-1}}{y_{dt}}\right|, \left|\frac{y_{dt} - y_{dt+1}}{y_{dt}}\right|\right] \leq \varepsilon_a, \quad (13)$$

where y_{dt} is the fetched load value of time t and ε_a is the predetermined threshold. The outlier data will be replaced by a simple interpolation manner, where y_{dt} is denoted as below:

$$y_{dt} = \frac{y_{dt-1} + y_{dt+1}}{2}. \quad (14)$$

A similar operation is implemented for the error data. It should be noted that the interpolation method is only effective in minor outliers, whereas the large amount of error data may require more specific manner to handle or prevent. Finally, the tool panda is used to split the data at one-minute intervals for one year from 1 July 2017 12:00 a.m. to 30 June 2018 11:59 a.m. At this point, the data pre-processing task finishes and ends.

Shenzhen is located in the southern part of China and belongs to the subtropical monsoon climate. The summer time spans more than six months. It is therefore not reasonable to analyze the data features on a quarterly basis. Figure 6 shows the load curve for half a month. It can be seen from the figure that the PEV charging load as a whole follows a certain fixed pattern, which is related to the fixed route of the bus and the commutes of passengers. Therefore, according to the climate characteristics of Shenzhen, the distribution of charging load is compared by the scenarios of the dry season and the rainy season. The rainy season is from April to September, while the dry season is between October and March. The box plot of charging load distribution per minute is shown in Figure 7. It can be seen that the peak load, median load, upper quantile and lower quantile in the rainy season are all higher than those of the dry season, and the peak load in the rainy season exceeds 40 kW. It shows that the rainy season is pleasant for people to travel and consume more power, due to which the amount of charging vehicles is larger.

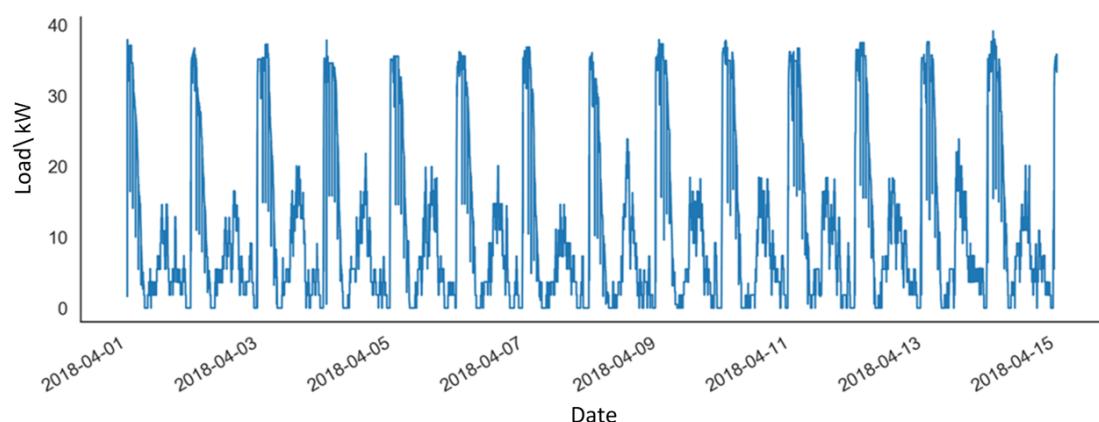


Figure 6. Load curve for half a month data.

In addition to the climate impact, the holidays and work days are normally completely different in charging profiles. Figure 7 also shows the box plot comparison of load distribution for the weekday and holiday in each month. Overall, the peak load and upper quantile of the holidays are higher than the weekdays. In February of 2018, the weekday load distribution has many outliers. The reason is that this month is the Chinese Spring Festival, when people are mostly reunited at home and the charging load is lower than other times. These points are judged as abnormal data in the box plot, but this indicates the authenticity of the data and the impact of holidays on the charging load.

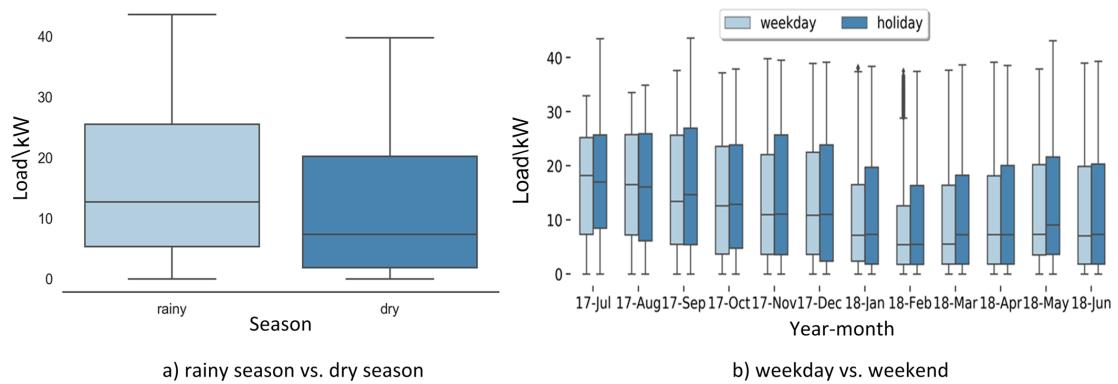


Figure 7. Box plot of EV charging load.

5. Numerical Results for Case Study

5.1. Evaluation Metrics and Error Function

Generally, three popular metrics are used to evaluate the performance of model, including root mean squared error (RMSE), mean absolute percent error (MAPE) and mean absolute error (MAE), etc. The equations of these metrics are shown as follows:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (15)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^n \left(\left| \frac{\hat{y}_i - y_i}{y_i} \right| \times 100 \right), \quad (16)$$

$$MAE = \frac{1}{N} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (17)$$

where N is the number of samples, \hat{y}_i is forecasting value, and y_i is actual value. In this paper, the EV charging facilities have been using the constant charging power. In this regard, to keep this commonly used unit, the minute based charging load of a whole charging station is calculated by the accumulated charging time slots of all the charging piles in each minute. The accumulated charging power could also denote the total charging amount for the overall charging station in every minute. Due to the durability of the PEV batteries, the drivers' driving habits and the period of frequent use of the vehicles have the regularity. Therefore, the sum of the charging load of the piles in a certain period of time may be 0. However, in Equation (16), the actual value is the denominator and should not be '0'. Therefore, in our study, the MAPE metric is not considered in the evaluation of forecasting accuracy. On the other hand, the RMSE and MAE are chosen as the metrics, and a coefficient of determination is adopted to estimate the goodness of fit, named R square, of which the formula is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}, \quad (18)$$

where \bar{y}_i is the average of all samples. The range of R^2 is (0,1), and the closer R^2 is to 1, the higher forecasting accuracy it would be.

5.2. Experimental Setup

The load forecasting model can be divided into univariate models and multivariate models. Some traditional load forecasting methods cannot increase the temperature and other characteristics to express a more complete relationship map because we cannot capture the change law from the individual load variables. The deep learning model is capable of capturing the load variation characteristics from univariate, which is a great improvement for the efficiency of real-time prediction. The final load prediction error confirms the validity of a univariate deep learning prediction model.

In this paper, we choose two datasets of PEV charging load of a whole year ranging from 1 July 2017 to 30 June 2018 including a charging station case and an official charging site aggregator case. For both cases, the charging load time interval is in minutes and an accumulation of 525,600 rows data are considered in total. The data are split into three subsets for different purposes in order to make full use of the data. The proportion of the three subsets are divided as 0.7/0.2/0.1, among which the training set is from 1 July 2017 to 31 January 2018 and from 22 April 2018 to 21 May 2018, the test set is from 1 February 2018 to 21 April 2018, and the validation set is from 22 May 2018 to 30 June 2018. The training set is firstly used to get the pre-training model, followed by the model performance improvement by adjusting the hyper-parameters on the validation set. Finally, the validation set is used for result evaluation.

In the application of the Deep Learning model, hyper-parameter tuning is essential to obtain the best performance. In this case study, six featured and popular models are selected for performance comparison including ANN, RNN [49], canonical LSTM, gated recurrent units (GRU) [50], stacked auto-encoders (SAEs) [51] and the bi-directional long short-term memory (Bi-LSTM) [52]. These methods have broadly covered the conventional neural network approaches as well as the state-of-the-art deep learning methods. For deep learning models, the adjustment of hyper-parameters relies heavily on repeated experiments. By observing the performance of the validation set, hyper-parameter adjustment is performed. If the number of neural network layers is too large, the model will be over-fitting, and too little will lead to under-fitting. Through repeated experiments, the ANN is assumed as a single hidden layer structure for parameter tuning simplification. The RNN, GRU, Bi-LSTM and LSTM have two hidden layers, whereas the SAEs have four hidden layers, and all of the models' hidden layers have 16 nodes. The learning rate in the deep learning model is also an important parameter. It is often set between 0.00001 and 0.01; if the learning rate is too high, the network cannot converge to the global optimal value; if the learning rate is too low, it cannot converge. Through extensive experiments, the learning rate is set to 0.001. The epoch and batch size of the training process are according to the data set size, and the network convergence can be achieved. In this experiment, we set them as 30 and 512, respectively. To prevent over-fitting of the network, we added the dropout layer, which is usually set to a range of 0.3–0.8. Through experimentation, the dropout is 0.3 to achieve the lowest prediction error. Other comparison models have achieved optimal results under this set of parameters and are discussed separately. The MAE is chosen as the loss function, while the RMS prop [53] is adopted as an optimizer. The time steps' parameter represents the second dimension of the input matrix. In the experimental study, the performance of the models in three different time steps are compared. The dimension conversion in each layer is shown in Figure 8. The dropout concept, first introduced by Hinton et al. [54], refers to the temporary discarding of neural network units from the network according to a certain probability during the training process to prevent the model from over-fitting, of which the value is set as 0.5.

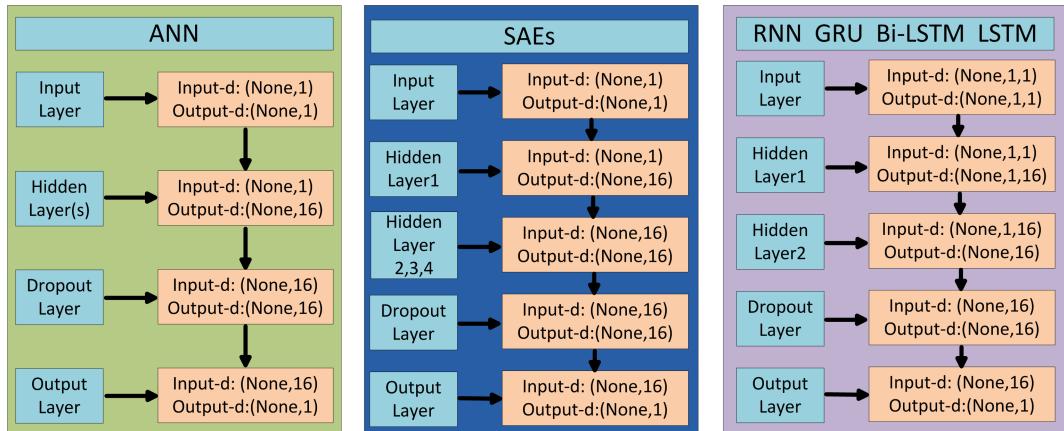


Figure 8. Dimension conversion in each layer.

5.3. Case 1: PEV Charging Station Case Study

The first step in the experimental process is to read the load data file and convert the load data into the required matrix. The data are then fed it into the model for pre-training. It is required to keep observing whether the loss function of the model on the training set converges. Once it converges, the loss of the validation set should be finally checked regarding whether the model converges or not. If the loss does not converge, an over-fitting phenomenon occurs. The hyper parameters of the model are not effective and should be adjusted. The procedure is repeated until the best performance on the validation set is obtained. The best hyper-parameters are shown in the previous subsection. The training loss and validation loss on some epochs are shown in Table 1. Multiple scenarios with three different time steps for each method are trained. As it can be seen from the table that the LSTM model has the minimum loss in the last epoch for three scenarios, which is 0.0068 (one time step), 0.0065 (five time steps), 0.0064 (15 time steps) in the training set, and 0.0031 (one time step), 0.0043 (five time steps), 0.0034 (15 time steps) in the validation set. The loss curve of three scenarios are shown in Figure 9, where the losses of each model are converged. Though the Bi-LSTM method has the comparatively well performance, the training time of the model in the training procedure is too long as the time steps increase, which need 58 s on each epoch for the 15 time steps, whereas LSTM only needs 8 s. Overall, the performance of LSTM is proved to be the best among all the counterparts.

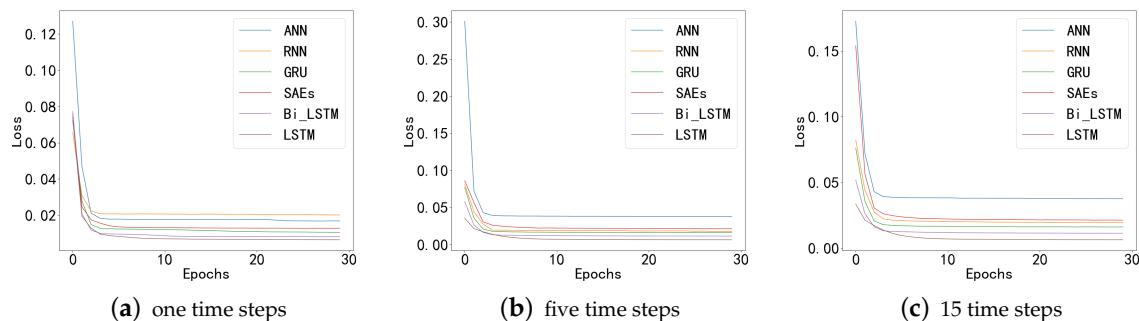


Figure 9. Loss curve comparison in three time steps.

Table 1. Training and validation loss of different algorithms in different time steps.

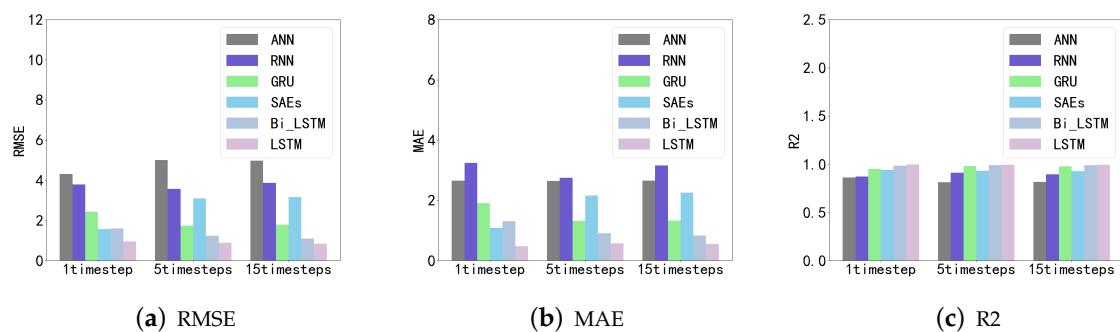
T-Step	Epoch	Loss	ANN	RNN	GRU	SAEs	Bi-LSTM	LSTM
1	1	Training Loss	0.4227	0.1007	0.1067	0.1421	0.1076	0.0746
	1	Validation Loss	0.1525	0.0253	0.0258	0.0583	0.0089	0.0136
	10	Training Loss	0.0540	0.0270	0.0193	0.0229	0.0142	0.0079
	10	Validation Loss	0.0483	0.0169	0.0098	0.0161	0.0072	0.0048
	20	Training Loss	0.0455	0.0271	0.0190	0.0212	0.0140	0.0070
	20	Validation Loss	0.0289	0.0166	0.0105	0.0116	0.0101	0.0033
	30	Training Loss	0.0399	0.0271	0.0188	0.0209	0.0133	0.0068
	30	Validation Loss	0.0153	0.0149	0.0096	0.0107	0.0068	0.0031
	1	Training Loss	0.3010	0.0810	0.0771	0.0862	0.0581	0.0356
	1	Validation Loss	0.0793	0.0240	0.0226	0.0293	0.0156	0.0253
5	10	Training Loss	0.0382	0.0195	0.0167	0.0222	0.0118	0.0074
	10	Validation Loss	0.0185	0.0100	0.0087	0.0149	0.0067	0.0077
	20	Training Loss	0.0380	0.0189	0.0163	0.0216	0.0116	0.0067
	20	Validation Loss	0.0196	0.0126	0.0066	0.0121	0.0052	0.0053
	30	Training Loss	0.0378	0.0186	0.0161	0.0214	0.0115	0.0065
	30	Validation Loss	0.0196	0.0087	0.0081	0.0118	0.0055	0.0043
	1	Training Loss	0.1726	0.0824	0.0763	0.1540	0.0519	0.0337
	1	Validation Loss	0.0360	0.0756	0.0373	0.0509	0.0156	0.0692
	10	Training Loss	0.0383	0.0195	0.0167	0.0224	0.0120	0.0073
	10	Validation Loss	0.0204	0.0120	0.0090	0.0129	0.0101	0.0072
15	20	Training Loss	0.0379	0.0197	0.0164	0.0217	0.0116	0.0067
	20	Validation Loss	0.0196	0.0167	0.0086	0.0120	0.0094	0.0084
	30	Training Loss	0.0377	0.0195	0.0162	0.0214	0.0115	0.0064
	30	Validation Loss	0.0194	0.0097	0.0075	0.0115	0.0049	0.0034

Then, the trained models are recalled to test the accuracy on the test set. The average error (MAE, RSME) and goodness of fit (R2) are calculated and shown as Table 2. In Table 2, we can find that the LSTM proves to be more accurate than the ANN, RNN, GRU, SAE, Bi-LSTM models for the minute-level load forecasting. The MAE of ANN model in three time steps is 2.3582, 3.0206 and 2.9988, and 0.4782, 0.5734 and 0.5500 for the LSTM model. In contrast, the MAE is reduced by 1.8800 in one time step, 2.4472 in five time steps, and 2.4489 in 15 time steps. It can be found that, with the time step increase, the MAE difference also increases in the two models. The average RMSE of LSTM in three scenarios is 0.8988. When the time step changes from 1 to 5, the RMSE of RNN, GRU, Bi-LSTM LSTM all decrease, and that of the ANN and SAE increase. Such results prove that the sequence models work better for long sequences. When the time step changes from 5 to 15, the RMSE of other models increase, while that of LSTM and Bi-LSTM decrease. This proves that LSTM can remember longer input sequences and efficiently process their information, providing a more robust solution to the sequences data with longer time intervals and longer delays. In addition, with the time steps increase, the R2 of ANN becomes smaller, which are 0.862306 for one time step and 0.816820 for 15 time steps. The R2 of LSTM is very close to one in three time steps, which indicates that the LSTM has been significantly competitive for the super-short-term EV charging load forecasting. The metrics and error comparison histograms in three scenarios are shown in Figure 10. It can be seen more intuitively from the histograms that LSTM has the lowest error and the best goodness of fit compared to five other methods. Compared with the best results obtained from all the counterparts, the LSTM method has an average of 30% lower errors on all the index criteria.

Table 2. Performance comparison of the MAE, RMSE, and R2 for all methods.

T-Step	Metrics	ANN	RNN	GRU	SAEs	Bi-LSTM	LSTM
1	MAE	2.3582	3.2397	1.9116	1.0886	1.3096	0.4782
	RMSE	4.3078	3.7915	2.4333	1.5689	1.5996	0.9546
	R2	0.8623	0.8716	0.9495	0.9403	0.9844	0.9953
5	MAE	3.0206	2.7457	1.3134	2.1616	0.9045	0.5734
	RMSE	5.0117	3.5703	1.7376	3.1042	1.2288	0.8937
	R2	0.8136	0.9104	0.9788	0.9323	0.9894	0.9944
15	MAE	2.9988	3.1559	1.3269	2.2516	0.8296	0.5500
	RMSE	4.9680	3.8630	1.7880	3.1556	1.0934	0.8452
	R2	0.8168	0.8941	0.9756	0.9292	0.9916	0.9950

Figure 11 shows the actual data curve and forecasting curve for each model in the three scenarios of a single day and a whole week. For the single day in the left column, 1440 points is described in the curve. It can be seen that the charging load began to increase sharply at 11:00 p.m. This is due to the fact that the public transportation in Shenzhen, such as taxis and buses, are all electric vehicles. These vehicles work during the day and can be recharged at night after being fully uncommitted by 11:00 p.m. In addition, it is the time period when the charging price is the lowest after 11:00 p.m., which is beneficial for reducing the charging cost. In the enlarged view, it could be observed that the change in per minute load is a more clear waveform. Both ANN and SAE models have slight fluctuations in the forecasting curve when the time step is 1, and the fluctuation disappears with the time step increases. Compared with the actual data curve especially near the peak, ANN and SAE models fail to capture the step changes in the load value. The LSTM model again perfectly predicts the load of each point and captures all the slight step changes.

**Figure 10.** RMSE, MAE and R2 comparison histograms.

For a whole week in the right column of Figure 11, the data from 15 February 2018 to 22 February 2018 are adopted and 10,080 points are considered in the curve. It could be easily seen that, due to the inherent working mode of public transports, the curve shows a certain periodicity. Moreover, the utilization of one minute as time intervals can effectively describe the nonlinear characteristics of the data; however, it makes it more difficult to the general model for accurate load forecasting. From the first figure, it can be observed that, when the time step is 1, the RNN model cannot fit the peak load well and the valley load, and the forecasting value is larger than the actual value. The ANN model can not capture the subtle changes of the load, in particular for the data near the peak load. In the partially enlarged graph, the forecasting effect of the SAEs model is basically a straight line near the curve initial position, and it is basically unable to capture the subtle changes. The GRU, Bi-LSTM and LSTM models can reasonably forecast the value of each point, and LSTM still shows the best forecasting accuracy among the three models. In the second figure that illustrates the curves of the five time step models, it can be seen that the RNN model forecasting effect is largely improved. However,

in the enlarged graph, the forecasting curve of SAE and ANN produces the straight line near the initial position. The LSTM model still obtains the best performance among all the counterparts. The third figure demonstrates the curves of the 15 time steps models and shows the same results with Table 2. In this scenario, the forecasting accuracy of the RNN model is reduced, where the forecasting value of peak and valley load is higher than the actual value in the one time step case and it is lower than the true data in the 15 time steps case.

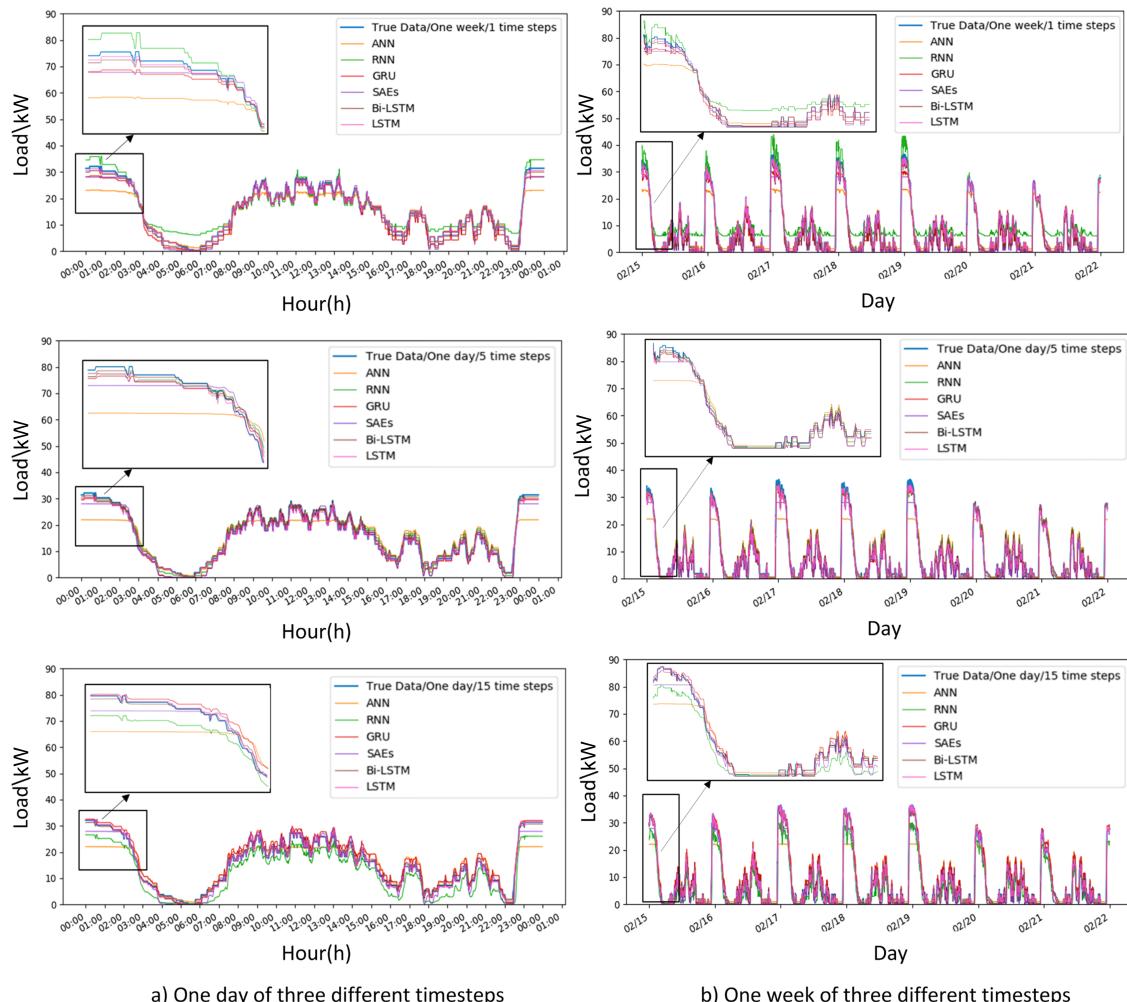


Figure 11. Different time steps load forecasting curve effect comparison graph of one week.

As shown in Figure 12, the forecast results of the holiday (dry season) and the working day (rainy season) are compared. From the forecast of the working day (rainy season), the charging load peaks at 11:00 p.m., begins to decrease at 3:00 a.m., and increases again at 8:00 a.m. In addition, the charging load is higher during working days, while the holiday (dry season) had a low charging load during the day. This is because people often choose to travel on holidays, and most private EVs are charged on working days, while most public transport EVs can only choose to charge at night. Again, due to the large amount of public transportation, the charging load increases greatly at 11:00 p.m. The prediction results of each model show that the prediction accuracy of LSTM on holidays and working days is better than the other five counterparts.

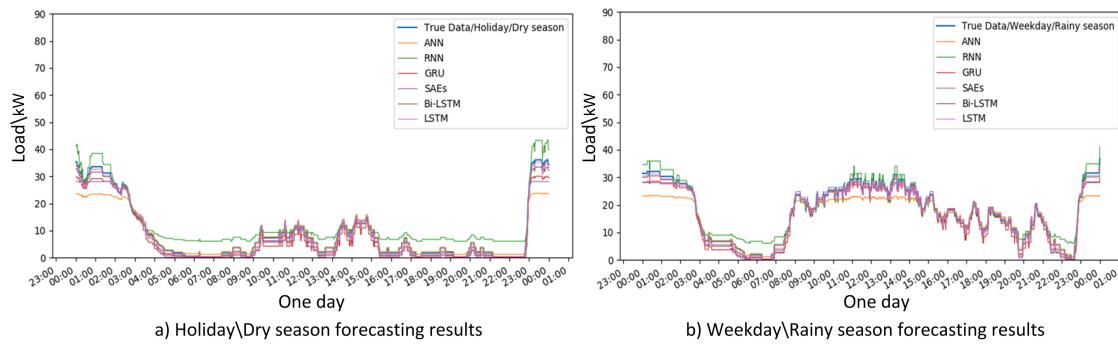


Figure 12. Load forecasting curve effect comparison graph of holiday (dry) and weekday (rainy).

In the results, the LSTM model exhibits strongly competitive performance in minute-level PEV charging load forecasting, and its variants GRU and Bi-LSTM also exhibit comparatively well accuracy. The effectiveness of the deep learning based model for minute-level super short-term load forecasting for PEV charging is promising and manifested.

5.4. Case 2: PEV Aggregator Case Study

The load curves of the PEV charging station persist highly periodical characteristics, which might not be convincing enough to demonstrate the superior performance of deep learning methods. In order to further verify the validity of the proposed model, another minute-level load dataset of a PEV aggregator for commercial building chargers in Shenzhen was used to validate the proposed super short-term model, and the prediction results' curves are shown in Figure 13. It could be observed that the charging behavior of a commercial charges aggregator is more random and fluctuating, which is completely different from the charging station profile. In this case study, all six of the methods are compared again and the results of performance comparison are shown in Table 3, also with three time-steps' options. The LSTM again shows the best performances on all the time-step tests, obtaining the MAE of the three time-steps as small as 0.3096, 0.4699 and 0.2864, respectively. Though the accuracy is slightly lower than the results of the periodical charging station case study, deep learning models again demonstrate competitive performance for the extra-short-term PEV charging load forecasting.

Table 3. Performance comparison of the MAE, the RMSE, and the R2 for all methods.

T-Step	Metrics	ANN	RNN	GRU	SAEs	Bi-LSTM	LSTM
1	MAE	0.9098	0.4751	0.4281	0.7008	0.5321	0.3096
	RMSE	1.2581	0.6890	0.6340	0.9551	0.7702	0.5095
	R2	0.8603	0.9581	0.9645	0.9195	0.9476	0.9771
5	MAE	0.8912	0.4830	0.5112	0.5529	0.6241	0.4699
	RMSE	1.2654	0.6761	0.7218	0.7638	0.8091	0.6219
	R2	0.8585	0.9596	0.9361	0.9484	0.9421	0.9658
15	MAE	0.8823	0.4659	0.6111	0.6576	0.8157	0.2864
	RMSE	1.2489	0.6506	0.8519	0.8956	1.0260	0.4418
	R2	0.8626	0.9627	0.9284	0.9293	0.9072	0.9828

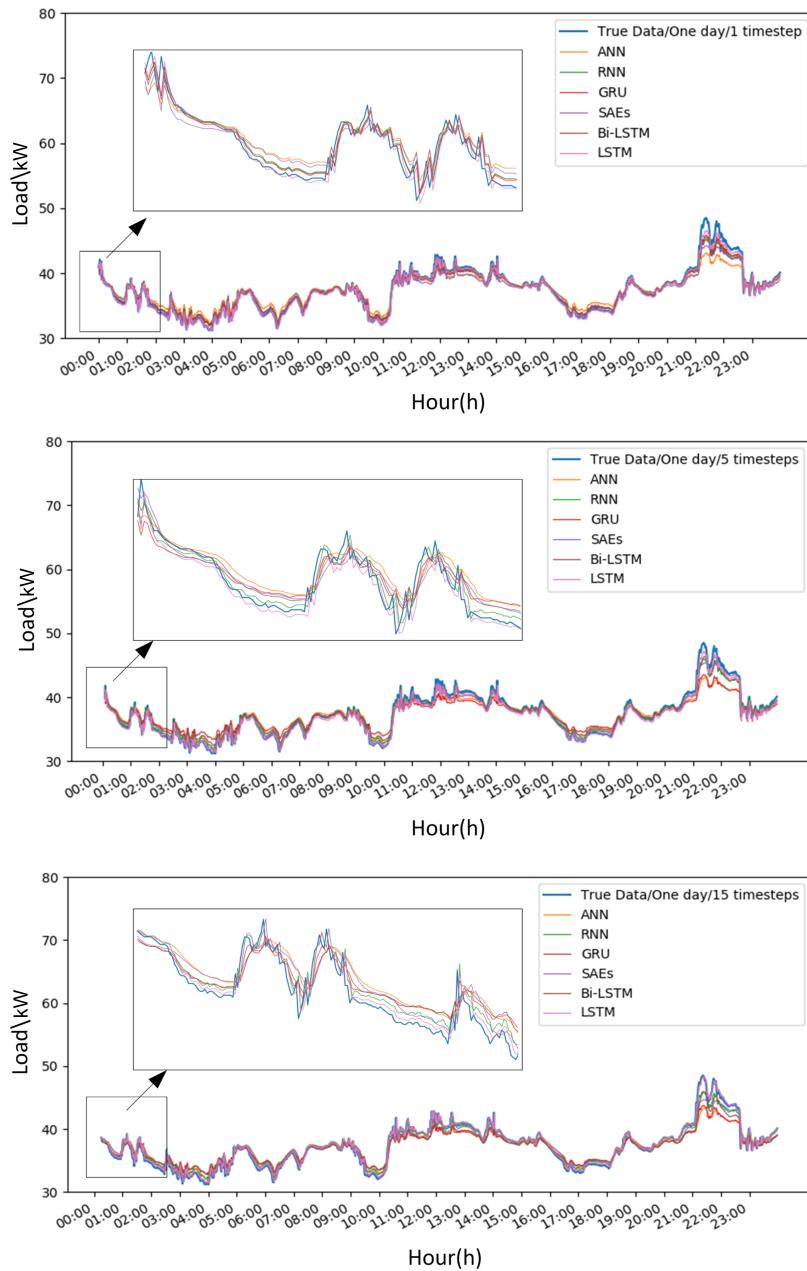


Figure 13. Load forecasting curve effect comparison graph of the PEV aggregator.

6. Conclusions and Future Work

In this study, deep learning approaches are for the first time utilized in super-short-term minute-level short-term PEV charging load forecasting. Unlike the previous shallow structure methods, the deep learning based models do not need many features in the model training. They effectively capture the potential load change features using only historical load data such as the nonlinear feature and temporal correlations. Comprehensive comparative studies including ANN, RNN, LSTM, GRU, SAEs and Bi-LSTM models are implemented in three scenarios, where the unsupervised learning algorithm is applied to pre-train the models. Fine-tuning and proper hyper-parameters are well investigated to achieve the best performance. Comprehensive metrics index including RMSE, MAE and R² are used to evaluate the model performances. The results show that deep learning models effectively forecast super-short-term PEV charging load for providing accurate prediction curves in

dynamic power system scheduling. Among the deep learning methods, the LSTM model is superior to the other methods and is competent in forecasting extra-short term PEV charging load.

With the quickly mass roll out of PEVs in the multiple levels of power grid, the accurate forecasting of PEV charging load has the potential to bring significant economic and social benefits. The proposed deep learning model provides an important tool to pave the way for the large penetration of PEVs integrating into the power system and provides a competitive artificial intelligence showcase in the low carbon energy systems.

Author Contributions: J.Z. and Z.Y. proposed the algorithms and drafted the paper. M.M., Y.G., Y.W. and S.F. prepared the data and modified the paper. Y.Z. and Y.C. run the data test and prepared the materials.

Funding: This research is financially supported by China NSFC (51607177), the Natural Science Foundation of Guangdong Province (2018A030310671, 2018A030313755), the China Post-Doctoral Science Foundation (2018M631005), the European Commission's Horizon 2020 project, 'Demand response integration technologies: unlocking the demand response potential in the distribution grid' (DRIvE, 774431), the Outstanding Young Researcher Innovation Fund of the Shenzhen Institute of Advanced Technology, the Chinese Academy of Sciences (201822), the Shenzhen Basic Research Fund under Grant No. JCYJ20160331190123578 and the Shenzhen Discipline Construction Project for Urban Computing and Data Intelligence.

Acknowledgments: The authors would like to thank Huikun Yang from Winline Co. Ltd. (Shenzhen, China) for kindly providing the charging data.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Raza, M.Q.; Khosravi, A. A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings. *Renew. Sustain. Energy Rev.* **2015**, *50*, 1352–1372. [[CrossRef](#)]
- Taylor, J.W. An evaluation of methods for very short-term load forecasting using minute-by-minute British data. *Int. J. Forecast.* **2008**, *24*, 645–658. [[CrossRef](#)]
- Yang, Z.; Li, K.; Foley, A. Computational scheduling methods for integrating plug-in electric vehicles with power systems: A review. *Renew. Sustain. Energy Rev.* **2015**, *51*, 396–416. [[CrossRef](#)]
- Zhang, C.; Yang, Z.; Li, K. Modeling of electric vehicle batteries using rbf neural networks. In Proceedings of the 2014 International Conference on Computing, Management and Telecommunications (ComManTel), Da Nang, Vietnam, 27–29 April, 2014; pp. 116–121.
- Box, G.E.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley and Sons: Hoboken, NJ, USA, 2015.
- Wei, L.; Zhen-gang, Z. Based on time sequence of arima model in the application of short-term electricity load forecasting. In Proceedings of the 2009 International Conference on Research Challenges in Computer Science, Shanghai, China, 28–29 December 2009; pp. 11–14.
- Haida, T.; Muto, S. Regression based peak load forecasting using a transformation technique. *IEEE Trans. Power Syst.* **1994**, *9*, 1788–1794. [[CrossRef](#)]
- Shankar, R.; Chatterjee, K.; Chatterjee, T. A very short-term load forecasting using kalman filter for load frequency control with economic load dispatch. *J. Eng. Sci. Technol. Rev.* **2012**, *5*, 97–103. [[CrossRef](#)]
- Park, D.C.; El-Sharkawi, M.; Marks, R.; Atlas, L.; Damborg, M. Electric load forecasting using an artificial neural network. *IEEE Trans. Power Syst.* **1991**, *6*, 442–449. [[CrossRef](#)]
- Chen, B.J.; Chang, M.W. Load forecasting using support vector machines: A study on eunite competition 2001. *IEEE Trans. Power Syst.* **2004**, *19*, 1821–1830. [[CrossRef](#)]
- Hippert, H.S.; Pedreira, C.E.; Souza, R.C. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Trans. Power Syst.* **2001**, *16*, 44–55. [[CrossRef](#)]
- Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
- Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*; NIPS: Grenada, Spain, 2012; pp. 1097–1105.

14. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
15. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems*; NIPS: Grenada, Spain, 2014; pp. 3104–3112.
16. Yang, D.; Pang, Y.; Zhou, B.; Li, K. Fault Diagnosis for Energy Internet Using Correlation Processing-Based Convolutional Neural Networks. *IEEE Trans. Syst. Man Cybern. Syst.* **2019**. [[CrossRef](#)]
17. Dommel, H.W.; Tinney, W.F. Optimal power flow solutions. *IEEE Trans. Power Appar. Syst.* **1968**, *87*, 1866–1876. [[CrossRef](#)]
18. Corpening, S.L.; Reppen, N.D.; Ringlee, R.J. Experience with weather sensitive load models for short and long-term forecasting. *IEEE Trans. Power Appar. Syst.* **1966**, *PAS-92*, 1966–1972. [[CrossRef](#)]
19. Hagan, M.T.; Behr, S.M. The time series approach to short term load forecasting. *IEEE Trans. Power Syst.* **1987**, *2*, 785–791. [[CrossRef](#)]
20. Juberias, G.; Yunta, R.; Moreno, J.G.; Mendivil, C. A new arima model for hourly load forecasting. In Proceedings of the Transmission and Distribution Conference, New Orleans, LA, USA, 11–16 April 1999.
21. Jie, W.U.; Wang, J.; Haiyan, L.U.; Dong, Y.; Xiaoxiao, L.U. Short term load forecasting technique based on the seasonal exponential adjustment method and the regression model. *Energy Convers. Manag.* **2013**, *70*, 1–9.
22. Pai, P.F.; Hong, W.C. Support vector machines with simulated annealing algorithms in electricity load forecasting. *Energy Convers. Manag.* **2005**, *46*, 2669–2688. [[CrossRef](#)]
23. Guo, Y.; Nazarian, E.; Ko, J.; Rajurkar, K. Hourly cooling load forecasting using time-indexed arx models with two-stage weighted least squares regression. *Energy Convers. Manag.* **2014**, *80*, 46–53. [[CrossRef](#)]
24. Lahouar, A.; Slama, J.B.H. Day-ahead load forecast using random forest and expert input selection. *Energy Convers. Manag.* **2015**, *103*, 1040–1051. [[CrossRef](#)]
25. Feng, Y.; Xiaozhong, X. A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved bp neural network. *Appl. Energy* **2014**, *134*, 102–113.
26. Kouhi, S.; Keynia, F. A new cascade nn based method to shortterm load forecast in deregulated electricity market. *Energy Convers. Manag.* **2013**, *71*, 76–83. [[CrossRef](#)]
27. Mahmoud, T.S.; Habibi, D.; Hassan, M.Y.; Bass, O. Modelling self-optimised short term load forecasting for medium voltage loads using tunning fuzzy systems and artificial neural networks. *Energy Convers. Manag.* **2015**, *106*, 1396–1408. [[CrossRef](#)]
28. Hinton, G.E.; Osindero, S.; Teh, Y.W. A fast learning algorithm for deep belief nets. *Neural Comput.* **2006**, *18*, 1527–1554. [[CrossRef](#)]
29. Salakhutdinov, R.; Hinton, G. Deep boltzmann machines. In *Artificial Intelligence and Statistics*; Addison-Wesley: New York, NY, USA, 2009; pp. 448–455.
30. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.A. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.
31. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
32. Marino, D.L.; Amarasinghe, K.; Manic, M. Building energy load forecasting using deep neural networks. In Proceedings of the IECON 2016-42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 23–26 October, 2016; pp. 7046–7051.
33. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-term residential load forecasting based on lstm recurrent neural network. *IEEE Trans. Smart Grid* **2017**, *10*, 841–851. [[CrossRef](#)]
34. Zheng, H.; Yuan, J.; Chen, L. Short-term load forecasting using emd-lstm neural networks with a xgboost algorithm for feature importance evaluation. *Energies* **2017**, *10*, 1168. [[CrossRef](#)]
35. Bouktif, S.; Fiaz, A.; Ouni, A.; Serhani, M. Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies* **2018**, *11*, 1636. [[CrossRef](#)]
36. Aziz, M.; Oda, T.; Mitani, T.; Watanabe, Y.; Kashiwagi, T. Utilization of electric vehicles and their used batteries for peak-load shifting. *Energies* **2015**, *8*, 3720–3738. [[CrossRef](#)]
37. Gerossier, A.; Girard, R.; Kariniotakis, G. Modeling and Forecasting Electric Vehicle Consumption Profiles. *Energies* **2019**, *12*, 1341. [[CrossRef](#)]

38. Mu, Y.; Wu, J.; Jenkins, N.; Jia, H.; Wang, C. A spatialtemporal model for grid impact analysis of plug-in electric vehicles. *Appl. Energy* **2014**, *114*, 456–465. [CrossRef]
39. Qian, K.; Zhou, C.; Allan, M.; Yue, Y. Modeling of load demand due to ev battery charging in distribution systems. *IEEE Trans. Power Syst.* **2011**, *26*, 802–810. [CrossRef]
40. Alizadeh, M.; Scaglione, A.; Davies, J.; Kurani, K.S. A scalable stochastic model for the electricity demand of electric and plugin hybrid vehicles. *IEEE Trans. Smart Grid* **2014**, *5*, 848–860. [CrossRef]
41. Luo, Z.; Song, Y.; Hu, Z.; Xu, Z.; Xia, Y.; Zhan, K. Forecasting charging load of plug-in electric vehicles in china. In Proceedings of the Power and Energy Society General Meeting, San Diego, CA, USA, 24–29 July 2011.
42. Lu, Y.; Li, Y.; Xie, D.; Wei, E.; Bao, X.; Chen, H.; Zhong, X. The Application of Improved Random Forest Algorithm on the Prediction of Electric Vehicle Charging Load. *Energies* **2018**, *11*, 3207. [CrossRef]
43. Zhu, J.; Yang, Z.; Guo, Y.; Zhang, J.; Yang, H. Short-Term Load Forecasting for Electric Vehicle Charging Stations Based on Deep Learning Approaches. *Appl. Sci.* **2019**, *9*, 1723. [CrossRef]
44. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [CrossRef] [PubMed]
45. Werbos, P.J. Backpropagation through time: What it does and how to do it. *Proc. IEEE* **1990**, *78*, 1550–1560. [CrossRef]
46. Chollet, F. Keras. Available online: <https://keras.io/> (accessed on 18 November 2018).
47. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. *Tensorflow: A System for Large-Scale Machine Learning*; OSDI: Boulder, CO, USA, 2016; Volume 16, pp. 265–283.
48. Wei, Y.; Zhang, X.; Shi, Y.; Xia, L.; Pan, S.; Wu, J.; Han, M.; Zhao, X. A review of data-driven approaches for prediction and classification of building energy consumption. *Renew. Sustain. Energy Rev.* **2018**, *82*, 1027–1047. [CrossRef]
49. Vermaak, J.; Botha, E. Recurrent neural networks for shortterm load forecasting. *IEEE Trans. Power Syst.* **1998**, *13*, 126–132. [CrossRef]
50. Cho, K.; Van Merriënboer, B.; Bahdanau, D.; Bengio, Y. On the properties of neural machine translation: Encoder-decoder approaches. *arXiv Preprint* **2014**, arXiv:1409.1259.
51. Bengio, Y.; Lamblin, P.; Popovici, D.; Larochelle, H. Greedy layer-wise training of deep networks. In *Advances in Neural Information Processing Systems*; NIPS: Grenada, Spain, 2007; pp. 153–160.
52. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural Networks* **2005**, *18*, 602–610. [CrossRef]
53. Tieleman, T.; Hinton, G. Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA Neural Netw. Mach. Learn.* **2012**, *4*, 26–31.
54. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv Preprint* **2012**, arXiv:1207.0580.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).