Short-term Forecasting for Utilization Rates of Electric Vehicle Charging Stations

Zuzhao Ye, Ran Wei and Nanpeng Yu

Department of Electrical and Computer Engineering
University of California, Riverside
Riverside, California 92521 USA
zye066@ucr.edu, ran.wei@ucr.edu, nyu@ece.ucr.edu

Abstract—Accurate forecasts for the utilization rates of electric vehicle charging stations (CSs) are crucial to coordinating the operations of on-site distributed energy resources. In this paper, we propose to forecast the CS utilization rates by considering key explanatory variables such as historical utilization rates, traffic flows, demographic properties, the number of EV registrations, and points of interest. Three machine learning models, namely random forest, feed-forward neural network, and long short-term memory (LSTM) are adopted for the forecasting task. The proposed algorithms are validated using the real-world utilization data collected from around 130 CSs in Contra Costa County, California. The numerical study results show that the LSTM model achieves the best prediction performance. The lagged CS utilization rates and traffic flows are the two most influential features. More interestingly, the traffic flow plays a more important role in predicting the utilization rates of DC Fast CSs than that of the level 1 (L1) and level 2 (L2) CSs.

Index Terms—Data-driven forecast, utilization rate, electric vehicle, charging station.

I. INTRODUCTION

The adoption of electric vehicles (EVs) has accelerated tremendously in the past decade. To support the continued growth of the EV market, a network of intelligent EV charging stations is in critical need. Accurate short-term forecast for charging station (CS) utilization rate is essential to peak power reduction and management of on-site distributed energy resources for intelligent CSs [1].

An early attempt to forecast the charging demand of Plug-in Hybrid EVs is carried out by [2], which models the charging demand of multiple vehicles based on queuing theories. The city-wide hourly EV charging demand is predicted based on the travel patterns and assumptions for initial charging time [3]. [4] predicts station-level hourly demand with the assistance of surveillance video in local traffic networks. A macro-scale charging demand model can be developed by assembling micro-scale models of EV driver behaviors, which can be derived from the American National Household Travel Survey (NHTS) dataset [5], [6]. The customer preferences of charging time can also be inferred from survey [7], which could be integrated into other models to determine the temporal patterns of charging demand.

The aforementioned works predict average hourly charging demands, which are important for siting of CSs. To facilitate

This research is funded by the University of California Institute of Transportation Studies under the project of UCITS-2021-47.

more efficient operations of CSs, we also need day-ahead and hour-ahead charging demand forecast at the CS or even customer level. Based on the charging data of South Korea in 2018 and 2019, [8] forecasts day-ahead, week-ahead, and month-ahead charging demands in three different scales: national, city, and station. [9] uses deep learning method to forecast traffic flow and 1-hour ahead charging demand of a CS. With complete charging session data of Nebraska, USA from 2013 to 2019 and unique customer ID, [10] predicts how much energy an EV will charge given the plug-in time, season and cost of electricity.

The existing studies discovered that the charging demand of CSs heavily depends on fast-varying factors such as traffic flows and slow-varying factors such as land characteristics [11] and socio-demographic properties [12], [13] of their neighborhood area.

However, most of the existing work for CS demand forecasts are performed either without considering the impact of real-time traffic flows or with simplified traffic flow estimates from coarse-grained models. Most of the existing time series-based CS demand forecast work do not incorporate slow-varying factors such as land-use and socio-demographic properties into the prediction model. This is acceptable if a large amount of historical data has been collected for the CS. However, for a newly installed CS, the lack of historical data will make it extremely difficult to build a time series-based model with satisfactory performance.

In this study, we propose a machine learning-based framework for short-term CS demand forecast by considering both fast-varying and slow-varying explanatory variables. This unique approach allows us to provide accurate hour-ahead charging demand forecast for a CS with limited historical data.

The contribution of this paper is summarized below.

1) We develop three machine learning models to perform hour-ahead charging demand forecast for CSs by using both fast-varying and slow-varying explanatory variables, which yields accurate prediction with limited historical data.

2) Our proposed models achieve great performance on real-world noisy charging data, dynamic traffic flow data, and socio-economic data. The numerical study results of the relative importance of input features reveal that EV traffic flows have a greater influence on the charging demand at DC Fast CSs than at L1/L2 CSs.

978-1-6654-4919-9/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: Univ of Calif Riverside. Downloaded on February 13,2022 at 22:02:35 UTC from IEEE Xplore. Restrictions apply.

2021 IEEE International Smart Cities Conference (ISC2) | 978-1-6654-4919-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISC253183.2021.9562948

The rest of this paper is structured as follows. Section II introduces the adopted machine learning models for charging demand forecast. Section III provides the source, collection method, and data preprocessing techniques for the real-world datasets. Section IV presents the numerical study to verify the performance of the proposed prediction models. Limitations and potential future work are provided in Section V. The conclusions are stated in Section VI.

II. MACHINE LEARNING MODELS FOR CHARGING DEMAND FORECAST

A. Machine Learning Models

To forecast the utilization rate of CSs, we adopt three different machine learning algorithms: random forest, feed forward neural network (FNN), and long short-term memory (LSTM). They are widely used prediction models and have been successfully demonstrated in many other domains.

We briefly introduce them in this subsection.

1) Random Forest. Originated from decision tree model, a random forest consists of a set of decision trees, which are constructed randomly and independently. Each of the decision trees will output a prediction, and the output of the random forest is the average of the outputs from individual decision trees.

2) FNN. An FNN consists of one input layer, one or multiple hidden layers, and one output layer. Each layer contains a certain number of nodes. Starting from the first hidden layer, each node receives weighted inputs from all of the nodes in the previous layer, and the sum of these inputs will be fed into a non-linear activation function before serving as the input of the next layer. The weights between any of the adjacent layers will be updated through back-propagation of gradients taken from minimizing a well designed loss function to match the goal of learning.

3) LSTM. LSTM is a recurrent neural network that is designed for learning time series data. An LSTM cell consists of an input gate, an output gate, and a forget gate. Each of the gates is a parameterized linear function followed by a non-linear activation function. The forget gate controls what information from the previous calculation should be kept for current calculation. The schematics of an unfolded LSTM is shown in Fig. 1. The input x is fed into the LSTM model by following the time-series order, i.e., the earliest input x_{t-n} is the first being processed, and the output hidden states h_{t-n} and cell state c_{t-n} are serving as the input to the next time step. The final output is obtained from a fully connected layer with inputs from the last hidden state. Similar to FNN, the parameters in an LSTM model are updated through the gradient backpropagation.

The hyper-parameters of the aforementioned models are summarized in Table I. The final reported performance is obtained for the best combination of hyper-parameters of each model on the validation dataset.

TABLE I: Summary of Model Configurations

Model	Model Configuration
Random Forest	Number of trees $\in \{500, 1000, 5000\}$;
FNN*	Two hidden layers with [128, 128] or [256, 256] neurons in each layers.
LSTM*	Two LSTM layers with [50, 100, 200] features in the hidden state.

* For FNN and LSTM, the following parameter combinations are tested: Learning rate $\in \{10^{-3}, 10^{-4}, 10^{-5}\}$. Optimizer $\in \{SGD, Adam\}$. Batchsize $\in \{32, 64, 128\}$.
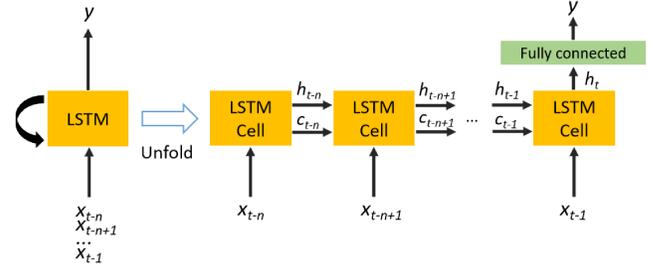

Fig. 1: Unfolded computational graph of an LSTM.

B. Interpret Machine Learning Models with the SHAP Method

While the adopted machine learning models are capable of making accurate predictions, it is difficult to directly interpret the trained models. We use the recently developed SHAP framework [14], [15] to evaluate and quantify the feature importance in the trained models.

The working principal of the SHAP method can be explained as follows: Suppose we have a trained model f that predicts y from $y = f(x)$, where $x \in X$ and $x = [x_1, x_2, ..., x_K]^T$ contains K input features. For the ith input $x^{(i)}$ sampled from X, the SHAP framework calculates the contribution vector of input features $\phi^{(i)} = [\phi_1^{(i)}, \phi_2^{(i)}, ..., \phi_K^{(i)}]$ such that $f(x^{(i)}) = \sum_{k=1}^{K} \phi_k^{(i)} + E_{x \in X}(f(x))$. The kth element in contribution vector ϕ, i.e., $\phi_k^{(i)}$, is also referred to the SHAP value of feature k for sample $x^{(i)}$.

We define the relative importance (RI) of a feature k to be:

$$\text{RI}_k = \frac{|\phi_k|}{\sum_{k \in K} |\phi_k|} \tag{1}$$

The SHAP method can be applied to various machine learning models as it is model agnostic. In this study, we will calculate the SHAP values of different input features and compare their relative importance.

III. DATA SOURCES AND PREPROCESSING STEPS

In this section, we introduce the datasets by data categories, including the data collection and preprocessing procedures.
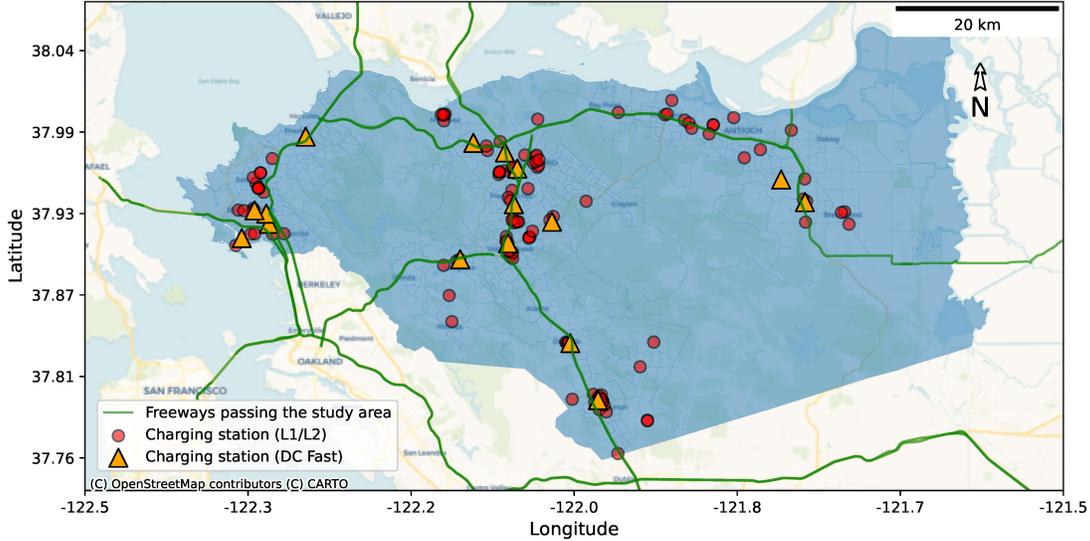
Fig. 2: The location of charging stations and freeways within the study area (Contra Costa, CA).
Source of the base map: OpenStreetMap.

A. Charging Stations

We obtain the locations of public CSs within the boundary of Contra Costa County (California, USA) from Google Map, which provides station specific information such as CS operator, number and power level of chargers, and real-time charging status of the individual chargers.

1) Utilization Rate: The number of chargers in use can represent the demand of a CS to a certain degree. However, the individual chargers at a charging station may have different power specifications. Thus, we measure the charging demand of a CS as the ratio of its actual energy usage to its maximum possible energy usage. Specifically, we formulate our dependent variable, utilization rate (UR) as follows:

$$\text{UR}(t) = \frac{\sum_{i \in I} n_i(t) p_i}{\sum_{i \in I} N_i p_i}, \qquad (2)$$

where I is the set of all types of chargers, N_i is the total number of type i chargers with power rating p_i, $n_i(t)$ is the number of occupied type i chargers at time t. $\sum_{i \in I} N_i p_i$ represents the maximum energy usage of the charging station, while $\sum_{i \in I} n_i(t) p_i$ represents the actual charging energy usage, assuming all the chargers are running at their nominal powers. The range of UR is $[0, 1]$. We collect real-time charger status every 5 minutes for DC Fast CSs and every 15 minutes for L1/L2 CSs. The hourly UR is estimated as the average of UR(t) of multiple readings during the hour.

2) Charging Station Properties: Based on the power specification of chargers as shown in Table II, we divide the CSs into two types: L1/L2 and DC Fast. As the power ratings are drastically different between these two types of CSs, we expect the utilization patterns to be different as well.

Since some CSs have both types of chargers, we treat such a CS as if it were two distinct CSs, corresponding to their charger types. The geographical location of L1/L2 and DC Fast CSs are shown in Fig. 2. The CS type is modeled with one-hot encoding, where 1 represents DC Fast CS and 0 represents L1/L2 CS.

In addition to power specifications, the operator of a CS could also have potential impacts on the quality of service. There are five CS operators involved in our study area, and they are represented by an one-hot encoded 5×1 vector.

TABLE II: Charger Types in the Study Area

Type	Charger Specifications
L1/L2	Wall outlet: 1.92 kW J-1772: 3.6 / 5.0 / 6.2 / 6.7 / 7.2 / 7.7 / 8.3 / 9.6 kW
DC Fast	CHAdeMO: 40.5 / 54.0 / 62.5 / 100 kW CSS: 28.5 / 29.8 / 50 / 62.5 / 100 kW

B. Traffic Flow (TF)

Previous studies have shown that utilization rates of CSs are highly correlated with traffic flows [2], [4]. We obtain traffic flow data from Performance Measurement System (PeMS) [16]. PeMS provides hourly traffic flow of major freeways in California. The freeways in our study area are shown in Fig. 2. Note that most of the CSs are located along the freeways.

In order to estimate the amount of traffic flow that can be potentially served by a CS, we first identify the closest freeway to the CS. Then we collect the amount of traffic flow from the

6 closest vehicle detector stations on this freeway, as shown in Fig. 3. Finally, we use the average traffic flow recorded by these 6 vehicle detector stations as one of the independent variables.

C. Service Area of Charging Stations

The geographic area that a CS can serve has a great impact on its utilization rate [17], [18]. Usually, drivers will be waiting at home or office while their EVs are charging at an L1/L2 station, as the charging time can be very long. In this case, the CS should be within reasonable walking distance. On the other hand, for the DC Fast CSs, the relatively short charging time makes it feasible for drivers to wait around the stations. Thus, driving to a DC Fast CS that is outside of reasonable walking distance is acceptable. A survey conducted by the Contra Costa Transportation Authority (CCTA) confirms this assumption and shows that people can tolerate a 0.5-mile walking distance and a 2-mile driving distance to access an L1/L2 station and a DC Fast station, respectively. We use these two distances to construct the service areas for the L1/L2 and DC fast CSs. Fig. 3 shows the service area of an L1/L2 station. As introduced later this section, the service area determines a set of features of the forecasting model.

D. Demographic and Socioeconomic Factors

We also include a variety of demographic and socioeconomic characteristics to forecast charging demand [13], [17]. The demographic and socioeconomic factors, including population, income, education, employment, percentage of work travel by private cars, and number of multifamily units, are available from National Historical Geographic Information System (NHGIS) [19] at the granularity level of census block groups (CBGs).

We assume that a CBG could be served by a CS if it geographically intersects with the service area of the CS. The total population, employment and multifamily units served by a CS is derived by summing up these factors for all CBGs served by the CS. The income, education level, and percentage of work travel by private cars are obtained from averaging the corresponding factors over the served CBGs.

E. EV Registration

The number of EV registrations in a CBG is expected to greatly influence the charging demand. We obtain the EV registration data from California Department of Motor Vehicles at the level of ZIP code tabulation area (ZCTA). To make EV registration data consistent with other demographic variables, we interpolate the number of EV registrations at each CBG using population weighting, as follows:

$$n_{EV}^c = \frac{pop^c}{pop^z} n_{EV}^z, \tag{3}$$

where n_{EV}^c and n_{EV}^z are the number of registered EVs in a CBG and the ZCTA that the CBG belongs to. pop^c and pop^z are the population in the CBG and ZCTA. Note that only CBGs whose median household income is higher than $ 48,500 [20]
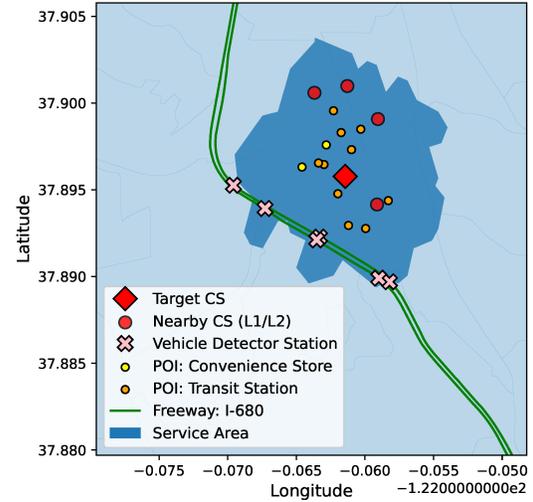

Fig. 3: An example service area covered by an L1/L2 CS, along with other nearby CSs, vehicle detector stations, and POIs

are considered in the interpolation process due to the relatively high price of EVs.

F. Point of Interest

Existing literature also found that charging demand is significantly impacted by certain types of point of interest (POI) such as transport, retail, and commercial [11]. Thus, we include the number of POIs in the service area of a CS as another feature of the prediction model, as shown in Fig. 3. In particular, we count the number of transit stations and convenience stores based on the raw data gathered from Google Places API.

G. Nearby Charging Stations

The utilization rate of a CS will be influenced by its nearby CSs. Thus, we include the number of CSs in the same service area as the target CS as an input feature of the prediction model as shown in Fig. 3.

H. Summary of Data

There are a total of 139 CSs in the study area, which are managed by 5 different operators. Among these stations, 123 CSs only have L1/L2 chargers and 12 CSs only have DC Fast chargers. The remaining 4 CSs have mixed L1/L2 and DC Fast chargers. As stated in Subsection A, we split each of them into two different CSs. As a result, we have 127 L1/L2 CSs and 16 DC Fast CSs. We collect and derive 48 days of hourly CS usage data from Dec 28, 2020 to Feb 14, 2021. Note that a few CSs have constant utilization rates in a long period of time (e.g. more than 24 hours), possibly due to equipment break-down or maintenance. Data points with such properties are eliminated and they account for less than 5%

of the overall dataset. Finally, the remaining data is split into training, validation, and testing sets with 30 days, 9 days, and 9 days of data respectively.

The input features of the machine learning model is summarized in Table III.

TABLE III: Summary of Model Input Features

Category	Feature Name	Short Name
Lagged UR	Lagged utilization rates of the past x hour[a]	UR-xh
Lagged TF	Lagged traffic flow of the past x hour	TF-xh
Temporal	Hour of the day	Hour[b]
	Weekday or weekends	Weekday[c]
Nearby CS	No. of nearby DC Fast CSs	NDC
	No. of nearby L1/L2 CSs	NLC
Station	DC Fast or L1/L2	Fast[c]
	Business Operator	Operator[c]
Transportation	No. of registered EV	NEV
	% of population travel to work by private cars	Pct.WTP
	Residential vehicle miles traveled	RVMT
	Work vehicle miles traveled	WVMT
Demographic	Population served	Population
	Median household income	Income
	% of population with bachelors or higher	Education
	No. of Employments	Employment
	No. multifamily units	NML
POI	No. of nearby transit stations	NTST
	No. of nearby convenience stores	NCVN

[a] Here we choose $x \in [1, 2, 3]$.
[b] Hour h is converted to $sin(h)$ and $cos(h)$ to reflect periodicity.
[c] One-hot encoded variables.

IV. CASE STUDY RESULTS

In this section, we compare the CS utilization rate prediction performance of the adopted machine learning models and analyze the relative importance of various input features.

A. Forecasting Performance of Machine Learning Models

We train the three adopted machine learning models and select the set of hyper-parameters for each model based on the results on the validation dataset. We also include a multivariate linear regression as the baseline method.

The performance of a forecasting model is quantified by the root mean square error:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i \in \text{test dataset}} (\text{UR}_{i,\text{predicted}} - \text{UR}_{i,\text{actual}})^2}, \quad (4)$$

where N is the number of data samples in the testing dataset.

The impact of using different feature sets on the performance of machine learning models is shown in Table IV. All of the models have significant improvements in performance when traffic flow and other explanatory variables are included as additional input features. The LSTM model achieves the best performance with all of the input features. Note that the RMSE is relatively low across all models. This is because a large portion of CSs have very low utilization rates during the period with stay-at-home orders. Thus, we also compare the model performances for daily peak hours and non-peak hours in Table V. Clearly, the three adopted machine learning models all have much better peak hour predictions compared to the benchmark. Random forest is slightly better than FNN and LSTM is the best in peak UR prediction.

TABLE IV: RMSE of 4 Models with Different Combinations of Input Features

	Model			
Input Features	LR[a]	RF[b]	FNN	LSTM
Lagged UR only	0.1571	0.1330	0.1308	0.1319
UR + TF	0.1453	0.1302	0.1285	0.1296
All features	0.1390	0.1258	0.1260	0.1211

[a] Linear regression.
[b] Random forest.

TABLE V: RMSE of 4 Models in Peak and Non-peak Hours[a] (All Features are Included)

	Model			
Study Period	LR	RF	FNN	LSTM
Peak hours	0.3301	0.2574	0.2589	0.2377
Non-peak hours	0.1305	0.1202	0.1199	0.1161

[a] For each day in the 9 day testing period, we identify the peak hour as the one with the highest UR for each CS. The remaining hours are identified as non-peak hours.

We also quantify the performance of machine learning models for different types of CSs. The hour-ahead CS utilization rate forecasting performance are shown in Table VI. The utilization rate prediction is more accurate for L1/L2 stations. Fig. 4 visualizes the utilization prediction results for three stations. CS 1 and 2 have L1/L2 chargers, and CS 3 have DC Fast chargers. The predicted utilization rates have a much better match with the actual utilization for L1/L2 CS compared with the DC Fast CS.
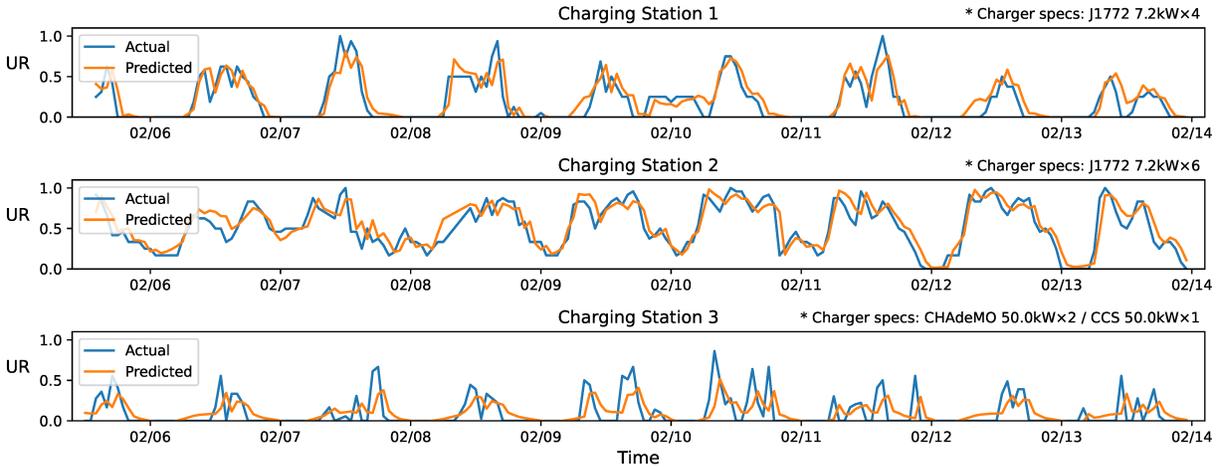
Fig. 4: Utilization rate forecast (LSTM) on three CSs. CS 1 and 2 have L1/L2 chargers and CS 3 has DC Fast chargers.

TABLE VI: RMSE of 4 Models for Different CS Types
(All Features are Included)

CS Type	Model			
	LR	RF	FNN	LSTM
L1/L2	0.1381	0.1257	0.1261	0.1201
DC Fast	0.1454	0.1281	0.1263	0.1267

B. Relative Importance of Input Features

Beyond the utilization prediction accuracy, we also want to understand how much each feature contributes to the outputs. To compare the relative importance of features, we calculate the SHAP values for each input feature. We then compare the absolute magnitude of the SHAP values to the ratio of the sum as an indicator of the relative importance of the input features. Here we choose the random forest model to quantify the input feature importance.

1) Relative Importance of Feature Categories: We first analyze the contribution of all feature categories. As L1/L2 and DC Fast CSs have different characteristics, we separately compare the relative importance of feature categories for them. As shown in Fig. 5, lagged utilization rate of the CS has the largest impact on UR predictions for both L1/L2 and DC Fast CSs. Lagged traffic flow is the second most important feature. Specifically, the relative importance of lagged UR of L1/L2 CSs (57%) is slightly higher than that of the DC Fast CSs (41%). On the other hand, the relative importance of lagged TF for L1/L2 CSs (14%) is slightly lower than that of the DC Fast CSs (19%). This phenomena can be explained by different charging behaviors of EVs at L1/L2 CSs and DC Fast CSs. In an L1/L2 CS, the EVs being charged in previous hours are very likely to extend their dwelling to the next hour because of the slow charging rates. The likelihood of an EV staying in a DC Fast CS for two consecutive hours is lower. The incoming EV traffic play a more important role in predicting the utilization
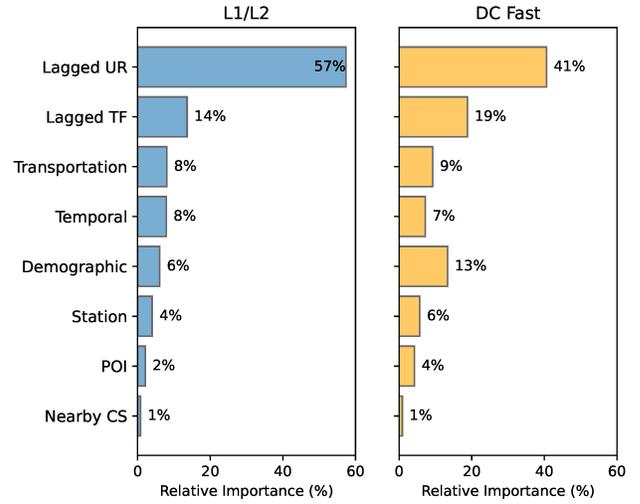

Fig. 5: Relative importance of different feature categories

rate of DC Fast CSs than that of L1/L2 CSs.

2) Relative Importance of Individual Features: The top ten individual features for predicting the utilization rate of CSs are reported for L1/L2 and DC Fast CSs separately in Table VII. As shown in the table, 1-hour lagged UR and TF are the most important individual features, which have stronger influence than 2-hour and 3-hour lagged variables. The other key features include Hour, Operator, and Population. Table VII reveals that work related features also contribute significantly to the CS utilization rates. For example, Pct. WTP and WVMT features rank 7th and 10th for L1/L2 CSs. Employment and Pct. WTP rank 3rd and 10th for DC Fast CSs. This finding could be useful to guide the siting of future CSs.

V. LIMITATIONS AND FUTURE WORK

This study has several limitations. First, the number of charging stations and the duration of data collection (48 days) are limited. Second, only sensor data along the freeway are

TABLE VII: Top Ten Features for the UR Prediction of L1/L2 and DC Fast CSs

Rank	L1/L2	R.I.[a]	DC Fast	R.I.
1	UR-1h	47.6%	UR-1h	33.3%
2	TF-1h	8.9%	TF-1h	13.3%
3	Hour	7.0%	Employment	6.8%
4	UR-2h	6.8%	Hour	5.5%
5	Operator	3.5%	UR-2h	4.6%
6	UR-3h	3.0%	Population	4.1%
7	Pct. WTP	2.9%	Operator	3.3%
8	TF-3h	2.5%	TF-3h	3.1%
9	TF-2h	2.2%	UR-3h	2.7%
10	WVMT	1.3%	Pct. WTP	2.4%
Sum	-	85.7%	-	79.1%

[a] R.I.- Relative importance

used to estimate traffic flows. In practice, traffic data for arterial roads could also serve as useful input features. Third, the data collection period overlaps with the local lockdown order due to the Covid-19 outbreak. The charging behavior of EVs during the study period could be significantly different from the pre-Covid period. In the future, we plan to perform another round of data collection with an expanded study area and longer duration when the pandemic is under control. To enhance the input features related to traffic flows, we plan to leverage a network model that can map freeway-level traffic flow to the arterial-level. The prediction model can also serve as the basis of a recommendation system, which matches EV drivers with the best CS by considering both travel and waiting cost.

VI. CONCLUSION

This paper develops short-term prediction models to estimate the utilization rates of charging stations. The proposed machine learning models leverage both slow-varying features (e.g., land-use and socio-demographic properties) and fast-varying features (traffic flow). The proposed machine learning models are trained based on real-time charging session data from around 130 charging stations in Contra Costa County, California, traffic data, and socio-economic data from the nearby census block groups. The numerical study results reveal that the LSTM model yields much lower prediction error than that of random forest, FNN, and linear regression models. The analysis of feature importance based on SHAP framework shows that the lagged utilization rates and the traffic flows are the two most important features for the model outputs. The trained machine learning models can be used to forecast real-time charging demand at both existing and planned charging stations. The charging station utilization prediction model serves as a critical tool in supporting the strategic planning and design of EV charging infrastructure.

REFERENCES

[1] J. Shi, Y. Gao, W. Wang, N. Yu, and P. A. Ioannou, "Operating electric vehicle fleet for ride-hailing services with reinforcement learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 11, pp. 4822–4834, 2019.

[2] G. Li and X.-P. Zhang, "Modeling of plug-in hybrid electric vehicle charging demand in probabilistic power flow calculations," *IEEE Transactions on Smart Grid*, vol. 3, no. 1, pp. 492–499, 2012.

[3] M. B. Arias and S. Bae, "Electric vehicle charging demand forecasting model based on big data technologies," *Applied energy*, vol. 183, pp. 327–339, 2016.

[4] M. B. Arias, M. Kim, and S. Bae, "Prediction of electric vehicle charging-power demand in realistic urban traffic networks," *Applied energy*, vol. 195, pp. 738–753, 2017.

[5] H. Lin, K. Fu, Y. Wang, Q. Sun, H. Li, Y. Hu, B. Sun, and R. Wennersten, "Characteristics of electric vehicle charging demand at multiple types of location-application of an agent-based trip chain model," *Energy*, vol. 188, p. 116122, 2019.

[6] Y. Xia, B. Hu, K. Xie, J. Tang, and H.-M. Tai, "An EV charging demand model for the distribution system using traffic property," *IEEE Access*, vol. 7, pp. 28089–28099, 2019.

[7] H. Moon, S. Y. Park, C. Jeong, and J. Lee, "Forecasting electricity demand of electric vehicles by analyzing consumers' charging patterns," *Transportation Research Part D: Transport and Environment*, vol. 62, pp. 64–79, 2018.

[8] Y. Kim and S. Kim, "Forecasting charging demand of electric vehicles using time-series models," *Energies*, vol. 14, no. 5, p. 1487, 2021.

[9] X. Zhang, K. W. Chan, H. Li, H. Wang, J. Qiu, and G. Wang, "Deep-learning-based probabilistic forecasting of electric vehicle charging load with a novel queuing model," *IEEE transactions on cybernetics*, pp. 1–14, 2020.

[10] A. Almaghrebi, F. Aljuheshi, M. Rafaie, K. James, and M. Alahmad, "Data-driven charging demand prediction at public charging stations using supervised machine learning regression methods," *Energies*, vol. 13, no. 16, p. 4231, 2020.

[11] G. Dong, J. Ma, R. Wei, and J. Haycox, "Electric vehicle charging point placement optimisation by exploiting spatial statistics and maximal coverage location models," *Transportation Research Part D: Transport and Environment*, vol. 67, pp. 77–88, 2019.

[12] X. Xi, R. Sioshansi, and V. Marano, "Simulation–optimization model for location of a public electric vehicle charging infrastructure," *Transportation Research Part D: Transport and Environment*, vol. 22, pp. 60–69, 2013.

[13] S. Y. He, Y.-H. Kuo, and D. Wu, "Incorporating institutional and spatial factors in the selection of the optimal locations of public electric vehicle charging facilities: A case study of Beijing, China," *Transportation Research Part C: Emerging Technologies*, vol. 67, pp. 131–148, 2016.

[14] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017.

[15] W. Wang, N. Yu, J. Shi, and N. Navarro, "Diversity factor prediction for distribution feeders with interpretable machine learning algorithms," in *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2020, pp. 1–5.

[16] "Caltrans PeMS," https://pems.dot.ca.gov/, accessed 8 May 2021.

[17] T. D. Chen, K. M. Kockelman, and M. Khan, "Locating electric vehicle charging stations: Parking-based assignment method for Seattle, Washington," *Transportation research record*, vol. 2385, no. 1, pp. 28–36, 2013.

[18] E. Kontou, C. Liu, F. Xie, X. Wu, and Z. Lin, "Understanding the linkage between electric vehicle charging network coverage and charging opportunity using GPS travel data," *Transportation Research Part C: Emerging Technologies*, vol. 98, pp. 1–13, 2019.

[19] "IPUMS NHGIS," https://www.nhgis.org/, accessed 8 May 2021.

[20] J. Bennett, R. Fry, and R. Kochhar, "Are you in the American middle class?" "https://www.pewresearch.org/fact-tank/2020/07/23/are-you-in-the-american-middle-class/", Pew Research Center, accessed 8 May 2021.