electronics

MDPI

Article

Factors Impacting Short-Term Load Forecasting of Charging Station to Electric Vehicle

Md Fazla Elahe [1,2], Md Alamgir Kabir [2,3,*], S. M. Hasan Mahmud [2,4] and Riasat Azim [5]

[1] Department of Software Engineering, Daffodil International University, Dhaka 1216, Bangladesh
[2] Centre for Advanced Machine Learning and Application (CAMLAs), Dhaka 1229, Bangladesh
[3] Artificial Intelligence and Intelligent Systems Research Group, School of Innovation, Design and Engineering, Malardalen University, Hogskoleplan 1, 722 20 Vasteras, Sweden
[4] Department of Computer Science, American International University-Bangladesh (AIUB), Dhaka 1229, Bangladesh
[5] College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China
* Correspondence: md.alamgir.kabir@mdu.se

Abstract: The rapid growth of electric vehicles (EVs) is likely to endanger the current power system. Forecasting the demand for charging stations is one of the critical issues while mitigating challenges caused by the increased penetration of EVs. Uncovering load-affecting features of the charging station can be beneficial for improving forecasting accuracy. Existing studies mostly forecast electricity demand of charging stations based on load profiling. It is difficult for public EV charging stations to obtain features for load profiling. This paper examines the power demand of two workplace charging stations to address the above-mentioned issue. Eight different types of load-affecting features are discussed in this study without compromising user privacy. We found that the workplace EV charging station exhibits opposite characteristics to the public EV charging station for some factors. Later, the features are used to design the forecasting model. The average accuracy improvement with these features is 42.73% in terms of RMSE. Moreover, the experiments found that summer days are more predictable than winter days. Finally, a state-of-the-art interpretable machine learning technique has been used to identify top contributing features. As the study is conducted on a publicly available dataset and analyzes the root cause of demand change, it can be used as baseline for future research.

Keywords: charging station; electric vehicle; forecast; power demand; privacy

Citation: Elahe, M.F.; Kabir, M.A.; Mahmud, S.M.H.; Azim, R. Factors Impacting Short-Term Load Forecasting of Charging Station to Electric Vehicle. *Electronics* **2023**, *12*, 55. https://doi.org/10.3390/electronics12010055

Academic Editor: Rui Castro

Received: 22 November 2022
Revised: 18 December 2022
Accepted: 19 December 2022
Published: 23 December 2022

1. Introduction

Climate change has become a serious concern, and countries are desperate to limit greenhouse gas emissions from the transportation sector. Internal combustion engine (ICE)-based vehicles contribute to greenhouse gas emissions. In the USA, CO2 emissions reached 5981 million metric tons in 2020, with the transportation sector contributing 27% (see Figure 1). People are moving away from ICE-based vehicles due to climate consciousness among vehicle users and government encouragement [1]. As EVs are getting popular, charging stations are built throughout the cities. The accurate short-term load forecasting of these EV charging stations is one of the critical concerns for the optimal operation of charging stations and power grids. Numerous factors are involved in determining the power need of EVs. The power demand of the charging station is extremely volatile, and intermittent charging behavior threatens the safety of the existing power system. PEV adoption increases daily peak demand, disrupts the distribution network, and results in an under-voltage situation. A study in [2] found that the increased penetration of EVs has an adverse effect on the transformer and may disrupt the service.

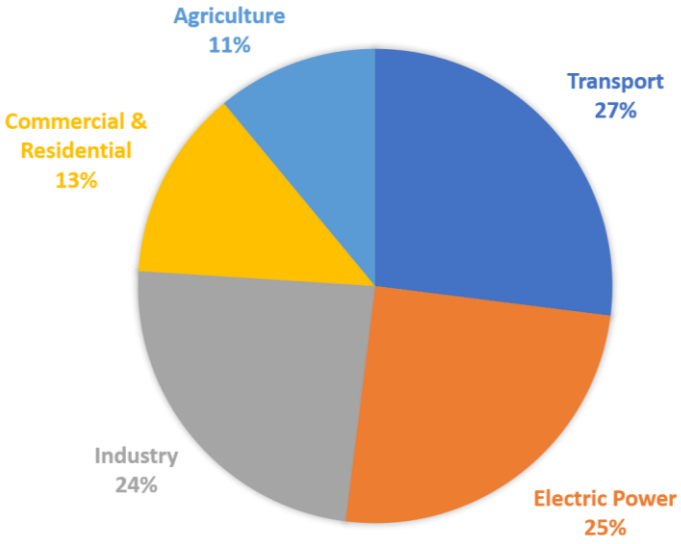

Figure 1. CO2 emissions by economic sector in the USA (source—United States Environmental Protection Agency).

It is challenging to forecast the power demand due to driving behavior, pricing process of utility, and various random factors [3]. Identifying valuable features from charging and surrounding data is crucial for improving forecasting accuracy. Several studies on EV charging station load forecasts have been conducted. There are two basic types of charging load forecasting methods reported in the literature. The first type makes use of electric vehicle (EV) driving features [4]. The forecasting model using driving characteristics is mainly formed from the state of charge (SOC), travel time, charging and discharging characteristics, vehicle model, etc. These characteristics are not the same for all EVs. The different characteristics of EVs are an obstacle for simulating a charging profile. It is computationally expensive to address each EV as a single entity for minimum charging rate because many decision variables are included [5]. One of the solutions for reducing computational complexity is the combined SOC-based methodology. A study in [6] proposed a combined SOC-based method that employs a simple mechanism: EVs with lower SOC charges at higher prices and EVs with higher SOC charges at lower prices. Reference [7] divides EV fleet into four categories to easily characterize charging behavior. However, such user profile-based forecasting techniques often violate the privacy of the user and expose private information. Global positioning system (GPS) data were used in [8] to collect the SOC information and model the EV charging. The second type of charging load forecasting method uses the charging station data to forecast the demand [9]. A comparative study of these two forecasting methods found comparable accuracy for both methods [10]. Most of the literature on EV charging station load forecasting uses common features such as SOC, driving distance, vehicle model, holiday, etc. However, as discussed earlier, privacy is the primary concern while modeling a forecasting technique based on user profile.

Conventional forecasting methods [11] are unable to take full advantage of external factors that influence the charging station's power demand. As a result, the effectiveness of these conventional procedures is not up to the expectations [12]. Later, power grids have benefited from the fast development of artificial intelligence (AI) technology. Algorithms such as support vector machines (SVMs), random forest (RF), artificial neural networks (ANNs), etc., have been employed to solve different challenges of the power grid. With no exception, these techniques have also been studied for power systems. The SVR model is used in [13] to forecast the power demand of an EV charging station. The schedulable capacity of EVs has also been investigated using random forest (RF) [14]. Performance can be improved for short-term forecasting using an ANN model [15].

To uncover the latent feature of EV charging and further improve the forecasting model's accuracy, deep learning (DL)-based techniques are dominating these days. One of the popular DL methods is recurrent neural networks (RNNs), which is often used for time series forecasting. The two most popular variations of RNNs: gated recurrent units (GRUs) [16], and the long short-term memory (LSTM) [17], are explored for time series forecasting. The performance of another DL approach, the convolutional neural network (CNN), is also investigated in [18]. In contrast, a comparative study of a few popular DL-based approaches has been presented in [19]. Numerical studies at various time steps show that DL techniques can accurately forecast super-short-term electricity demand. Moreover, the study found that LSTM performs best for reducing forecasting error among the various DL approaches.

With the increased adaption of electric vehicles, electricity demand is also increasing. Short-term load forecasting of charging station plays a big role in uninterrupted power supply. However, short-term load forecasting has been rarely explored in the literature. Moreover, it is necessary to understand the load-impacting factors of charging stations. This paper presents the forecasting result on publicly available charging station power demand datasets, using recent popular forecasters. Moreover, the data of two charging stations are simultaneously analyzed and used for forecasting purposes to exclude the analysis' randomness effect. Furthermore, the effect of the external features for forecasting and forecasting results on different day types are analyzed. Thus, the case study can serve as a state-of-the-art technique for future research.

The terms 'energy delivered', 'power demand', and 'load demand' are used interchangeably throughout the paper. Moreover, Caltech and JPL stand for Caltech EV charging station and JPL EV charging station, respectively. The rest of the paper is organized as follows: Section 2 discusses the system modeling. Dataset characteristics are given in Section 3. Section 4 analyzes the load demand of charging station to identify key features. The experimental setup is given in Section 5 and the results of the forecasters are presented in Section 6. Finally, Section 7 draws the conclusion.

2. System Modeling

One of the essential methods to improve forecasting accuracy is to include as many demand-changing factors as possible. So far, the demand-changing factor has been studied as part of improving forecasting performance where the focus on the demand-changing factor was limited. This paper analyzes different types of demand-changing factors so that they will be beneficial for modeling power demand. Another drawback is that most existing charging station load forecasting techniques focus on profiling charging behavior to forecast the power demand. Such techniques often use features such as SOC, driving distance, vehicle model, etc. SOC is the remaining capacity of the battery and can be measured as follows:

$$SOC(d) = \frac{C(d)}{C_m} \tag{1}$$

$$C(d) = C_m - C_u(d) \tag{2}$$

$$C_u(d) = x(d) * r + C_u(d-1) \tag{3}$$

where C_m is the maximum amount of charge that can store in the vehicle, and $C(d)$ is the remaining capacity of the battery after day d. $C_u(d)$ is the used capacity of day d and it depends on the distance driven on day d ($x(d)$), discharging rate r, and remaining capacity of the previous day ($d-1$). The charging time of an EV on day d can be expressed as follows:

$$T(d) = \frac{C_u(d)}{P_{battery}} \tag{4}$$

$$P_{battery} = P_{charger} * \eta \tag{5}$$

where $P_{battery}$ is the power consumed by battery, $P_{charger}$ is the power rate of the charger, and η is the efficiency of the charger. As seen from Equations (1)–(5), the EV's power demand

and charging time depend on features that can violate user privacy. The collection of these features is impractical for a public charging station. Instead of forecasting individual EVs, one alternative solution to this problem is to forecast the aggregated load. Individual load profiles are not required for aggregated load forecasting; therefore, privacy is preserved while predicting. This paper forecasts aggregated load using external features such as weather, historical, and calendar features.

3. Dataset Characteristics

ACN-Data [9], a dataset obtained from a charging network in Los Angeles, California, was used in this experiment. Caltech and JPL charging station data were chosen for the experiment out of the data of three charging stations in ACN-Data. Caltech is a research university with 55 electric vehicle supply equipment (EVSE) in the parking garage. Although EV charging is available to the general public at Caltech, most users are students, teachers, and staff. As Caltech's parking garage is close to the campus gymnasium, many EV users use the charging station while working out in the gymnasium. The charging station of JPL is only open to employees. As a result, the charging data of JPL is representative of workplace charging. In contrast, Caltech charging data is representative of a hybrid between public use and workplace charging.

The basic information of the considered charging stations is given in Table 1. A search on the dataset download website found the first session of Caltech is 25 April 2018, and JPL's is 5 September 2018. The last date selected for the search was 30 June 2021. For Caltech, the number of claimed sessions (claimed session is associated with user input and unclaimed is not associated with user input) is almost similar to the number of unclaimed sessions. However, for JPL, the number of claimed sessions is much higher than the number of unclaimed sessions. Another notable difference between the two charging stations is that Caltech's average charging time per session is smaller than JPL's. The EV penetration of JPL is relatively high; as a result, the average energy consumed per day for each EVSE, the average energy consumed per session, and the number of sessions per day are higher than Caltech's.

Table 1. Basic information of the charging stations.

Property	Caltech	JPL
Location	Caltech campus, California.	JPL campus, California.
Duration	1162 days	1024 days
Charging session found	1008 days	889 days
Number of sessions	30,052	31,543
Number of claimed sessions	15,167	29,477
Number of unclaimed sessions	14,884	2065
Avg. charging time per session (hour)	2.93	3.99
Number of EVSE	55	52
Avg. energy con. per session (kW)	4.95	9.63
Avg. energy con. per day per EVSE (kW)	9.14	14.12
Number of sessions per day	29.81	35.48
Min. power delivered in a session (kW)	0.501	0.50
Max. power delivered in a session (kW)	77.7	68.60
Idle occupied time per session (hour)	2.81	2.87

4. Analysis of EV Charging Station Load Demand

EV charging behavior can be changed due to different periodic and random effects. For accurate forecasting, these factors should be included in the forecasting model. One of the fundamental questions is what are the influential factors that affect load forecasting of a charging station. From the literature and experiments, different types of influential factors are found. Some load forecasting factors such as SOC and drive distance often violate users' privacy. In this paper, such influential factors are excluded from analysis and discuss the

factors that preserve the privacy of the user. A full calendar year with the same policy (price and charging policy for claimed and unclaimed sessions) is used for the analysis. As a major policy change occurred on 1 November 2018, the period from 1 January 2019 to 31 December 2019 is selected for our study. Mathematically, the cumulative power demand p of a charging station at time t can be described as follows:

$$p_t = \sum_{i=1}^{n} v_{it} \tag{6}$$

where v_i is the power demand of vehicle i and n is the number of vehicles connected to the charging piles of the charging station.

4.1. Seasonal Variation

Power demand exhibits different characteristics in different seasons for residential and industrial sectors. The analysis here is carried out to find the seasonal variation of the power demand of the EV charging station. A full seasonal year is different from a calendar year. A complete seasonal year spans from March to next year's February. There, the period from 1 March 2019 to 28 February 2020 is selected for seasonal analysis. A complete seasonal year has four seasons: spring (March, April, May), summer (June, July, August), fall (September, October, November), and winter (December, January, February). The daily average power demand of Caltech and JPL for four seasons is shown in Figure 2. Figure 2 shows that the highest daily demand is in the spring season, followed by the fall season for both charging stations. Such power demand may result from pleasant weather conditions in those two seasons and outgoing human activity. However, the lowest demand for Caltech is in winter but for JPL it is in summer. The lowest power demand in those seasons may result from the user activity of the charging station as Caltech is a hybrid charging station, and JPL is a workplace charging station.

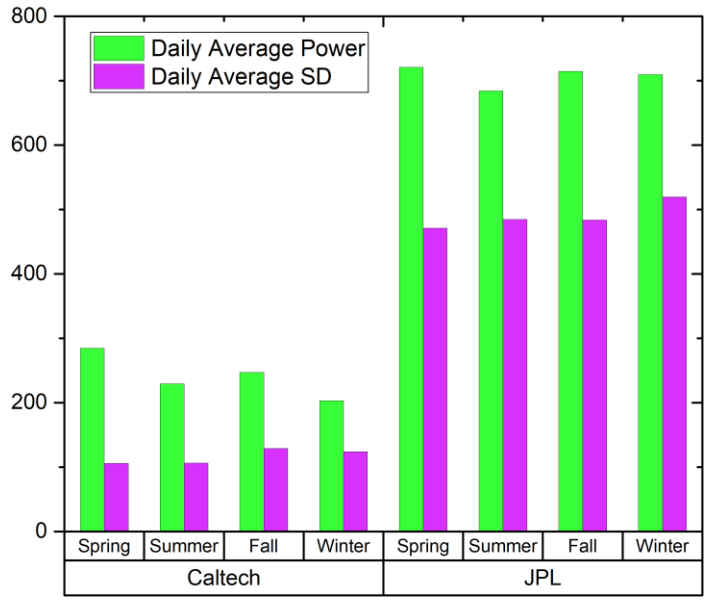

Figure 2. Seasonal variation of power demand.

Moreover, if we carefully examine the daily average standard deviation of the four seasons (Figure 2), both the charging stations exhibit a few similar behaviors. Although daily consumption in spring is highest among all seasons, the standard deviation in spring is the lowest. In other words, the daily consumption in spring follows a consistent power demand. Another vital point to notice is that the average daily demand in winter is low, whereas the standard deviation is high. Such characteristic indicates the inconsistent power demand in winter.

4.2. Weather Conditions

Weather conditions are important for load forecasting because they are related to HVAC and control human activity. The analysis is carried out to determine whether the weather conditions affect the charging station's power demand and, if so, what are the most affecting weather variables. Fifteen weather variables and the data of two charging stations are considered. The Pearson correlation coefficient is the de facto statistical indicator for measuring the relationship between two variables. In the case of two n-dimensional variables $X = \{x_1, x_2, \ldots, x_n\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$, their Pearson correlation coefficient r_{xy} is as follows:

$$r_{xy} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - (\sum x_i)^2}\sqrt{n\sum y_i^2 - (\sum y_i)^2}} \quad (7)$$

The result of the Pearson correlation coefficient among weather variables and daily demand is given in Table 2. A negative correlation coefficient indicates the weather variable is negatively correlated with the daily demand, whereas a positive correlation coefficient indicates the weather variable is positively correlated with the weather variable. It is shown from the table that the considered weather variables have negligible influence over the power demand of the charging station. The power demand of Caltech is more sensitive to weather variables than that of JPL. Moreover, the two charging stations exhibit different characteristics for the same weather variable.

Table 2. Correlation scores of weather variables with power delivered.

Weather Variable	Caltech	JPL
Maximum temperature	−0.070750	0.016431
Average temperature	−0.103542	0.015418
Minimum temperature	−0.082004	0.039647
Maximum dew point	−0.062207	0.038575
Average dew point	−0.070037	0.016149
Minimum dew point	−0.075114	0.018542
Maximum humidity	−0.011453	0.007398
Average humidity	0.002043	0.025363
Minimum humidity	−0.023895	0.021847
Maximum wind speed	0.038898	0.000188
Average wind speed	0.039866	−0.019258
Minimum wind speed	0.005172	0.005300
Maximum pressure	0.176864	0.011223
Average pressure	0.103767	−0.031547
Minimum pressure	−0.042822	−0.039377

4.3. Calendar Features

Keeping in mind that human activity changes over the days of the week and months of the year, the power demand of charging stations is investigated for calendar features. The average power demand of different days of the week is given in Figure 3. It is clear from the figure that the power demand on Saturday and Sunday drop significantly for both charging stations. Moreover, the power demand of JPL charging station on Friday is relatively low among all other weekdays. Reference [20] finds very low intra-week variations of charging habits, whereas our analysis finds a significant difference between weekdays' and weekends' power demand. The intra-week difference may be the result of the charging station location. Our experimental datasets are collected from charging stations, which office employees mostly use. As a result, the demand for electricity on weekends is less than the weekdays. It can be concluded from the above observation that the day of the week and the charging station's location are important for forecasting the demand of a charging station.

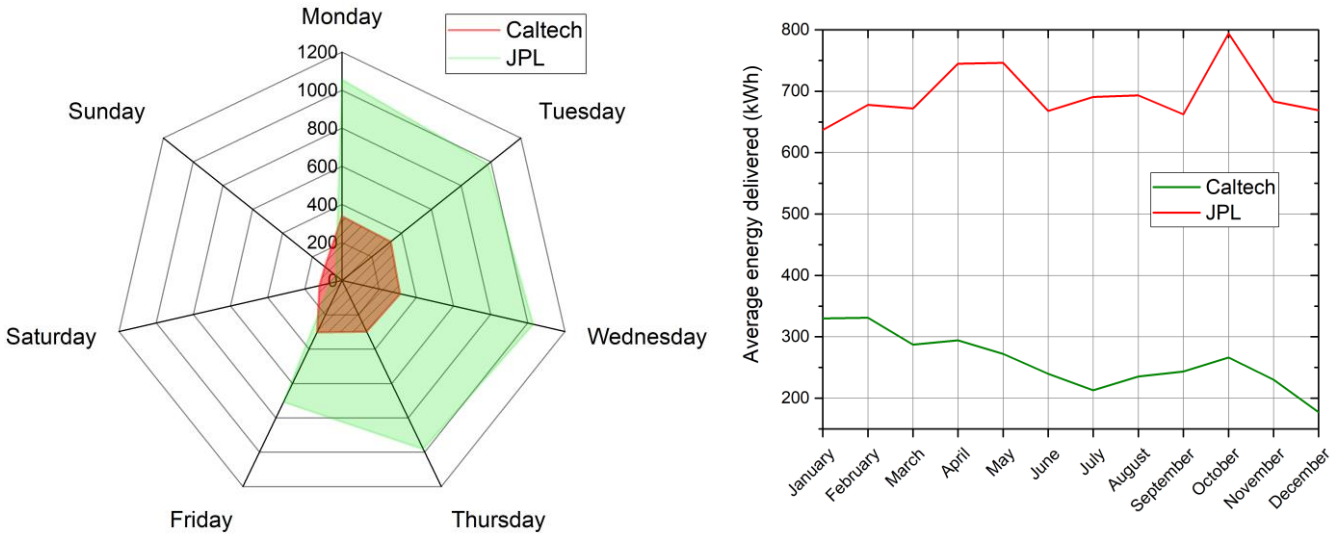

Figure 3. Average power demand of different days and months.

The monthly average energy delivered is presented in Figure 3. It is observed from the figure that the monthly average energy delivered in Caltech is lower than in JPL. JPL has similar power demands in April–May and July–August, and Caltech has similar power demands in January–February and March–April. A surge in demand for both the charging stations was observed in October. The lowest demand is found in January for JPL and in December for Caltech.

Caltech's average daily holiday and non-holiday demands are 256.89 and 259.64, respectively. JPL's average daily holiday and non-holiday demands are 662.51 and 696.22, respectively. The list of holidays for Los Angeles is obtained from [21]. The holiday power demand for both charging stations is slightly lower than the non-holiday power consumption. If holiday and non-holiday power demand are compared to weekend and weekday power demand, it can be seen that there is a considerable demand difference even though weekends and holidays are off days. The power demand on holiday indicates many more charging sessions than on the weekend.

4.4. Arrival and Departure

The arrival and departure distribution for the considered charging stations are shown in Figure 4. Weekends' and weekdays' distribution of both charging stations is plotted in the figure. The figure shows a significant peak for morning arrival on weekdays. As both of the charging stations are in the workplace, the peak indicates the arrival of the employees. However, the charging station near residences may show a peak for evening arrival for the returnee. The distribution of departure is analogous to arrivals. The number of vehicles leaving charging station increases as the workday ends and peaks near 5–6 pm on weekdays. As expected, the weekend has a completely different distribution than the weekdays. The arrival and departure for the weekend are much lower for both charging stations. Although the EV penetration in JPL is higher than in Caltech for weekdays, it is lower on the weekend because Caltech is open for public use, and JPL is available to employees only. Another important point to note is that the arrival and departure distribution on weekends is much more uniform (no sharp peak like weekday for arrival and departure) for both charging stations. This is most likely due to the wildly heterogeneous weekend work schedules.

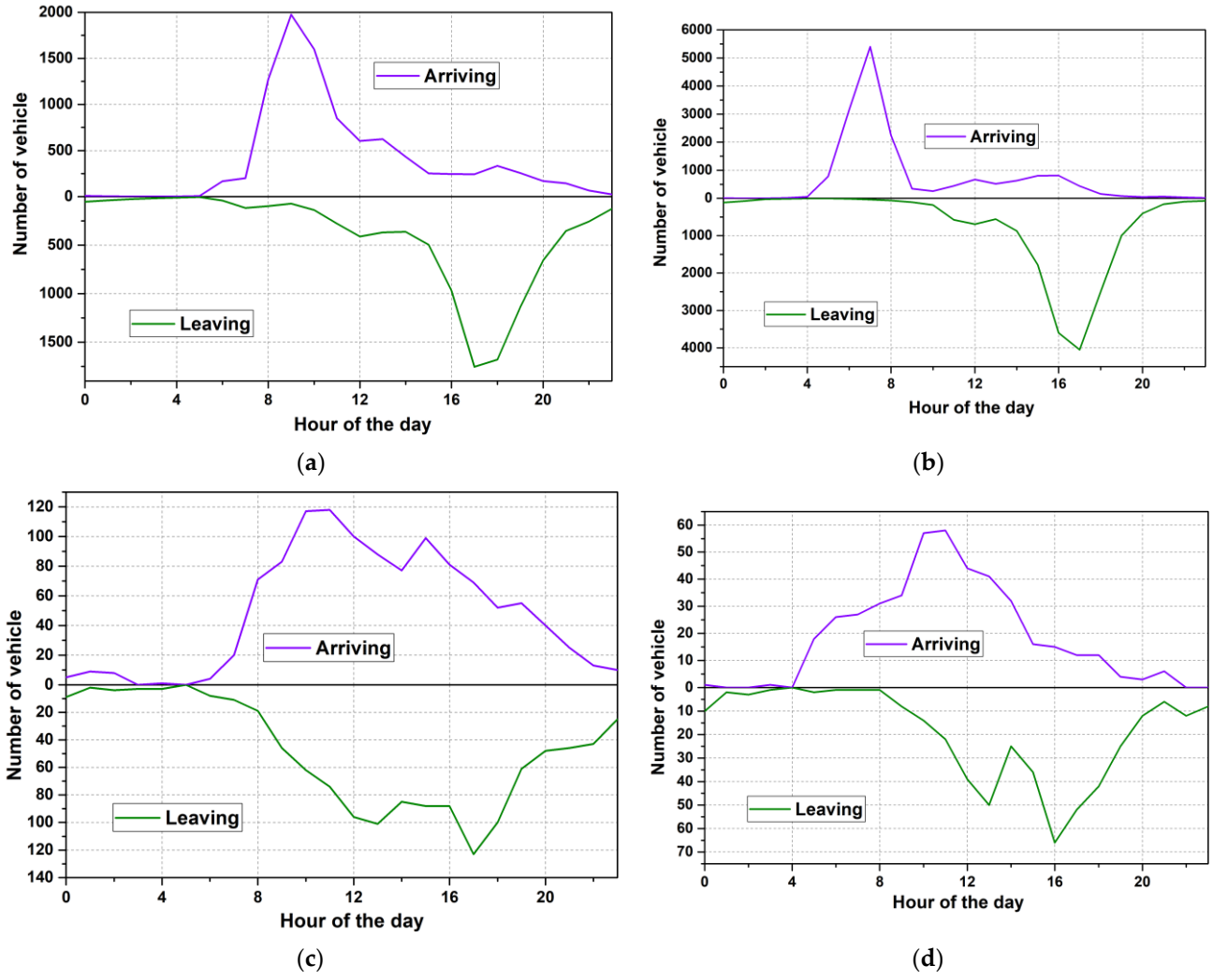

Figure 4. Hourly arrival and departure of vehicles. (**a**) Caltech weekday. (**b**) JPL weekday. (**c**) Caltech weekend. (**d**) JPL weekend.

4.5. Charging Session

In this section, charging sessions throughout the days of the week were analyzed to find the difference in intra-week charging behavior. Figure 5 depicts the charging behavior of different days of the week. It shows the number of average sessions, average session duration, and average energy delivered per session throughout the days of the week. The average session and average session duration for both charging stations drop for Saturday and Sunday. However, the number of average sessions for Caltech is more than the JPL on weekends. The reason behind such behavior is that the Caltech charging station is open to the public, but the JPL one is only open to the employees. As Caltech is open to the public, it welcomes the user from the community on weekends. Another important observation is that the number of average sessions and average session duration decreased for JPL on Friday. The same characteristic is also shown for JPL's average daily power demand on Friday (Figure 3). The smaller charging session, shorter session duration, and less energy delivery on Friday for JPL can be considered the implicit charging characteristics for the specific day and specific charging station. Although the session duration for JPL on Monday is similar to other weekdays, the average energy delivery is highest among all other days, as shown in the Figure of average energy delivered per session. The lower SOC level after the weekends may be the reason for higher power demand on those days. Similarly, the session duration for Caltech on Sunday is the lowest, but the average energy delivery is the highest among all other days. Such characteristics of the sessions can be the

combined effect of lower SOC after a Saturday (weekend) and a smaller number of active sessions compared to the capacity of the station.

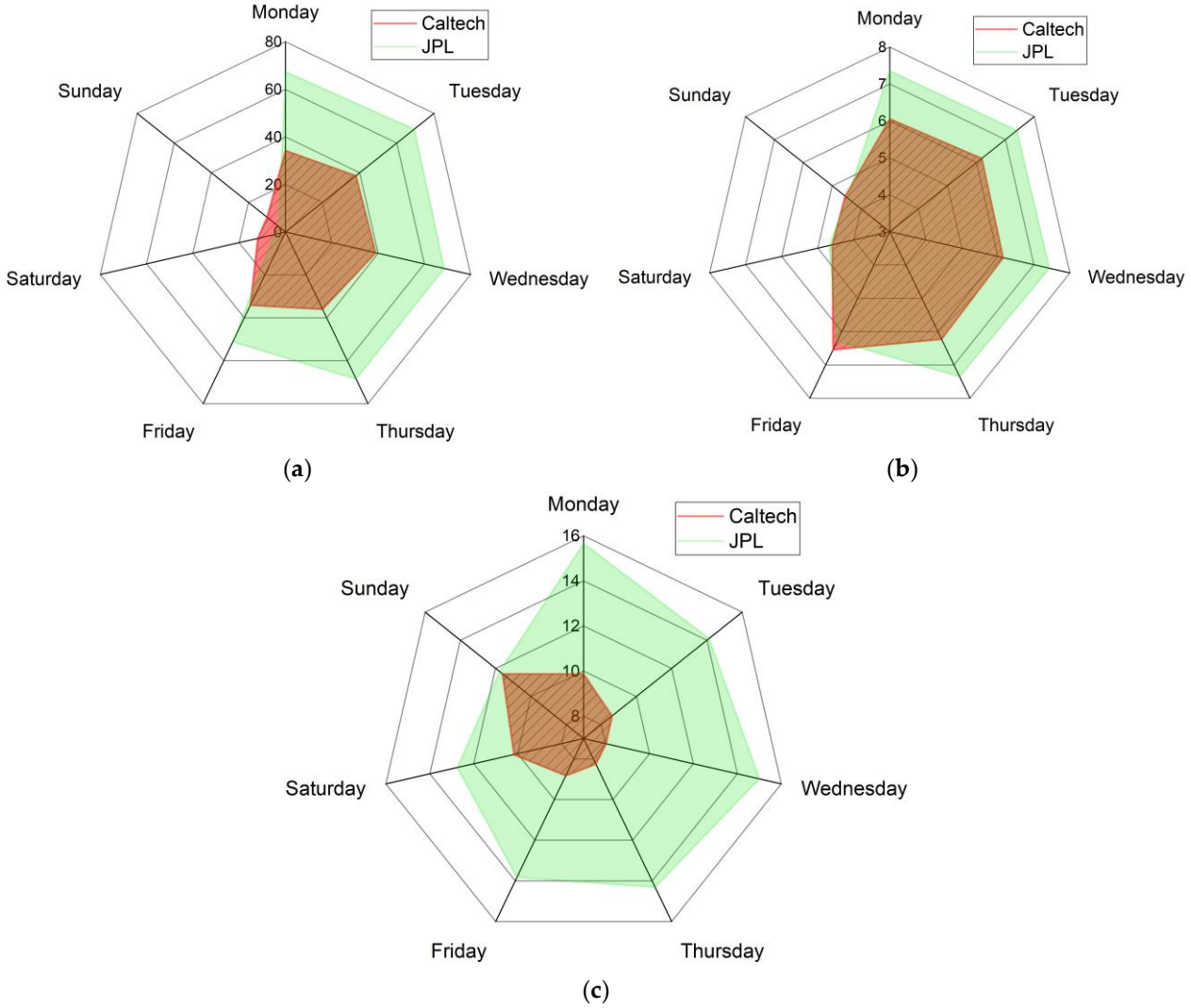

Figure 5. Change of charging behavior on different days of the week. (**a**) Number of average sessions. (**b**) Average session duration. (**c**) Average energy delivered per session.

4.6. Policy

Policy changes such as price, subsidy scheme, etc., are often considered load-affecting factors for demand response programs. This analysis here aims to identify how policy change affects the energy demand of the charging station. On 1 November 2018, Caltech and JPL implemented a major policy change. Figure 6 shows the power demand from 1 September 2018 to 31 December 2018. Claimed sessions refer to the session when a user specifies the energy requirement and duration of the session, whereas unclaimed sessions use default values. Before 1 November, an unclaimed session at Caltech received 14 kWh over 12 h and the claimed sessions were free. After the policy change, the claimed session became USD 0.12/kWh, and the unclaimed session was set to terminate after 30 min in Caltech. For JPL, claimed sessions have always been USD 0.10 per kWh, whereas unclaimed sessions have changed from 17.5 kWh/4 h to terminate 30 min. The energy demand for Caltech drops after 1 November, and JPL power demand does not show any big difference. Due to the fact that the cost became comparable to in-home charging after the policy change, Caltech's drop in power demand can be described as decreasing community users. However, no policy change has been implemented for claimed sessions

at JPL, and our intuition is that the high EV penetration overshadows the policy change for unclaimed sessions.

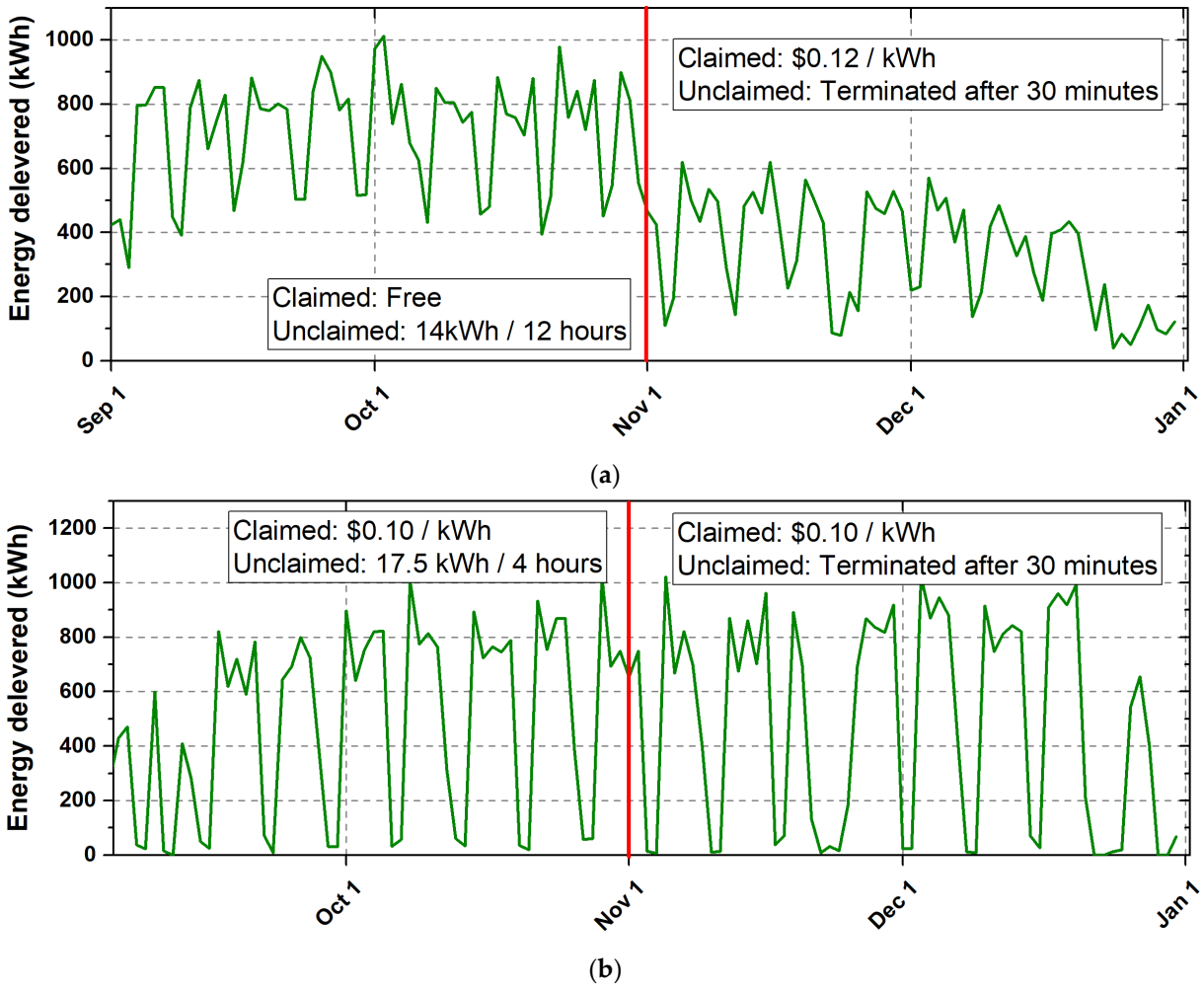

Figure 6. Policy change vs. power demand. (**a**) Caltech. (**b**) JPL.

4.7. Idle Occupancy

Another important fact that should be paid special attention to for forecasting the power demand is the capacity of the charging station. If the demand exceeds the charging station's capacity, there should be some mechanism to mitigate the demand. Charging station utilization can play a major role in this regard. One of the crucial indicators of charging station utilization is the analysis of idle occupancy. The maximum utilization of a charging station means the use of the maximum capacity of the charging station. When the maximum utilization obtained once the charging is completed, the EV should free the EVSE as soon as possible and give a chance to the EV waiting for charging. However, in reality, EVs are left for hours after charging is completed. Such scenarios demand huge investment for infrastructure development and capacity increase. A thorough analysis of the capacity utilization can help avoid extra expenses. Generally, the electricity demand of a charging station at a specific time is used to obtain the utilization of the charging station at that specific time. However, the analysis of idle occupancy can give more insight into this. Figure 7 shows the daily idle occupancy and energy delivered in January 2019. An important point to note from the figure is that the idle occupied time increases with the energy delivery for most days. Generally, the idle occupied time of a charging station increases when the charging station is away from maximum utilization because when the maximum utilization is achieved, idle occupancy should be 0. Moreover, the idle occupancy per session for different days of the week is presented in Table 3. The table contains a few

notable points. First, on weekends, idle occupancy is less than on the weekdays. Such characteristics may be the result of spending less time in the office on weekends. Second, idle occupancy on Friday for JPL is less than on other weekdays. Average power demand (Figure 3), average session duration (Figure 5b), and idle occupancy (Table 3) indicate a different charging behavior for JPL on Friday. Third, on Monday, the idle occupancy is below the average of the weekday idle occupancy for both charging stations. The average session duration (Figure 5b) and average energy delivered per session (Figure 5c) suggest that the energy requirement on Monday is higher. As a result, the session comprises less idle occupied time before leaving the station.

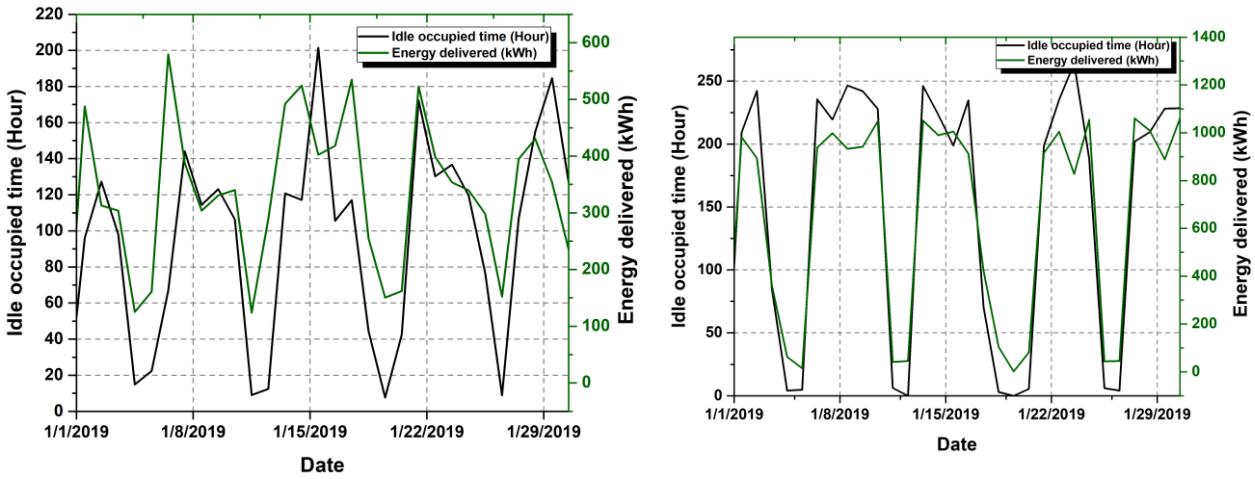

Figure 7. Idle occupancy vs. energy delivered. (**a**) Caltech. (**b**) JPL.

Table 3. Average idle occupancy (hour) per session.

Day	Caltech	JPL
Monday	3.12	2.78
Tuesday	3.39	3.24
Wednesday	3.39	3.07
Thursday	3.24	3.17
Friday	3.48	2.50
Saturday	2.17	1.74
Sunday	1.89	1.87

4.8. Historical Data

Generally, a function relates the predicted power with the historical consumed power. The predicted power $\hat{P}$ at time t can be expressed as follows:

$$\hat{P}(t) = f(P(t-1), P(t-2), \ldots) \tag{8}$$

where f is the function that relates predicted power with past consumed power and $P(t-i)$ is the consumed power at $(t-i)$. The correlation score of the power delivered of a day with the previous seven days' power delivered is given in Table 4. It is seen from the score that the electricity demand is highly correlated with historical demand data. The same day of the previous week, which is seven days before the demand data, has the highest correlation. The finding can be described as the demand for electricity for EV charging follows a periodic manner. It is also observed that the second-highest correlation score was found for the previous day. This is logical as the demand for electricity for residential houses and industries has similar characteristics.

Table 4. Correlation score of historical power delivered.

Day	Caltech	JPL
1 day before	0.341762	0.352255
2 days before	−0.057642	−0.233426
3 days before	−0.030332	−0.404988
4 days before	−0.020157	−0.437741
5 days before	−0.089821	−0.237613
6 days before	0.266087	0.311596
7 days before	0.647215	0.707768

5. Forecasting of Power Demand

5.1. Input Selection and Processing

The analysis of EV charging station load demand uncovers different influential factors. The holiday feature for the charging station is obtained from [21], and the weather condition data have been collected from [22]. According to the requirement of the forecaster, seasonal variation, month of the year, and the day of the week features are one-hot encoded [23]. The summary of different types of features is given in Table 5. The arrival and departure are not included in the feature summary list. There are two important observations from the analysis of arrival and departure. First, the number of vehicles on weekends is less than on the weekdays. The feature is included as day of the week (Monday–Sunday). Second, the time of arrival and departure of the vehicles. As the aim of this paper is to forecast the daily power demand of the charging station, the time of arrival and departure of the vehicle could not be included. However, this feature can play a major role in hourly load demand forecast.

Table 5. Summary of the features.

Feature Type	Number of Features	Feature
Seasonal variation	4	spring, summer, fall, winter
Weather conditions	15	Average temperature, maximum temperature, minimum temperature, average pressure, maximum pressure, minimum pressure, average dew point, maximum dew point, minimum dew point, average wind speed, maximum wind speed, minimum wind speed, average humidity, maximum humidity, minimum humidity
Calendar features	20	Monday–Sunday (7) January–December(12) Holiday
Charging session	2	Number of sessions Session duration
Policy	1	Policy
Idle occupancy	1	Idle occupancy
Historical data	7	Day 1–Day 7

Among the seven different types of features, some features contain repetitive information. For example, seasonal variation and average daily demand in different months (January–December) may result from the season's weather conditions. The repetitive inclusion of the same information increases the model complexity and deteriorates the model's performance. To overcome this challenge, a scikit learn-based univariate feature selection technique is used and the 15 best features are selected. For Caltech, the selected features are Policy, Day 1, Day 2, Day 3, Day 4, Day 5, Day 6, Day 7, Avg. temperature, Min. temperature, Max. dew point, Avg. dew point, Idle occupancy, Number of sessions, and Session duration. The selected features are then used to train the model. For JPL, the selected features are fall, Saturday, Sunday, Tuesday, Day 1, Day 2, Day 3, Day 4, Day 5,

Day 6, Day 7, Max. dew point, Idle occupancy, Number of sessions, and Session duration. The selected features are then used to train the model. All the features are normalized according to the following equation:

$$x_i' = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{9}$$

where x_i is the original value and x_i' is the normalized value. The normalization converts the feature values into the interval [0,1].

5.2. Forecaster Selection

Data that have been preprocessed are ready to be used for forecasting. Recent publications found that deep learning-based forecasters can learn the latent features and improve forecasting accuracy. In this paper, five deep learning-based forecasters are selected. The first selected forecaster is LSTM [17]. It has plenty of applications for time series forecasting and has shown its capability in various learning and prediction problems. Our second selected forecaster Bi-LSTM is able to consider both past and future information [24]. Researchers have explored the model performance combining CNN and LSTM for extracting temporal and spatial features in recent years. The CNN–LSTM forecaster has also shown promising results for energy consumption prediction, acknowledging that CNN removes the noise considering the correlation between multivariate variables, and LSTM extracts temporal information to generate prediction [25]. The next selected forecaster ConvLSTM has also shown its ability for prediction problems [26]. The last selected forecaster is GRU. It is an improved version of RNN. The most important benefit of using GRU is handling long-term dependency while keeping the structure simple [16].

5.3. Model Performance Evaluation

In this paper, two popular evaluation metrics are utilized to evaluate the effectiveness of the forecasters: root mean square error (RMSE) and mean absolute error (MAE). The equations of the evaluation metrics are given below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{10}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{11}$$

where n is the number of samples, y_i is the actual value, and $\hat{y}_i$ is the forecasting value. However, another popular evaluation metric, mean absolute percent error (MAPE), is not considered for this paper because power demand on weekends is extremely low and sometimes has even no power required on the weekend. In such a case, the following equation will produce an undefined result:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left(\left|\frac{\hat{y}_i - y_i}{y_i}\right| \times 100\right) \tag{12}$$

5.4. Experimental Setup

In this experiment, the data of two publicly available EV charging stations have been used to exclude the randomness effect from the forecasting result analysis. Three cases have been developed for the forecasting result analysis. The first case is developed to analyze the effect of including features for EV charging load forecasting. The second case is designed to analyze the forecasting performance for weekdays and weekends. Finally, the third case demonstrates the forecasting performance in summer and winter. The preprocessed dataset is divided into a training set and a test set. The training set is then fed to the selected forecaster, and the test set evaluates the forecasting performance.

One of the crucial points to improve the forecasting performance of forecasters is the fine-tuning of the hyper-parameter. This experiment applies expert opinion and trial and error methods to set the hyper-parameter. One common thing in the selected forecaster is that all the forecasters are based on neural networks. The most important hyper-parameter for the neural network-based forecaster is the number of layers. The dilemma arises while setting the layer number because if it is set too big, it will cause overfitting, and too small will cause underfitting. The number of hidden layers for GRU, Bi-LSTM, and LSTM is set to three, whereas the hidden layer for CNN–LSTM and convLSTM is set to two. A dropout layer with a value of 0.2 is added before the output layer to prevent overfitting. For all the forecasters, the number of nodes is set to 50, and the filter size is set to 64. Relu, adam, and MSE are chosen as activation, optimizer, and loss functions, respectively. The number of epochs is set to 100 for all the forecasters. Finally, the time step is set to 3 for better learning and prediction. Important parameter of the algorithms is given in Table 6.

Table 6. Summary of the parameters.

Algorithm	Hidden Layers	Number of Nodes	Filter Size	Activation Function	Optimizer	Loss Function
GRU	3	50	64	Relu	adam	MSE
Bi-LSTM	3	50	64	Relu	adam	MSE
LSTM	3	50	64	Relu	adam	MSE
CNN–LSTM	2	50	64	Relu	adam	MSE
convLSTM	2	50	64	Relu	adam	MSE

6. Results

6.1. Case 1 (Effect of External Features in Forecasting)

This case is designed to show the effect of using external features. The input of the forecaster with external features is discussed in the input selection and preprocessing section. As a few time-dependent features are considered for forecasting, as this type of forecaster is often termed as multivariate forecaster. The forecaster without external features is often called a univariate forecaster. Deep learning models can be used for univariate forecasting if multiple timesteps are used. The timesteps help the deep learning model map the input to predict the future demand. For this case, the period from 1 September 2018 to 28 February 2021 is selected as the training set, and the period from 1 March 2021 to 30 May 2021 is selected as the test set. The test set is used to evaluate the predicting performance, and the training set is used to train the model. The experimental results are shown in Table 7. It is evident from the result that the forecaster with external features performs better than the forecaster without external features regardless of forecaster, performance metrics, and charging station. The average accuracy improvement with the help of external features is 42.73% and 41.76% for RMSE and MAE, respectively. The degree of improvement the external feature brings to the model is different for the two charging stations. Due to external feature inclusion, the forecasting error reduces by 70.33% for the JPL charging station, whereas Caltech shows a 13.19% accuracy improvement. We believe that the location of the charging station plays an important role in predicting the power demand.

6.2. Case 2 (Forecasting Performance on Weekdays and Weekends)

We have already discussed that the power demand of a workplace charging station is not the same for weekdays and weekends. This case is designed to compare the effect of external features on weekdays' and weekends' power demand forecasting. Like in the previous case, the training set is selected as the period from 1 September 2018 to 28 February 2021. Weekdays from 1 March 2021 to 30 May 2021 are selected as the test set of weekdays, and weekends of the same duration are selected as the test set of weekends. The result of the experiment is given in Figure 8. There are two important points to note. First, the accuracy with external features is 56.20% higher than without external features for all scenarios for weekdays and weekends. External features are the cause of increased

accuracy because external features allow for a better understanding of load-changing rules. Second, the improvement of forecasting accuracy for weekends is 46.89%, which is lower than the weekday forecasting improvement by 65.52% due to there being fewer charging sessions on weekends than on weekdays.

Table 7. Effect of external feature in forecasting.

Station	Forecaster	Without External Features		With External Features	
		RMSE (kWh)	MAE (kWh)	RMSE (kWh)	MAE (kWh)
Caltech	LSTM	78.65	63.16	60.53	46.83
	Bi-LSTM	78.54	64.01	64.95	51.52
	CNN–LSTM	70.19	54.03	60.51	47.42
	ConvLSTM	74.93	57.52	68.41	54.44
	GRU	71.60	56.67	6.51	54.97
JPL	LSTM	196.81	135.34	52.27	38.48
	Bi-LSTM	204.61	138.89	57.87	43.97
	CNN–LSTM	160.06	120.37	47.16	36.55
	ConvLSTM	189.76	130.18	48.66	37.08
	GRU	165.81	120.98	47.26	35.56

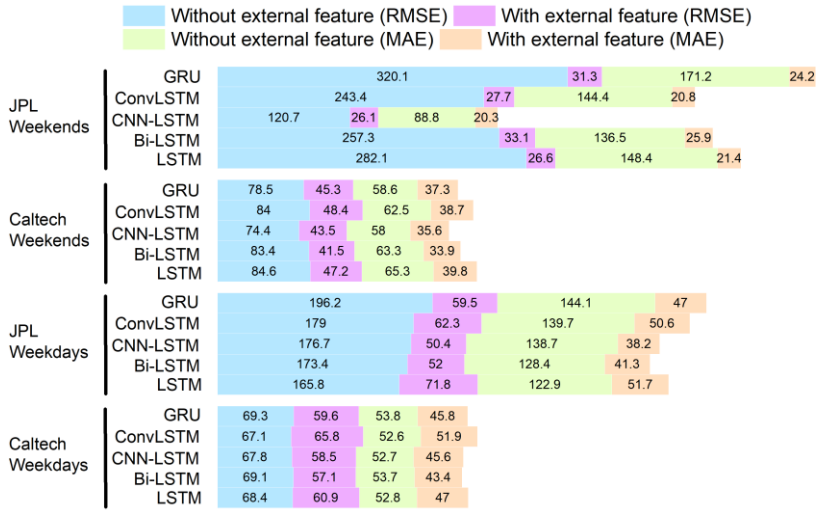

Figure 8. Forecasting performance on weekdays and weekends.

6.3. Case 3 (Forecasting Performance for Summer and Winter Days)

According to the literature, the performance of load forecasting for the grid and household loads varies in summer and winter days. As EV also has an HVAC system and human activity changes from the summer to winter season, power demand changes as well. This case is designed to compare the forecasting performance for the summer and winter seasons. The training set for the summer season is the period from 1 September 2018 to 31 May 2020, and the test set is the period from 1 June 2020 to 31 August 2020. The training set for the winter season is the period from 1 September 2018 to 30 November 2020, and the test set is the period from 1 December 2020 to 28 February 2021. The result of the experiment is presented in Figure 9. The result shows that summer days are more predictable than winter days without external features. Previously, we have seen that the standard deviation value in winter is bigger than in summer. As a result, the winter is less predictable than the summer. The uncertainty and diverse driving behaviors in winter do not allow for the model to predict well enough. However, with the help of external features, the forecasting error becomes 40.62% lower for summer and 48.27% lower for winter.

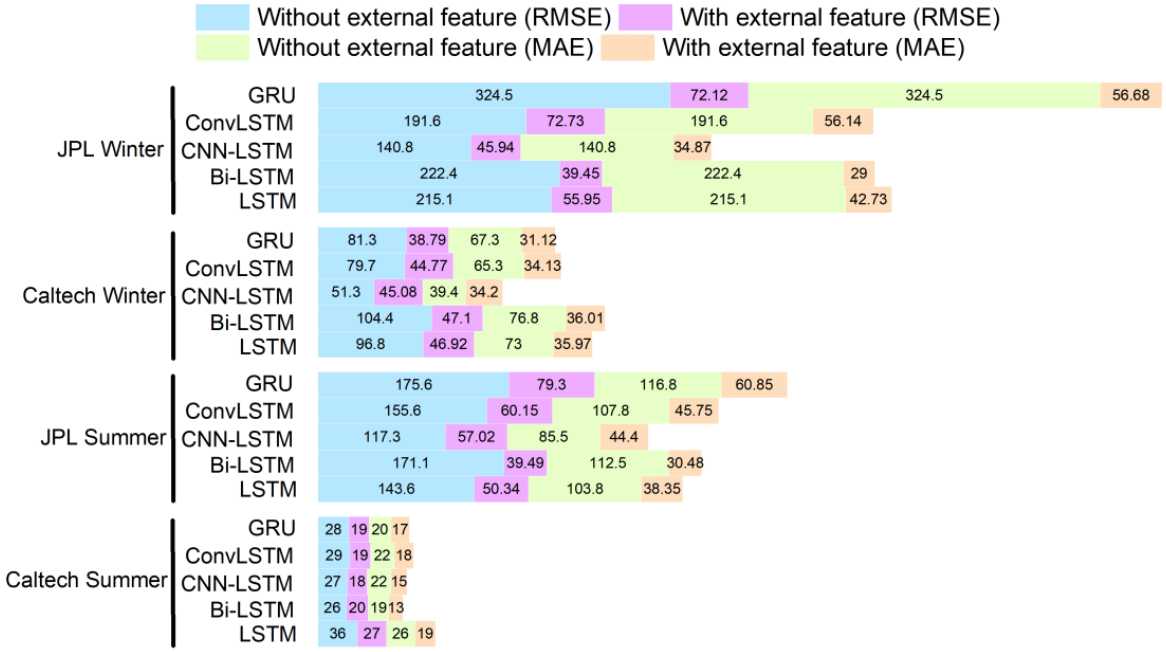

Figure 9. Forecasting performance for summer and winter days.

6.4. Interpretability Analysis of Factors

In Section 4, we attempted to investigate the relationship between various external factors and the load demand of the charging station. Moreover, in Section 5, the experiments are conducted to explore the combined effect of external features in forecasting the load demand of charging stations. However, the degree of improvement each individual feature brings to the forecasting model is not known from the analysis and experiment. Generally, machine learning models are known as the black box technique because it is hard to understand the decision-making procedure once trained. With continuous research, IML (interpretable machine learning) is able to interpret and analyze each factor. The contribution of each feature for a model could be calculated through a state-of-the-art IML technique called SHapley Additive exPlanations (SHAP) [27]. Although correlation can give us hints about the potential feature, the high correlation of a feature with the target variable does not necessarily mean that it has good ability in accuracy improvement because correlation does not consider the combined impact of the features. On the other hand, SHAP is based on the idea that the outcome of each possible combination of features should be taken into account when determining the importance of a single feature. The SHAP summary plot of both charging stations is given in Figure 10, which shows the feature density and its impact on model output. Each dot on the SHAP summary plot corresponds to a day in the study. The impact of that feature on the model's prediction for that day is indicated by the position of the dot on the x-axis. Multiple dots that land at the same x position pile up to show density. Furthermore, the color of the dot shows the value of the feature (red high, blue low). The top two contributing features for both the charging stations are number of sessions and session duration. When the value of these two features increases, the SHAP value increases as well.

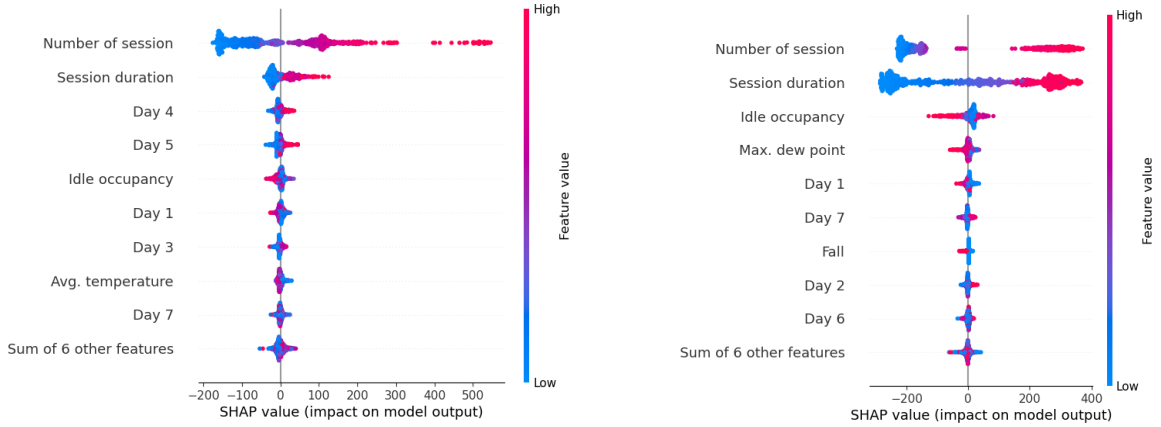

Figure 10. Feature density and its impact on model output. (**a**) Caltech. (**b**) JPL.

7. Conclusions

This paper analyzes eight different types of load-affecting factors and studies two EV charging stations, keeping the user's privacy in mind. The analysis uncovers a few dominant features such as number of sessions, session duration, and idle occupancy. The average accuracy improvement with these features is 42.73% in terms of RMSE. Moreover, this paper utilizes a few popular load forecasting models to forecast the power demand of the EV charging station. Three cases are designed to explore the effect of external features for improving forecasting accuracy. The results found that the inclusion of the dominant features in forecasting models improves forecasting accuracy in all three cases. Finally, the load-affecting factors are interpreted using a state-of-the-art IML technique called SHAP to identify the crucial features for improving the accuracy of the forecasting model. As the analysis is performed on the data of two publicly available charging stations, it can serve as a baseline for future research.

The scope of this research was limited to forecasting the daily electricity demand of EV charging stations. In future research, we will examine the effect of features on more granular-level forecasting. Moreover, long-term forecasting may help urban planners make better decisions. Deep learning algorithms are often criticized for the high number of parameters and training times. Research on minimizing the parameter and training times will certainly improve the efficiency of the algorithm.

Author Contributions: Conceptualization, M.F.E.; methodology, M.F.E.; software, M.F.E.; validation, M.F.E., M.A.K., S.M.H.M. and R.A.; investigation, M.F.E.; writing—original draft preparation, M.F.E.; writing—review and editing, M.A.K., S.M.H.M. and R.A.; visualization, M.F.E., M.A.K., S.M.H.M. and R.A.; funding acquisition, M.A.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to acknowledge Mälardalen University for supporting the publication charges.

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

References

1. Rominger, J.; Farkas, C. Public charging infrastructure in Japan—A stochastic modelling analysis. *Int. J. Electr. Power Energy Syst.* **2017**, *90*, 134–146. [CrossRef]
2. Lou, Y.; Wu, C.; Shi, Z.; Yang, R. Evaluation of EV penetration level limit in distribution system applying charging and scheduling strategies. *Sustain. Energy Grids Netw.* **2022**, *32*, 100922. *in press* [CrossRef]

3. Ahmadian, A.; Mohammadi-Ivatloo, B.; Elkamel, A. A Review on Plug-in Electric Vehicles: Introduction, Current Status, and Load Modeling Techniques. *J. Mod. Power Syst. Clean Energy* **2020**, *8*, 412–425. [CrossRef]
4. Li, X.; Zhang, Q.; Peng, Z.; Wang, A.; Wang, W. A data-driven two-level clustering model for driving pattern analysis of electric vehicles and a case study. *J. Clean. Prod.* **2019**, *206*, 827–837. [CrossRef]
5. Einaddin, A.H.; Yazdankhah, A.S. A novel approach for multi-objective optimal scheduling of large-scale EV fleets in a smart distribution grid considering realistic and stochastic modeling framework. *Int. J. Electr. Power Energy Syst.* **2020**, *117*, 105617. [CrossRef]
6. Islam, S.; Mithulananthan, N.; Hung, D.Q. A Day-Ahead Forecasting Model for Probabilistic EV Charging Loads at Business Premises. *IEEE Trans. Sustain. Energy* **2018**, *9*, 741–753. [CrossRef]
7. Zheng, Y.; Shao, Z.; Zhang, Y.; Jian, L. A systematic methodology for mid-and-long term electric vehicle charging load forecasting: The case study of Shenzhen, China. *Sustain. Cities Soc.* **2020**, *56*, 102084. [CrossRef]
8. Yang, T.; Xu, X.; Guo, Q.; Zhang, L.; Sun, H. EV charging behaviour analysis and modelling based on mobile crowdsensing data. *IET Gener. Transm. Distrib.* **2017**, *11*, 1683–1691. [CrossRef]
9. Lee, Z.J.; Li, T.; Low, S.H. ACN-Data: Analysis and applications of an open EV charging dataset. In Proceedings of the e-Energy 2019—ACM International Conference on Future Energy Systems, Phoenix, AZ, USA, 25–28 June 2019; pp. 139–149.
10. Majidpour, M.; Qiu, C.; Chu, P.; Pota, H.R.; Gadh, R. Forecasting the EV charging load based on customer profile or station measurement? *Appl. Energy* **2016**, *163*, 134–141. [CrossRef]
11. Louie, H.M. Time-Series Modeling of Aggregated Electric Vehicle Charging Station Load. *Electr. Power Compon. Syst.* **2017**, *45*, 1498–1511. [CrossRef]
12. Paterakis, N.G.; Mocanu, E.; Gibescu, M.; Stappers, B.; van Alst, W. Deep learning versus traditional machine learning methods for aggregated energy demand prediction. In Proceedings of the 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Torino, Italy, 26–29 September 2017; pp. 1–6. [CrossRef]
13. Wu, J.; Wang, Y.-G.; Tian, Y.-C.; Burrage, K.; Cao, T. Support vector regression with asymmetric loss for optimal electric load forecasting. *Energy* **2021**, *223*, 119969. [CrossRef]
14. Mao, M.; Zhang, S.; Chang, L.; Hatziargyriou, N.D. Schedulable capacity forecasting for electric vehicles based on big data analysis. *J. Mod. Power Syst. Clean Energy* **2019**, *7*, 1651–1662. [CrossRef]
15. Nakabi, T.A.; Toivanen, P. An ANN-based model for learning individual customer behavior in response to electricity prices. *Sustain. Energy, Grids Networks* **2019**, *18*, 100212. [CrossRef]
16. Chen, Q.; Wang, H. Time-adaptive transient stability assessment based on gated recurrent unit. *Int. J. Electr. Power Energy Syst.* **2021**, *133*, 107156. [CrossRef]
17. Wu, H.; Liu, H. Non-intrusive load transient identification based on multivariate LSTM neural network and time series data augmentation. *Sustain. Energy Grids Netw.* **2021**, *27*, 100490. [CrossRef]
18. Li, Y.; Huang, Y.; Zhang, M. Short-Term Load Forecasting for Electric Vehicle Charging Station Based on Niche Immunity Lion Algorithm and Convolutional Neural Network. *Energies* **2018**, *11*, 1253. [CrossRef]
19. Chen, X.; Wu, W.; Gao, N.; Liu, J.; Chung, H.S.H.; Blaabjerg, F. Electric Vehicle Charging Load Forecasting: A Comparative Study of Deep Learning Approaches. *Energies* **2019**, *12*, 1–19.
20. Gerossier, A.; Girard, R.; Kariniotakis, G. Modeling and Forecasting Electric Vehicle Consumption Profiles. *Energies* **2019**, *12*, 1341. [CrossRef]
21. Time and Date AS 1995–2021. Time and Date AS. Available online: https://www.timeanddate.com/ (accessed on 12 September 2021).
22. TWC Product and Technology LLC 2014-2020. Weather Underground. Available online: http://api.wunderground.com/history/ (accessed on 26 October 2017).
23. Yu, Z.; Niu, Z.; Tang, W.; Wu, Q. Deep Learning for Daily Peak Load Forecasting–A Novel Gated Recurrent Neural Network Combining Dynamic Time Warping. *IEEE Access* **2019**, *7*, 17184–17194. [CrossRef]
24. Wang, S.; Wang, X.; Wang, S.; Wang, D. Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting. *Int. J. Electr. Power Energy Syst.* **2019**, *109*, 470–479. [CrossRef]
25. Ozcanli, A.K.; Baysal, M. Islanding detection in microgrid using deep learning based on 1D CNN and CNN-LSTM networks. *Sustain. Energy Grids Netw.* **2022**, *32*, 100839. [CrossRef]
26. Senesoulin, F.; Ngamroo, I.; Dechanupaprittha, S. Estimation of dominant power oscillation modes based on ConvLSTM approach using synchrophasor data and cross-validation technique. *Sustain. Energy Grids Netw.* **2022**, *31*, 100731. [CrossRef]
27. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [CrossRef] [PubMed]