

Research Article

Vijay Paidi*

Short-term prediction of parking availability in an open parking lot

<https://doi.org/10.1515/jisys-2022-0039>

received March 07, 2022; accepted April 01, 2022

Abstract: The parking of cars is a globally recognized problem, especially at locations where there is a high demand for empty parking spaces. Drivers tend to cruise additional distances while searching for empty parking spaces during peak hours leading to problems, such as pollution, congestion, and driver frustration. Providing short-term predictions of parking availability would facilitate the driver in making informed decisions and planning their arrival to be able to choose parking locations with higher availability. Therefore, the aim of this study is to provide short-term predictions of available parking spaces with a low volume of data. The open parking lot provides parking spaces free of charge and one such parking lot, located beside a shopping center, was selected for this study. Parking availability data for 21 days were collected where 19 days were used for training, while multiple periods of the remaining 2 days were used to test and evaluate the prediction methods. The test dataset consists of data from a weekday and a weekend. Based on the reviewed literature, three prediction methods suitable for short-term prediction were selected, namely, long short-term memory (LSTM), seasonal autoregressive integrated moving average with exogenous variables (SARIMAX), and the Ensemble-based method. The LSTM method is a deep learning-based method, while SARIMAX is a regression-based method, and the Ensemble method is based on decision trees and random forest to provide predictions. The performance of the three prediction methods with a low volume of data and the use of visitor trends data as an exogenous variable was evaluated. Based on the test prediction results, the LSTM and Ensemble-based methods provided better short-term predictions at multiple times on a weekday, while the Ensemble-based method provided better predictions over the weekend. However, the use of visitor trend data did not facilitate improving the predictions of SARIMAX and the Ensemble-based method, while it improved the LSTM prediction for the weekend.

Keywords: exogenous variables, LSTM, parking availability, prediction

1 Introduction

Parking of personal vehicles around areas of public interest is a well-known global problem, and it exacerbates when there are a smaller number of available empty parking spaces for many vehicles. The availability of empty parking spaces varies over space and time. This problem is often encountered at locations, such as airports and shopping centers, where there is high demand for empty parking spaces [1]. Lack of a sufficient number of empty parking spaces during peak hours leads to problems, such as congestion, pollution, excess cruising, and driver frustration [2,3]. Approximately 30% of congestion in urban areas was caused by drivers looking for empty parking spaces [4]. Population in urban areas is expected to increase by 12% by 2050, which exacerbates these parking-associated problems.

* **Corresponding author: Vijay Paidi**, Department of Computer and Information Management, School of Information and Engineering, Dalarna University Borlänge, Borlänge, Sweden, e-mail: vpp@du.se

Excess cruising and other parking-associated problems occur primarily due to a lack of information on parking availability. According to ref. [5], parking spaces are better utilized if occupancy information is available to the driver. Real-time parking availability can be displayed to the driver either using parking guidance systems or using smart parking applications. However, such systems are only available for closed parking lots where there is a return on investments. Closed parking lots provide paid parking spaces, while open parking lots provide free parking spaces for a limited duration of time. There are no similar applications available for open parking lots due to a lack of return on investments [6]. Providing real-time occupancy information for open parking lots would not be of much use as parking lots are available only for a short duration due to high demand [7]. However, providing short-term predictions of parking spaces would facilitate the driver in making informed decisions on their arrival at the parking lot [8]. Thus, this study aims to provide a short-term prediction of parking availability in an open parking lot using limited historical data. Disseminating short-term predictions of parking availability would help the driver to avoid peak times and choose parking areas with higher availability and reduce additional cruising. Reduction in cruising for empty parking spaces also reduces CO₂ and other harmful emissions [9]. There are not many previous studies providing short-term predictions of parking availability for open parking lots primarily due to a lack of return on investments. This article facilitates in addressing this research gap.

Historical parking availability data need to be collected to train forecasting methods. It is not economically viable to install underground sensors in open parking lots since sensors incur high maintenance costs [6]. An optical camera facilitates in capturing collective parking occupancy data. However, due to issues with lighting conditions and privacy, a thermal camera was utilized to collect data in an open parking lot. Generally, historical data spanning over months or years would be used to train forecasting methods. However, to evaluate short-term predictions with a low volume of data, parking availability data collected over 21 days were utilized in this study. The parking availability data were collected using a thermal camera in an open parking lot beside a shopping center.

Previous studies like ref. [10] focused on predicting parking spaces in commercial spaces or office spaces. Similarly, studies like refs. [7,11] evaluated prediction methods using a large volume of parking data collected from parking providers. However, there were not many studies predicting parking spaces in open parking lots. Parking availability differs from each location of interest, such as restaurants, shopping centers, residential areas, or office buildings. There can be several variables that facilitate understanding the parking availability, such as holidays, super sales, day of the week, and office hours. Therefore, previous studies utilize some of the relevant and available variables to improve the prediction performance of algorithms. In a study by Richter et al. [12], historical trends data were utilized to improve the accuracy of the forecasting model. However, to capture historical trends data, a large volume of data is necessary. Therefore, in this study, since there was a limited volume of data available, historical trends were categorized using visitor trends data from Google popular times. Regression-based machine learning algorithms are suitable for stationary data [13]. The collected parking availability data have a constant mean and variance over a period of time and are considered stationary. Therefore, a regression-based algorithm, namely Ensemble, was utilized in this study. Recurrent neural network-based algorithms, such as long short-term memory (LSTM), are deep learning-based forecasting methods suitable for short-term predictions, capable of learning complex patterns and therefore utilized in this study to predict parking availability. LSTM is suitable for forecasting stationary and non-stationary data. Similarly, seasonal autoregressive integrated moving average with exogenous variables (SARIMAX) is another suitable method for short-term predictions. It learns patterns from previous values considering seasonality and exogenous variables [14]. In this study, exogenous variables, such as time, day of the week, visitors trend from Google, holidays, and rain or snow, were utilized to facilitate forecasting. These are further discussed in Section 3.

The contribution of this article is as follows.

- The article provides short-term predictions using a limited volume of parking availability data collected on an open parking lot.
- This article evaluates the use of visitor trend data from Google, which is an open-source visitor trends data.
- Finally, prediction methods, such as LSTM, SARIMAX, and an Ensemble-based method, were evaluated to provide short-term predictions at multiple times of the day.

The remaining sections of the article are organized as follows. Section 2 discusses relevant literature on predicting parking availability. Section 3 focuses on the data collection and prediction algorithms. Section 4 presents results and discussion, while the article ends with a conclusion in Section 5.

2 Related work

The prediction of parking spaces in previous literature was performed using several methods, such as neural networks, multivariate spatiotemporal models, deep learning, and machine learning algorithms. This section discusses these methods and their suitability for this article.

In ref. [7], parking lot forecasting of on-street and off-street parking was performed using a multivariate model. The on-street and off-street parking lots are paid parking spaces that fall under the category of a closed parking lot. The data for the study were obtained from SFpark. The multi-variate model also utilized time variants, such as accidents and maintenance, in forecasting parking occupancy information. The model achieved a 95% accuracy for a 20 min forecast horizon. It performed well with off-street parking lots compared to on-street parking lots, which have more variance. The model needs several spatial and temporal variables to forecast parking occupancy information. Also, the model is suited for large cities with several parking lots within the vicinity. The data provided by SFPark consist of additional variables, which were not available for the open parking lot dataset collected in this article.

In ref. [15], a forecast of free occupancies was provided based on drivers' preferences, such as cost, walking distance, parking rules, duration, and availability. The preferred parking location was forecast based on the cost function. A forecast of parking occupancy was provided based on historical and real-time data. The cost is not available for open parking lots as parking spaces are provided free of charge. Therefore, the model proposed in ref. [15] is not suitable for this study.

Information on historical parking availability with 5 min intervals over 5 months was utilized in ref. [12]. The clustering method was used to reduce the storage of data in providing forecasting information. Due to this, an accuracy level of 70% was achieved in its predictions. The average values along with categories of availability were utilized for training the algorithm. Categories of availability were low, medium, and high. Utilizing average values facilitates improving accuracy, even during holidays. Therefore, this article also employs a similar approach, utilizing average occupancy values and availability categories.

Simple recurrent neural networks have problems with exploding and vanishing gradients, unlike LSTM's. The architecture of LSTM consists of a memory cell that overcomes the vanishing gradient problem and facilitates learning sequential data [16,17]. In ref. [13], a prediction of parking availability was performed using LSTM and clustering techniques. Parking occupancy and duration were provided as inputs for the algorithm. A large parking lot was divided into clusters with similar patterns and a recurrent neural network is used to generate forecasting metrics, which are fed into the parking model. The proposed model is not suitable for this study as only one region of a parking lot was selected, and the duration of the parking was not compiled. In ref. [17], a multi-step prediction approach using LSTM was utilized to predict long-term parking availability information. Large amounts of historical data were utilized to predict parking availability for 60 min. The predictions of LSTM performed better compared to regression-based models. Similarly, LSTM was utilized in this study as well, to provide predictions for 60 min but with data available for 3 weeks.

As discussed in ref. [14], the parking availability forecast for 6 months ahead was performed using SARIMAX and LSTM algorithms. SARIMAX performed better than LSTM, leading to lower RMSE values. The inclusion of exogenous variables resulted in reduced errors by up to 27%. Hence, SARIMAX was also utilized in this study to perform the forecasting of parking availability. According to another study [18], SARIMAX was utilized to predict cement demand. The data consist of 178 points, which were divided into train and test datasets. SARIMAX produced fewer errors in predicting cement demand. In refs. [14,18], a large volume of data was utilized to train the SARIMAX method. However, in this article, a limited volume of data was utilized that might impact the accuracy of predictions.

The Ensemble-based model is another popular method used for prediction and classification. In ref. [19], the Ensemble method was used to predict the availability of empty parking spaces, achieving a mean absolute error (MAE) value of 0.06 in its prediction. Bagging and boosting approaches were evaluated in this study. Similarly, in another study [20], an Ensemble approach combines multilayer perceptron, K-nearest neighbors, decision trees, and random forest to predict parking occupancy. A voting classifier was utilized to make decisions based on individual predictions. Less complex learners, such as decision trees, random forest, and KNN, performed better, compared to multilayer perceptron. Hence, the Ensemble-based method with the bagging approach is also utilized in this study to predict parking availability.

There were no studies that collected data from an open parking lot. Several previous studies utilized data collected from closed parking lots where sensors and other tools were used to obtain parking availability and duration information for larger periods, such as months or years. However, in this study, the data were collected manually using a thermal camera and, thus, the volume of data is limited. The relevant variables utilized in this article are open-source data and not collected using external sensors or sources. Popular prediction algorithms, such as LSTM, SARIMAX, and Ensemble methods, were suitable for prediction purposes and, thus, they were utilized and evaluated in this study.

3 Materials and methods

This section discusses the location of the parking lot and the method utilized to predict the availability of parking occupancy.

3.1 Case study

The parking lot is located beside a shopping center in a mid-sized city in Sweden. A fraction of that parking lot is the region of interest (ROI) in this study, and it is highlighted by a green color line, as illustrated in Figure 1. A thermal camera is installed in the shopping center and is utilized to collect videos from the ROI. It was possible to collect data from an open parking lot without heavy installation and maintenance costs due to the use of the thermal camera. The thermal camera also avoided privacy concerns that would have been caused by the use of an optical camera [21]. The data for the entire parking lot could not be considered due to the limited field view of the camera. The ROI has four entrances, E1, E2, E3, and E4, and since it is



Figure 1: Observed parking lot and the region of interest highlighted by the green line.

located near the entrance of the shopping center, it is expected to have higher vehicle traffic [22]. The thermal camera views the vehicle from the side instead of the rear or front side of the vehicle, which hinders the visibility of the vehicle. Therefore, the parking occupancy data were collected manually.

3.2 Data collection

There are previous studies that forecast parking occupancy information using months of historical data combined with real-time parking occupancy data. Such data can be collected for closed parking lots, where underground sensors can be utilized to collect long-term data. However, it is expensive to obtain such data for an open parking lot due to the lack of Parking Guidance and Information systems. Therefore, data collection was performed using videos from a thermal camera installed on the dome of the shopping center. Parking occupancy data were extracted manually using these videos. The number of parking spaces covered in this study is 47. The data for the forecast are collected over 21 days, between January 7 and January 27, 2020, as shown in Table 1. The first 19 days of data were used to train the model, while the remaining 2 days of data were used to evaluate the model. As mentioned in ref. [23], at least 5 days of hourly data are necessary to provide reasonable prediction values. Hence, the allocated training data size was sufficient to produce better prediction values. The parking spaces in an open parking lot are in higher demand and frequent changes within short intervals are expected. Therefore, the parking occupancy data were collected every 5 min. Since this article utilizes limited data for parking availability forecasting, it uses additional exogenous variables to improve the accuracy of forecasting. Exogenous variables as shown in Table 1 were utilized in this article. This consists of a total of six columns where the parking availability output variable is to be forecasted and the remaining are exogenous variables or predictors. Variables, such as level of traffic to the shopping center, day, holiday, and rain or snow, are exogenous as they are not dependent on the availability of parking spaces.

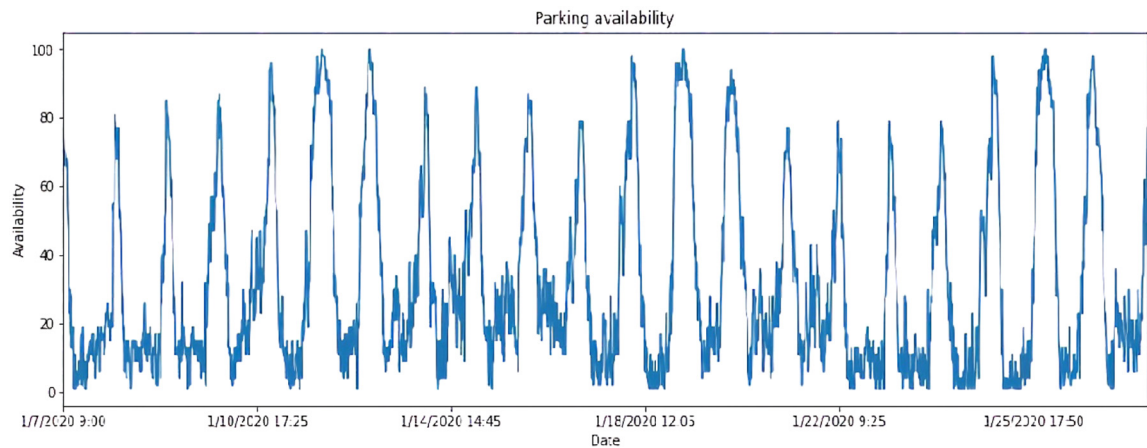
Table 1: Metadata of the dataset

Date	Day	Holiday	Date	Day	Holiday
January 07, 2020	Tuesday	No	January 18, 2020	Saturday	No
January 08, 2020	Wednesday	No	January 19, 2020	Sunday	No
January 09, 2020	Thursday	No	January 20, 2020	Monday	No
January 10, 2020	Friday	Yes	January 21, 2020	Tuesday	No
January 11, 2020	Saturday	No	January 22, 2020	Wednesday	No
January 12, 2020	Sunday	Yes	January 23, 2020	Thursday	No
January 13, 2020	Monday	Yes	January 24, 2020	Friday	No
January 14, 2020	Tuesday	No	January 25, 2020	Saturday	No
January 15, 2020	Wednesday	No	January 26, 2020	Sunday	No
January 16, 2020	Thursday	No	January 27, 2020	Monday	No
January 17, 2020	Friday	No			

- **Local time:** This column provides the timestamp of the collected data, where the date and time are displayed. The data, as shown in Table 2, are captured between 9:00 and 20:00, with a frequency of 5 min. Each day represents 133 data points.
- **Availability:** This column is the parking spot availability represented in percentage, i.e., $\text{Availability} = \frac{\max\{E_j\}}{47} \times 100$; $j \in w_i$, where E_j is the number of empty parking spots out of 47 available spots at any point in time j , when the change in the empty parking spots occurs within a specific 5 min, the observation window w_i , $i = 1, 2, n$. During peak periods, the availability becomes low, and during non-peak periods, it would be high. This is also the output or forecasted variable. The data on availability rate are illustrated through a time series plot in Figure 2.

Table 2: Example of time series input data for the model

Local Time	Availability (%)	Level	Day	Holiday	Rain or snow
2020-01-07 09:00	77	Low	Tuesday	No	No
2020-01-07 9:05	75	Low	Tuesday	No	No
2020-01-07 9:10	72	Low	Tuesday	No	No

**Figure 2:** Illustration of the parking availability of the total dataset.

- **Level:** This column is based on the visitor trend data from Google. However, the collected data are restricted to people using Google-based applications, such as Google Maps. Parking occupancy differs from each location of interest, such as restaurants, shopping centers, residential areas, or office buildings. Therefore, as mentioned in ref. [24], usage of parking profiles would facilitate improving the accuracy of the forecasting model. Popular times data are not available for certain periods, such as between 9:00 and 10:00, or between 19:00 and 20:00, due to few people visiting the shopping center. Since this only represents people using Google-based applications, these intervals are labeled as low. The periods where visitors were less than 60% are designated as medium. The periods near and around the peak, between 60 and 100%, are labeled as high. This variable is exogenous as visitor traffic may, or may not, affect parking availability. While visitor traffic is similar on weekdays, it varies on the weekend, as shown in Table 3.

Table 3: Hourly level data during weekday and weekend

Hour	Friday	Saturday
9	Low	Low
10	Medium	Medium
11	High	Medium
12	High	High
13	High	High
14	High	High
15	High	High
16	High	High
17	High	Medium
18	Medium	Low
19	Low	Low
20	Low	Low

- **Day:** This column represents days of the week. The use of this column facilitates forecasting parking availability based on the day of the week, and not on timestamp. As shown in Figure 3, the availability decreases by 12:00 during weekdays, while it increases by 13:00 during weekends. Availability during weekdays also sharply increases after 17:00 but gradually increases from 16:00 over the weekend.

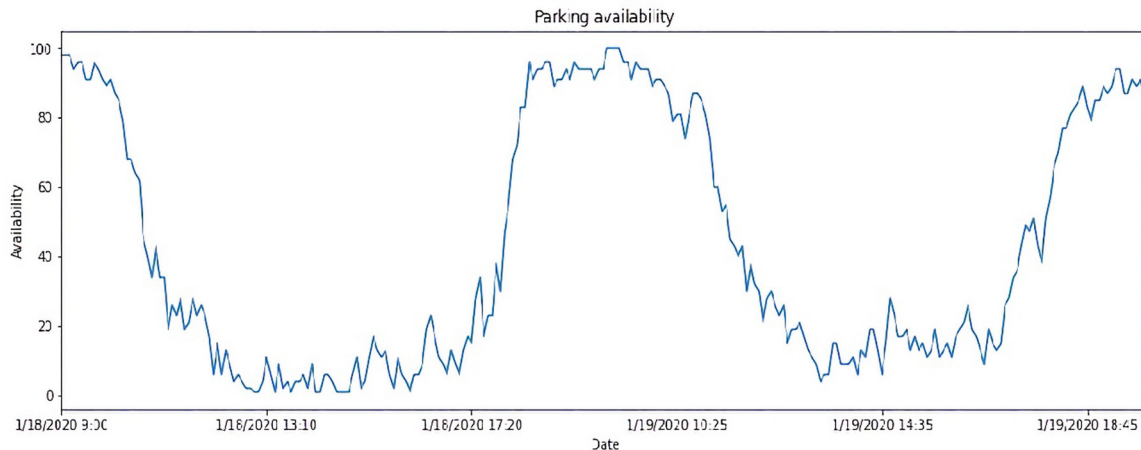


Figure 3: Parking availability during a weekday (January 18, 2020) and a weekend (January 19, 2020).

- **Holiday:** This column represents public holidays in Sweden. Parking occupancy varies on holidays. Therefore, it is necessary to consider this variable to forecast parking availability during holidays. However, traffic variation during all the holidays is not the same. Traffic during Christmas is high when compared to other public holidays. This variation is not observed in this study due to the limited dataset.
- **Rain or snow:** This column describes if there is rain or snow. Weather conditions usually impact travel behavior. However, in a country like Sweden, we expect this impact to be minimal unless there is rain or snowstorm.

Approximately 2,400 data points were utilized to train the selected algorithms, while 400 data points were utilized to evaluate the predictive performance of the algorithms. Training and testing were performed with and without visitor trend data to evaluate its effect on the performance of the algorithms. Two sets of prediction methods were trained, where one set includes visitor trend data during training and testing, while the other set does not include visitor trend data. Each set consists of LSTM, SARIMAX, and Ensemble-based methods. The exogenous variables, such as holiday, rain or snow, and day of the week, are included as they were found to improve the accuracy of the algorithms in several previous studies [25–28]. As parking availability varies between weekdays and weekends, the test dataset consists of data from a weekday and a weekend. Prediction methods were evaluated multiple times, such as 09:00, 13:00, and 18:00 during these days. Dell workstation with Quadro P5200 GPU was utilized in this study along with Matlab and Anaconda platforms to prepare data and evaluate prediction methods.

3.3 Algorithms for short-term prediction

Multiple algorithms, such as LSTM, SARIMAX, and Ensemble-based methods, are evaluated in this article with and without visitor trend data. The exogenous variables mentioned in Section 3.1 are the predictors, while the parking availability information is the response of the prediction methods.

3.3.1 LSTM

Recurrent neural networks are well known for prediction capability, as the algorithm understands the patterns in sequential data. LSTM is a recurrent-based neural network that consists of a memory cell and is ideal for prediction purposes [13]. This architecture consists of sequential, LSTM, dropout, and fully connected layers. Sequential data were fed into multiple LSTM and dropout layers. LSTM layer was used to learn patterns in sequential data, and the dropout layer was used to avoid overfitting [17]. Finally, the fully connected layers were utilized to produce the output. The batch size of the LSTM is 12, while the optimizer is stochastic gradient descent. Since the data have a limited volume, the number of epochs for training is 10.

3.3.2 SARIMAX

SARIMAX is a time series model incorporating seasonality and exogenous variables in forecasting. As ARIMA, it consists of an autoregressive polynomial of order p , moving average polynomial of order q , and non-seasonal difference d to produce stationarity. Due to the inclusion of seasonality, it consists of a seasonal autoregressive polynomial of order P , a seasonal moving average polynomial of order Q , and seasonal difference D , while the length of periodicity is given by m [18,29]. The seasonal and non-seasonal difference is zero as the data are stationary. In this study, the length of periodicity is 12, as each hour consists of 12 data points.

3.3.3 Ensemble-based model

This Ensemble method utilizes the bagging approach, which combines several decision trees, and random forests to individually provide predictions [19]. A weak learner, such as a decision tree, is taken over multiple selected samples and learns from them independently. Finally, it combines all the predictions based on deterministic average, and weights, to predict the output. The minimum leaf size is between 1 and 1,197, the number of learners is from 10 to 500, and the learning rate is 0.001.

3.4 Evaluation metrics

Parking availability predictions were provided for 60 min. The predicted results were evaluated using the following metrics, where y_i is the actual value at i the instance, while $\hat{y}_i$ is the predicted value at instance i .

MAE:

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

Root mean squared error (RMSE):

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

4 Results and discussion

The parking occupancy data collected from an open parking lot were evaluated using LSTM, SARIMAX, and Ensemble methods. The sample prediction values of the three methods for weekday and weekend at

multiple periods, namely 09:00, 13:00, and 18:00 are illustrated in this section. The predicted values are compared with the actual values for illustration purposes. These algorithms were trained and evaluated including and excluding visitor trend data.

4.1 Prediction with visitor trend data

Figure 4 illustrates prediction methods forecasting at multiple periods, namely 09:00, 13:00, and 18:00 during a weekend. The Ensemble-based method provided better short-term predictions than all other methods. As illustrated in Figure 4(b) and (c), the LSTM provided better predictions at 13:00 and 18:00, compared to SARIMAX. The sudden change in parking availability was learned better by Ensemble and the LSTM method. However, as shown in Figure 4(a) at 09:00, LSTM produced high errors. The SARIMAX method only produced better short-term predictions at 13:00, performing better short-term predictions when there is no sudden change in parking availability.

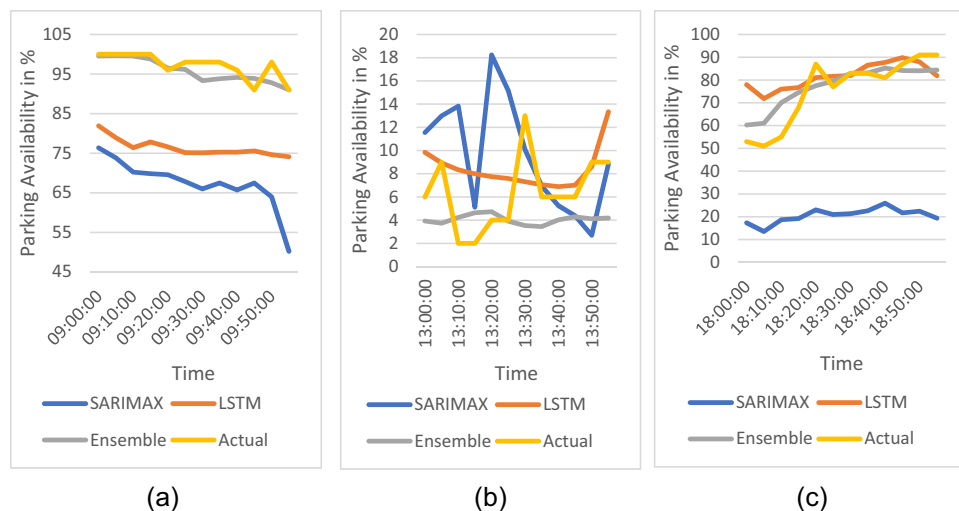


Figure 4: Prediction plot of methods with visitor trend data on the weekend. (a) Predictions with visitor trend data at 9 AM on weekend. (b) Predictions with visitor trend data at 1 PM on weekend. (c) Predictions with visitor trend data at 6 PM on weekend.

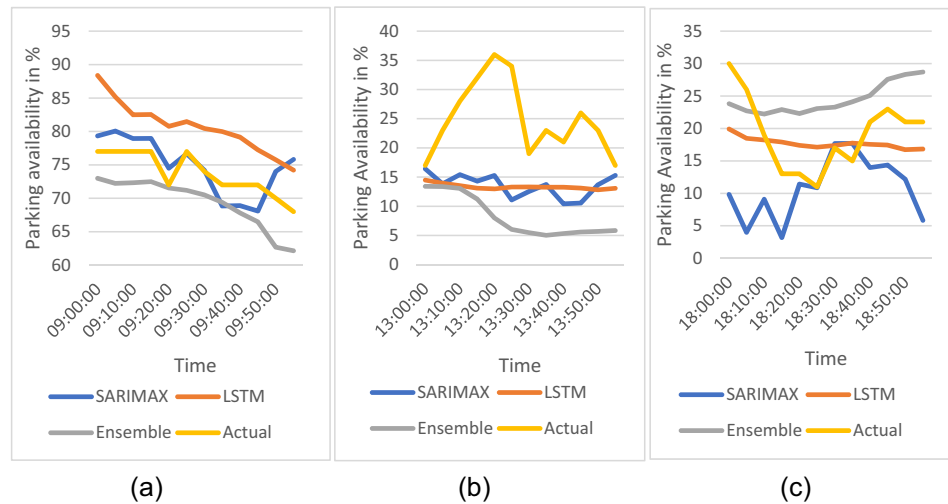
As shown in Table 4, the Ensemble-based method provided better predictions for multiple periods on a weekend with lower MAE and RMSE values. The total dataset consists of 3 weeks of data, where only 5 days representing a weekend were utilized for training the prediction methods. Despite the limited volume of data available for weekends, the Ensemble-based method produced better predictions. LSTM also performed better with fewer errors when compared to SARIMAX.

The overall performance of the three prediction methods was better for a weekday, as shown in Figure 5(a) and (c). However, when there is a sudden change in the rate of parking availability at 13:00 as shown in Figure 5(b), all three methods were unable to predict that change. Within 10 min, 10% of empty parking spaces were occupied. This random change in parking availability is expected for an open parking lot. However, due to limited data available for training and a lack of other potential exogenous variables, the sudden change in parking availability was not captured.

As shown in Table 5, SARIMAX produced lower MAE and RMSE values than other methods at 09:00 on a weekday. However, all three methods produced high errors at 13:00. LSTM produced fewer errors in predicting at 18:00. No method produced lower errors for all the mentioned periods. Ensemble-based method predicts values based on aggregation and there are not enough historical data to predict the change

Table 4: Evaluation of methods with visitor trend data on weekend

Method	Weekend at 09:00		Weekend at 13:00		Weekend at 18:00	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
LSTM	20.77	20.93	3.08	3.79	9.37	12.22
SARIMAX	29.79	30.13	5.22	6.93	55.1	56.44
Ensemble	1.97	2.63	3.2	4	5.9	7.2

**Figure 5:** Prediction of parking availability with visitor trend data on weekday. (a) Predictions with visitor trend data at 9 AM on a weekday. (b) Predictions with visitor trend data at 1 PM on a weekday. (c) Parking predictions with visitor trend data at 6 PM on a weekday.**Table 5:** Evaluation of methods with visitor trend data on a weekday

Method	Weekday at 09:00		Weekday at 13:00		Weekday at 18:00	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
LSTM	6.88	7.12	11.58	13.19	4.53	5.21
SARIMAX	2.86	3.42	11.37	13.16	8.89	11.28
Ensemble	4.43	4.74	16.75	18	6.92	7.43

in parking availability during peak hours. LSTM would have produced better predictions with more training data.

4.2 Prediction without visitor trend data

In this section, the prediction methods were evaluated without the use of visitor trend data for multiple periods mentioned in Section 4.1.

Figure 6 illustrates the prediction values without the use of visitor trend data on a weekend. As illustrated in Figure 6(a), the Ensemble-based method produced better prediction results at 09:00, and LSTM and SARIMAX produce predictions with high errors. As shown in Figure 6(b), all three methods provided predictions with lower errors. However, at 18:00, as illustrated in Figure 6(c), SARIMAX

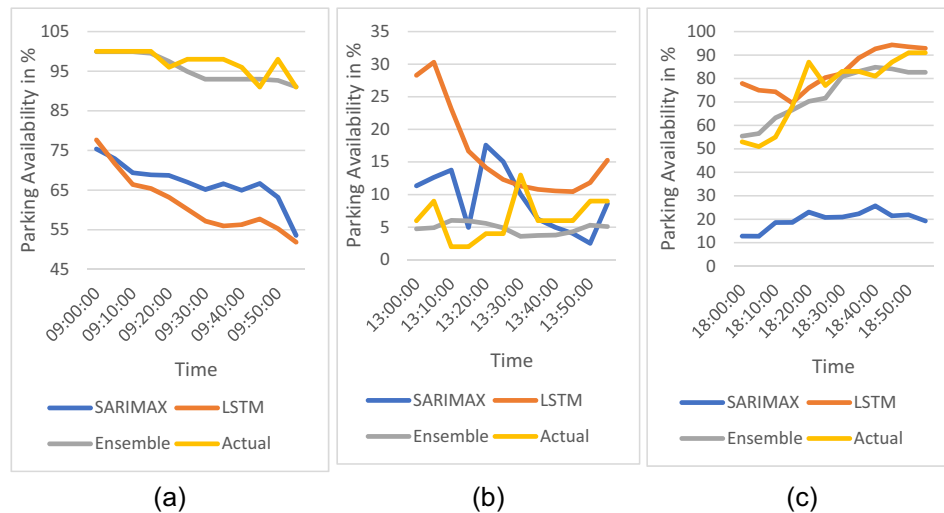


Figure 6: Prediction plot of methods without visitor trend data for the weekend. (a) Predictions without visitor trend data at 9 AM on weekend. (b) Predictions without visitor trend data at 1 PM on weekend. (c) Predictions without visitor trend data at 6 PM on weekend.

predictions produced high errors. This performance is similar to the performance with visitor trend data. Therefore, there is no change in the performance of SARIMAX and Ensemble without the visitor trend data.

As shown in Table 6, the LSTM performed slightly better with visitor trend data when compared to without visitor trend data. The MAE and RMSE values were higher for LSTM without the use of visitor trend data. However, SARIMAX and Ensemble-based methods provided similar MAE and RMSE values with and without visitor trend data.

Table 6: Evaluation of algorithms without visitor trend data on weekend

Method	Weekend at 09:00		Weekend at 13:00		Weekend at 18:00	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
LSTM	35.6	36	10.2	12.58	9.53	12.71
SARIMAX	30.3	30.53	5.12	6.78	55.76	56.95
Ensemble	2.13	2.92	3.2	3.9	5.44	6.9

As illustrated in Figure 7, the predictions of the Ensemble-based method and SARIMAX were similar to visitor trend data. Predictions of LSTM were the only predictions affected when excluding the visitor trend data. Contrary to the performance of LSTM on a weekend, the predictions slightly improved over a weekday without the use of visitor trend data, as shown in Figure 7(a) and (c). The variations in parking availability over a weekday were better learned by the LSTM without the use of visitor trend data. As illustrated in Figure 7(c), at 13:00, the sudden change in parking availability was not captured by any prediction methods without visitor trend data as well. As shown in Figure 7(c), at 18:00, LSTM performed better, compared to other methods. SARIMAX still maintained lower parking availability at 18:00, which is the reason for larger errors.

As shown in Table 7, LSTM provided better predictions with lower MAE and RMSE values compared to the Ensemble-based method and SARIMAX. The LSTM predictions were slightly improved for multiple periods on a weekday when compared to predictions with visitor trend data. For methods, such as SARIMAX and the Ensemble-based method, predictions had no impact with or without visitor trend data.

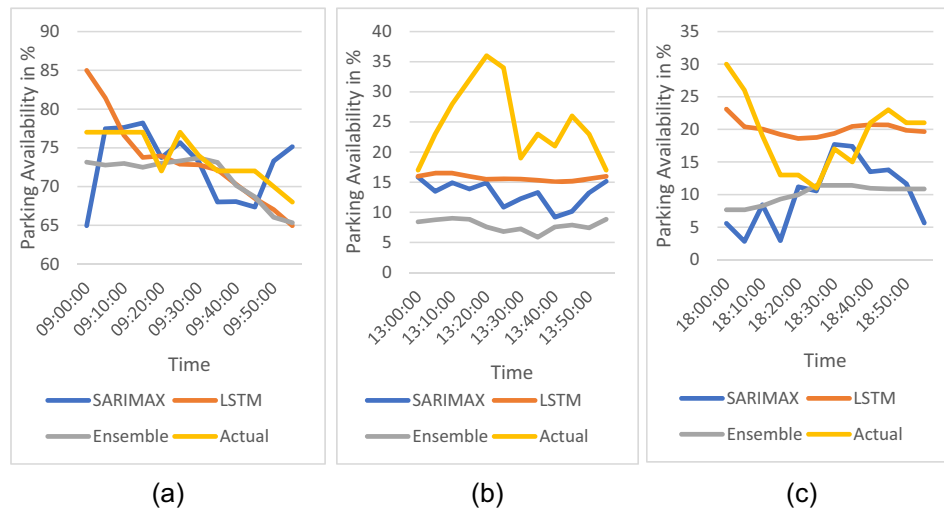


Figure 7: Prediction plot of methods without visitor trend data for weekday. (a) Predictions without visitor trend data at 9 AM on a weekday. (b) Predictions without visitor trend data at 1 PM on a weekday. (c) Predictions without visitor trend data at 6 PM on a weekday.

Table 7: Evaluation of algorithms without visitor trend data on a weekday

Method	Weekday at 09:00		Weekday at 13:00		Weekday at 18:00	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
LSTM	2.9	3.54	9.2	11	3.84	4.6
SARIMAX	3.44	4.73	11.82	13.55	9.59	12.32
Ensemble	2.88	3.1	17	18	9.1	11

Despite the less volume of training data, the Ensemble-based method and LSTM learned the patterns more efficiently similar to results observed in refs. [17,19]. Since there is a similarity in parking availability over weekends, the Ensemble-based method performed well, despite there being less volume of data. The Ensemble-based method utilizes multiple samples of the training dataset to generate aggregated values from decision trees and random forest, which is the reason for the improved prediction, but when there are quick changes in parking availability, which are not based on previous data, then the prediction methods would produce high errors in their predictions. However, this can be addressed by adding relevant exogenous or endogenous variables that affect this behavior. The LSTM and SARIMAX methods' predictions are based on previous values. However, due to the usage of a memory cell in LSTM, it understood the time series data efficiently compared to SARIMAX. SARIMAX produced better prediction in ref. [14]. However, in this study, due to less volume of data, SARIMAX produced high errors when there are sudden changes in the rate of parking availability. By increasing the volume of the dataset, the performance of the methods can further be improved. Another reason for the poor performance of SARIMAX is due to the lack of parking time information.

5 Conclusion, limitations, and future work

In this article, short-term predictions of parking availability with a low volume of data were evaluated using LSTM, SARIMAX, and Ensemble-based methods. Despite being trained with a low volume of data, the Ensemble-based method and LSTM produced short-term predictions with lower errors compared to

SARIMAX. Abrupt changes in parking availability were not captured by any evaluated prediction methods. The use of visitor trend data only improved LSTM performance on a weekend, while SARIMAX and Ensemble-based methods were not affected by it. The results from this study are also applicable to other similar regions of open parking lots that are near the location of interest.

The regions of the open parking lot, which are near the location of interest, are in high demand and usage of these prediction methods facilitated providing a short-term prediction of parking availability. If the orientation of the camera is pointed toward the front or rear end of the vehicles, detection algorithms can be utilized to collect vehicle occupancy data. The ROI is near the shopping center, which is in high demand. However, there are other regions in the parking lot, which were not covered due to the limited view of the thermal camera.

Simulated data can be utilized to further improve the performance of prediction. The training dataset contains occupancy data from weekdays, weekends, holidays, and rain or snow conditions. There can be other scenarios, such as special events, and severe weather conditions, which were not covered in this study. Adding relevant predictors can facilitate in further improving the accuracy of predictions. Duration of parking can also be added as a variable to improve the performance of algorithms. The parking lot can be divided into clusters to provide predictions as the rate of parking occupancy with these clusters can vary based on their proximity to the location of interest.

Conflict of interest: Author states no conflict of interest.

References

- [1] Tiedemann T, Vögele T, Krell MM, Metzen JH, Kirchner F. Concept of a data thread based parking space occupancy prediction in a berlin pilot region. In: Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence; 2015.
- [2] Paidi V, Håkansson J, Fleyeh H, Nyberg RG. CO₂ emissions induced by vehicles cruising for empty parking spaces in an open parking lot. *Sustainability*. 2022;14(7):3742.
- [3] Song J, Zhao C, Lin T, Li X, Prishchepov AV. Spatio-temporal patterns of traffic-related air pollutant emissions in different urban functional zones estimated by real-time video and deep learning technique. *J Clean Prod*. 2019;238:117881.
- [4] Shoup D. Free parking or free markets, in parking and the city. England, UK: Routledge; 2018. p. 270–5.
- [5] Van Der Waerden P, Timmermans H, Barzeele P. Car drivers' preferences regarding location and contents of parking guidance systems: stated choice approach. *Transp Res Record*. 2011;2245(1):63–9.
- [6] Paidi V, Fleyeh H, Håkansson J, Nyberg RG. Smart parking sensors, technologies and applications for open parking lots: a review. *IET Intell Transp Syst*. 2018;12(8):735–41.
- [7] Rajabioun T, Ioannou PA. On-street and off-street parking availability prediction using multivariate spatiotemporal models. *IEEE Trans Intell Transp Syst*. 2015;16(5):2913–24.
- [8] Caicedo F, Blazquez C, Miranda P. Prediction of parking space availability in real time. *Exp Syst Appl*. 2012;39(8):7281–90.
- [9] Caicedo F. Real-time parking information management to reduce search time, vehicle displacement and emissions. *Transp Res Part D Transp Environ*. 2010;15(4):228–34.
- [10] Zhao Z, Zhang Y. A comparative study of parking occupancy prediction methods considering parking type and parking scale. *J Adv Transp*. 2020;1620:2020.
- [11] Pozo RF, Gonzalez ABR, Wilby MR, Diaz JJV, Matesanz MV. Prediction of on-street parking level of service based on random undersampling decision trees. *IEEE Trans Intell Transp Syst*. 2021;1–10.
- [12] Richter F, Di Martino S, Mattfeld DC. Temporal and spatial clustering for a parking prediction service. In: 2014 IEEE 26th International Conference on Tools with Artificial Intelligence. Limassol, Cyprus: IEEE; 2014.
- [13] Shao W, Zhang Y, Guo B, Qin K, Chan J, Salim FD. Parking availability prediction with long short term memory model. In: International Conference on Green, Pervasive, and Cloud Computing. Springer; 2018.
- [14] Fokker ES, Koch T, van Leeuwen M, Dugundji ER. Short-term forecasting of off-street parking occupancy. *Transp Res Record*. 2022;2676(1):637–54.
- [15] Rajabioun T, Foster B, Ioannou P. Intelligent parking assist. In: 21st Mediterranean Conference on Control and Automation. Platanias, Greece: IEEE; 2013.
- [16] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Compu*. 1997;9(8):1735–80.
- [17] Fan J, Hu Q, Xu Y, Tang Z. Predicting the vacant parking space availability: an LSTM approach. In: IEEE Intelligent Transportation Systems Magazine; 2020.

- [18] Fişkın CS, Cerit AG. Forecasting domestic shipping demand of cement: comparison of SARIMAX, ANN and hybrid SARIMAX-ANN. In: 2019 4th International Conference on Computer Science and Engineering (UBMK). Samsun, Turkey: IEEE; 2019.
- [19] Tekouabou SCK, Cherif W, Silkan H. Improving parking availability prediction in smart cities with IoT and ensemble-based model. *J King Saud Univ Comput Inf Sci*. 2020;34(3):687–97.
- [20] Awan FM, Saleem Y, Minerva R, Crespi N. A comparative analysis of machine/deep learning models for parking space availability prediction. *Sensors*. 2020;20(1):322.
- [21] Asghar MN, Kanwal N, Lee B, Fleury M, Herbst M, Qiao Y. Visual surveillance within the EU general data protection regulation: a technology perspective. *IEEE Access*. 2019;7:111709–26.
- [22] Paidi V, Håkansson J, Fleyeh H, Nyberg RG. CO₂ emissions induced by vehicles cruising for empty parking spaces in an open parking lot. *Sustainability*. 2021;14(7):3742,1–18.
- [23] Balmer M, Weibel R, Huang H. Value of incorporating geospatial information into the prediction of on-street parking occupancy – a case study. *Geo Spat Inf Sci*. 2021;24(3):438–57.
- [24] Ionita A, Pomp A, Cochez M, Meisen T, Decker S. Where to park? Predicting free parking spots in unmonitored city areas. In: *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*; 2018.
- [25] Basu P, Little TD. Wireless ad hoc discovery of parking meters. In: *MobiSys Workshop on Applications of Mobile Embedded Systems (WAMES'04)*; 2004.
- [26] Ghosal SS, Bani A, Amrouss A, El Hallaoui I. A deep learning approach to predict parking occupancy using cluster augmented learning method. In: *2019 International Conference on Data Mining Workshops (ICDMW)*. Beijing, China: IEEE; 2019.
- [27] Provoost JC, Kamilaris A, Wismans LJJ, van der Drift SJ, van Keulen M. Predicting parking occupancy via machine learning in the web of things. *Internet Things*. 2020;12:100301.
- [28] Fabusuyi T, Hampshire RC, Hill VA, Sasanuma K. Decision analytics for parking availability in downtown Pittsburgh. *Interfaces*. 2014;44(3):286–99.
- [29] Cools M, Moons E, Wets G. Investigating the variability in daily traffic counts through use of ARIMAX and SARIMAX models: assessing the effect of holidays on two site locations. *Transp Res Record*. 2009;2136(1):57–66.