



# Predictions of Taxi Demand Based on Neural Network Algorithms

Chung-Yi Lin<sup>1,2</sup> · Shen-Lung Tung<sup>1,3</sup> · Po-Wen Lu<sup>1,4</sup> · Tzu-Cheng Liu<sup>1,5</sup>

Received: 18 December 2019 / Revised: 16 July 2020 / Accepted: 16 December 2020 / Published online: 24 July 2021  
© Springer Science+Business Media, LLC, part of Springer Nature 2021

## Abstract

To increase the profit both of taxi drivers and operators, this paper proposes an approach that efficiently collects the features of a customized-shape dispatch area to build the multivariate time-series prediction models for forecasting taxi demands. We also considered population distribution obtained from IMSI (International Mobile Subscriber Identity) data as the spatial correlations feature. The predictive models are built on some neural network algorithms and analyzed statistically. The experiments show that the predictions of the taxi demand in the next 30 minutes are successfully achieved. It is noteworthy that our approach outperforms the forecasting accuracy proved by a real-world error metric.

**Keywords** Taxi demand prediction · Taxi dispatch area · Neural networks

## 1 Introduction

Artificial intelligence technologies develop vigorously in recent years and they are powerful and suitable tools to forecast trends on interested data. In ITS (Intelligent Transportation System) industry, related resources can be allocated, if the trends of the road traffic, like peak hours, traffic amount and periodicity, are identified. Taking the taxi business as an example, the operator of a taxi fleet can dispatch its drivers with an schedule regulated according to the identified trends. Therefore, the fleet can be maneuvered efficiently, where both drivers' and passengers' waiting time are reduced and the revenue are increased [1]. To achieve precise predictions, a taxi-demand forecast must be considered in many different features, from the short-term factors (such as accidents and

activities) to the long-term factors (such as traffic in rush hour, weather information, etc.) [2].

The traffic forecast is a time-series data prediction problem, where the amount of taxi demand inside a region was forecast by using historical taxi demand data [3]. Traditional methods were mostly focused on linear models, and they performed well under normal conditions; however, their predictions were not good enough under some extreme conditions. A recent paper [4] further considered the background data (such as locations, weather, and events), but they still could not capture the complex nonlinear spatial relation and the temporal sequential relation. Some researches [5, 6] pointed out that the prediction of taxi demand was still a great challenge affected by multiple types of dependencies, such as the characteristics of locations. Inspired by previous works, we analyze the taxi demand in dispatch areas with different location characteristics:

### 1) Geographical area dependence:

The taxi demand is highly related to locations. For example, in Fig. 1, the number of taxi demand in the area around MRT (Mass Rapid Transit) station is much higher than that in the hospital area.

### 2) Time period dependence:

In general, the road traffic of a certain place has a fixed pattern, where peaks occur at a certain timing periodically.

✉ Chung-Yi Lin  
gn027759681@cht.com.tw

<sup>1</sup> Chunghwa Telecom Laboratories, Ltd., No.99, Dianyan Rd., Yangmei District, Taoyuan City 326402, Taiwan

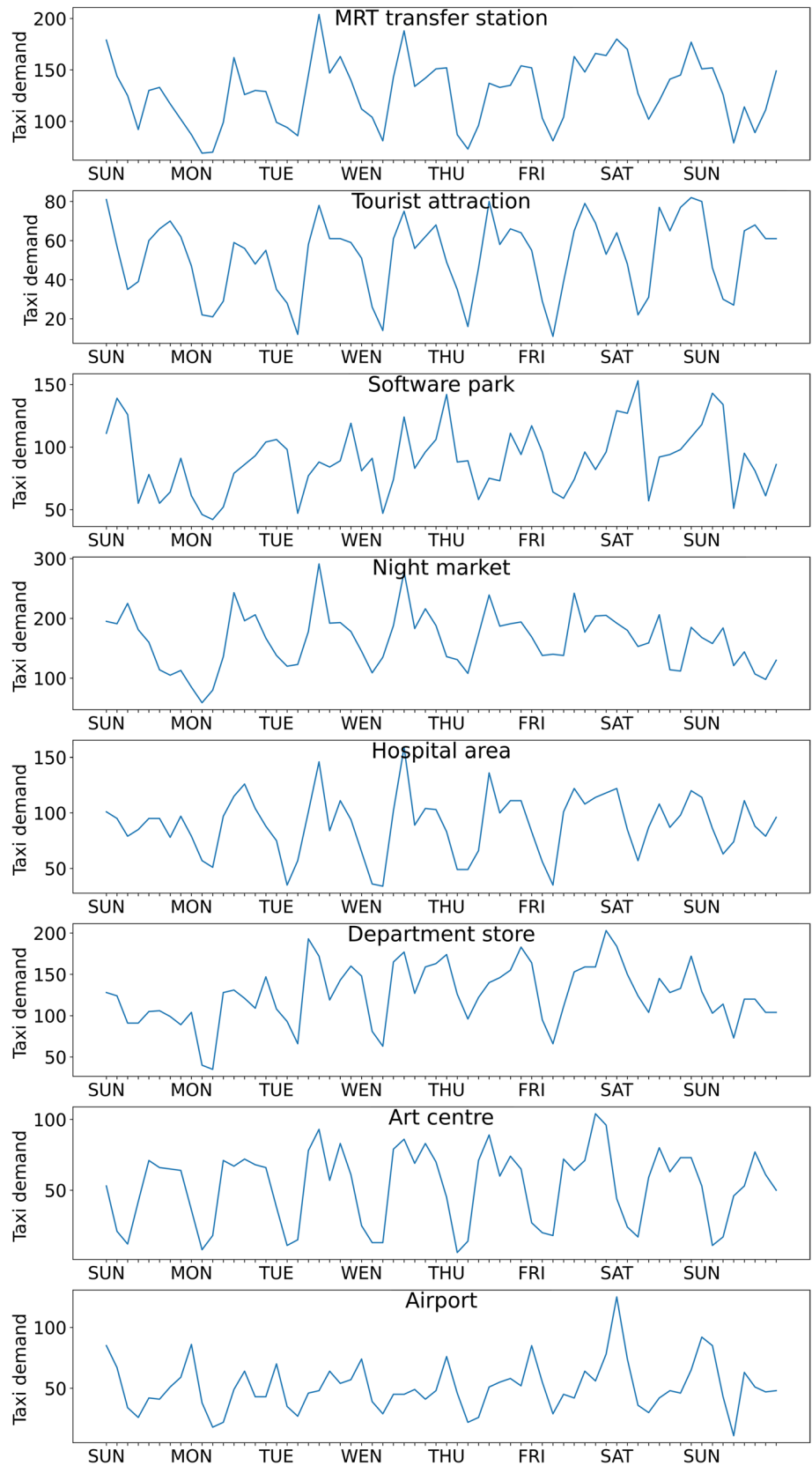
<sup>2</sup> Graduate Institute of Networking and Multimedia, National Taiwan University, Taipei, Taiwan

<sup>3</sup> Department of Electrical Engineering, National Central University, Taoyuan City, Taiwan

<sup>4</sup> Department of Power Mechanical Engineering, National Tsing Hua University, Hsinchu, Taiwan

<sup>5</sup> Department of Information Management, National Taiwan University of Science and Technology, Taipei, Taiwan

**Fig. 1** Taxi demand patterns in various geographical positions



Similarly, the time period dependence takes place in the taxi demand trend, as depicted in Fig. 1. Though the numbers of taxi demand vary around the clock, they show strong regular patterns for weekdays and weekends.

Because the functionality of the district dominates the people's activities taking place on it, both factors will affect the results of prediction enormously.

Deep learning can not only learn automatically from various data, but perform well in sequential prediction based on data with drastic changes. De Brébisson *et al.* adopted RNN (Recurrent Neural Network) for taxi destination prediction based on the GPS trajectory of the taxi, and the results showed the outstanding prediction performance of RNN than linear models [7]. Other special forms of RNN: LSTM (Long Short Term Memory networks) [8] and GRU (Gated Recurrent Unit) [9] were widely used. While containing more gating mechanisms to properly preserve the correlation in the sequence and learning long-term dependencies, these two approaches have achieved great success in learning traffic patterns through capturing the sequential dependency for the taxi demand service forecasting.

Although CNN (Convolutional Neural Network) is commonly used for image processing tasks, it was adopted in some related researches [10, 11]. In our experiments, the training convergence of CNN is faster than RNN more than 5 times, therefore we can utilize the CNN model to initially verify the impact of input features and improve the feature selection. Among those existing methods on traffic prediction, while some researches considered spatial relation (e.g., using CNN) or temporal relation (e.g., using LSTM) alone, some researches [12] considered both characteristics and proposed the CNN with a long short-term memory network, named “CNN-LSTM” herein for short. This architecture uses CNN as a front-end layer for feature extraction on input data and passes the features information to LSTM for sequence learning and prediction.

While existed methods made some great progress in forecasting the taxi demand, they missed some important features to make the forecast better. This research will take them into account. First, a customized-shape dispatch area. In reality, the shape of a dispatch area is designed to cover the most population around a certain spot such as an MRT station, and a customized shape does a better job. For instance, taxi operators in Kaohsiung City, Taiwan, may build dispatch areas along the crowd gathering spots to match the rides efficiently. Because the crowd population is not distributed uniformly, the range of dispatch area may be defined as customized-shape, such as polygons, to cover the concerned area, where passengers' demands for taxi services can be satisfied in that area within a reasonable waiting time. By Chunghwa Telecom's Taxi Dispatching System, the taxi companies can define their own dispatch areas by themselves. For instance, in the neighborhood of Formosa

Boulevard of Kaohsiung, three areas are circled for taxi dispatch areas as depicted in Fig. 2, where Area 1 is a tourist attraction area of Pier-2 Center, Area 2 is for Liuhe Night Market, and Area 3 is Sanduo Shopping District.

Second, the locations of vehicles. Since the time, how long a passenger will wait for a taxi, is an important factor, the location of the taxi is highly related and is provided by the global positioning systems (GPS) in this paper. Third, the weather condition. While in bad weather, the population may consider a more convenient way to leave. Therefore, the weather condition could be an incentive factor. Last, the population information, based on cell phones in the demand area is also considered. The International Mobile Subscriber Identity (IMSI), a number to uniquely identify a user in the cellular network, is used in this paper. As the largest telecom company in Taiwan, Chunghwa Telecom has about 11 million mobile users, which stands for over 1/3 of the population in Taiwan. Therefore, the collected IMSI data are representative of the population distribution in the Kaohsiung area. In this paper, we only consider the IMSI located outdoors related to the demand for taxis, and the IMSI data are partitioned by grid-cells, where the grid-cell size is designed as 500 m × 500 m and we partitioned the Kaohsiung into equal regions of 500 m × 500 m to follow with the geographical property.

The rest of the paper is organized as follows. In Section 2, we will introduce how to combine scattered GPS information of passengers' pick-up event, and weather information into the corresponding dispatch area data. In Section 3, we briefly explain how to build predictive models with multivariate and multi-step time series, and then forecast the taxi demand for the next 30 min. In this paper, the time-step of demand prediction is set as 30 minutes according to the experience of the taxi administrator. The reason we have discussed that 30 minutes is available for the response time of the taxi center and drivers. In Section 4, we show the experimental setup and forecasting results with different regions and input features. Lastly, a summary is given in Section 5.

## 2 Data Preparation of Dispatch Areas

The trend of taxi demands in downtown Kaohsiung during 2019 is proposed in this paper. The data is mainly obtained from Chunghwa Telecom and the Central Weather Bureau of Taiwan, where personal or client-related information is excluded from the data. The taxi demand dataset is obtained from Chunghwa Telecom Taxi Dispatching System and comprises timestamps of passenger demand and actual GPS data of passengers' pick-up points, where the dataset is real-world data generated by about 1,700 taxis. Information on the weather is included in the paper since weather conditions can affect the taxi demand. We obtain



**Fig. 2** Example of dispatch areas in Kaohsiung based on Chunghwa Telecom Taxi Dispatching System

the weather information of Kaohsiung from the open data website of the Central Weather Bureau, Taiwan. The weather conditions with an hourly sampling rate are provided on the website. The weather conditions include rain, temperature, clouds, and humidity. We select hourly rain and temperature. Besides, the outdoor IMSI volume in the dispatch area is also evaluated. According to our data, the higher the number of outdoor IMSI in the dispatch area, the higher is the taxi demand. The IMSI volume in this paper is obtained from the IMSI which is a signaling data of mobile collected at cell sites constructed by Chunghwa Telecom Company.

### 2.1 Forecasting the Taxi Demand in Dispatch Areas

The dataset contains both the GPS location and the time-stamp for each taxi demand event. To analyze the trend of taxi demand by using time steps and dispatch areas, the taxi demand for each time step and dispatch area must be summarized. Because a dispatch area is often a customized polygon, we reconstruct a polygon dispatch area into a mosaic area comprising multiple smaller square grid regions. The approach is proposed as follows.

Step 1. Partition the city map to equal regions:

The map of Kaohsiung is divided into many equal regions based on the approach proposed in [13, 14]. The region size is set as  $500\text{ m} \times 500\text{ m}$ . This process allows us to paper various relative geographical positions.

Step 2. Map the data to partition regions:

The location of a taxi demand event is distributed to a partition region if the GPS information of the demand event is within the range of that partition region.

Step 3. Count the taxi demand at each time step:

To construct time-series data, half an hour is assumed as one time step, that is, 48 pieces of time-step data are collected in a day.

Step 4. Group partition regions into the dispatch area:



Through steps 1–3, Kaohsiung is partitioned into multiple regions, and each partition region had its taxi demand in time steps. Fig. 3 displays the downtown area of Kaohsiung. The map containing three dispatch areas is partitioned into equal regions of 500 m × 500 m.

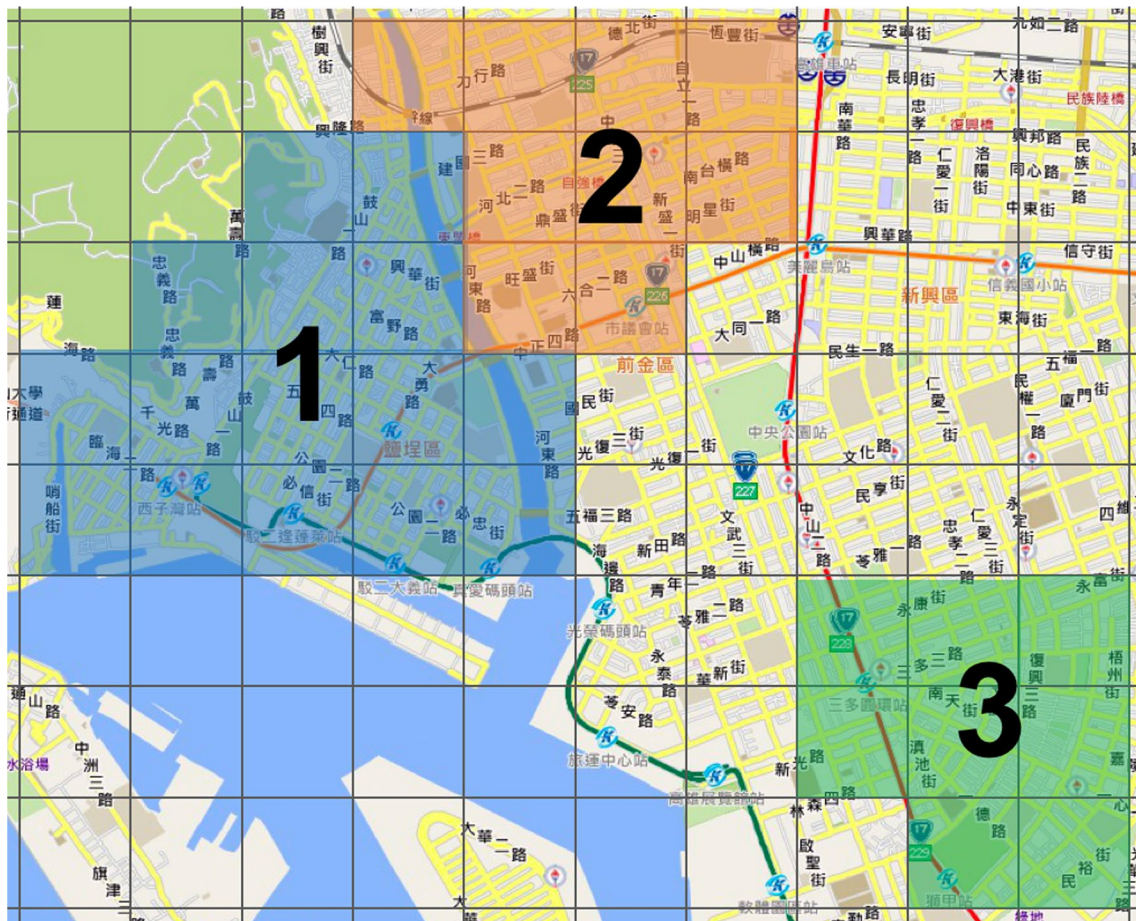
In this manner, we group the partition regions. According to the taxi dispatching system, the range of each dispatch area are predefined and each partition region is tagged and mapped to one dispatch area. When a region partly or fully overlaps with a dispatch area, the region belongs to the dispatch area. For each time step, the demands on all regions of the same dispatch area group are summarized.

## 2.2 Dispatch Area Selection

Because each dispatch area has its taxi demand pattern, we develop models according to various regional characteristics to achieve accurate predictions. We select the following dispatch areas as targets to evaluate the model performance:

- 1) Night market
- 2) Mass rapid transport transfer station (MRT transfer station)
- 3) Tourist attraction
- 4) Software park
- 5) Hospital area
- 6) Department store
- 7) Art centre

Among the dispatch areas, Liuhe Night Market is a tourist attraction and opens every day. Although night market culture is a special lifestyle in Taiwan, the opening days for different night markets vary from place to place. Unlike common restaurants or food courts, the peak time of the night market usually occurs at night. The crowds attracted to this lifestyle have generated a unique trend of taxi demand.



**Fig. 3** Through steps 1–3, Kaohsiung is partitioned into multiple regions. Then we group the partition regions based on the predefined range of each dispatch area

### 3 Taxi Demand Forecasting Model

#### 3.1 Forecasting Model Building

In this section, we briefly explain the algorithm used in the paper, namely the long short-term memory (LSTM), gated recurrent unit (GRU), convolutional neural network (CNN) and CNN-LSTM algorithms.

##### 3.1.1 LSTM

LSTM, which is a type of recurrent neural network, is commonly used for time-series prediction [8]. Because of the gating mechanism, the LSTM algorithm has achieved considerable success in sequential prediction and is effective in learning traffic patterns [2, 6].

A typical LSTM structure is presented in Fig. 4. In the figure,  $\sigma$  is the sigmoid function,  $h_t$  is the hidden state,  $C_t$  is the cell state at each time step  $t$ . The forget gate considers both the hidden state  $h_{t-1}$  and input  $X_t$  depending on the information from the current cell state. The input gate decides what information is going to be updated and stored in the cell state. The output gate refers to the cell state and then decides what should be sent to the output.

We refer to the concept of data modeling proposed by Uber laboratory [15] and extend for the multivariate and multi-step time series forecasting. Before the LSTM model can be used for time series forecasting, the problems must be re-framed as supervised learning problems, that is, time-series data must be converted to supervised learning format. After the original data are converted to the time-series data of dispatch areas, we re-construct the supervised learning problem as predicting the demand in dispatch areas at the future time step ( $t+1$ ) given the history features such as

taxi demand data, weather conditions, etc, from the prior time step ( $t-n$ ) to time step  $t$ , where the input time window length is  $n$ . We create the training input ( $X$ ) from the prior time steps and output ( $Y$ ) with different sliding window lengths, that is, the LSTM model predicts the data of the next time step according to a certain range of historical data, as depicted in Fig. 5. We explore the impacting of the input time window length on the model through the experimental section.

##### 3.1.2 GRU

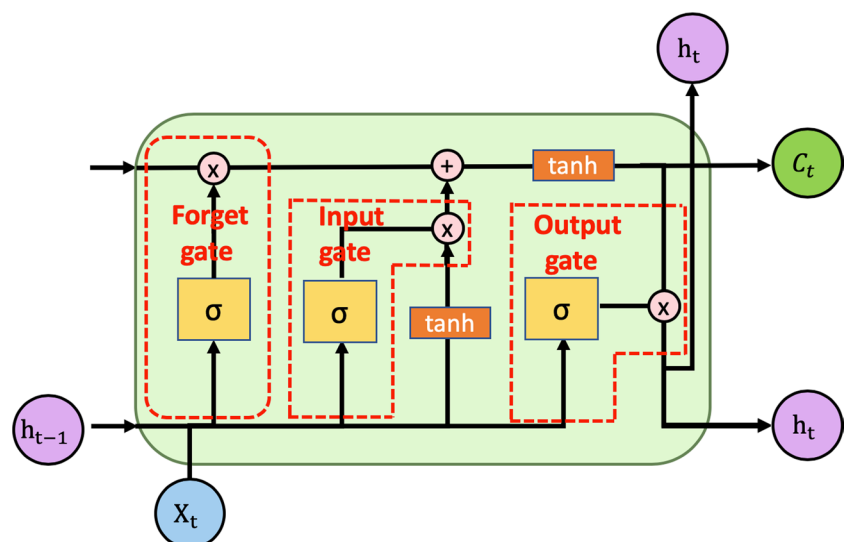
The GRU algorithm, which was introduced by Cho et al. [9], can solve the vanishing gradient problem associated with a standard recurrent neural network. In [16], a clear distinction was drawn between the GRU and LSTM algorithms. The GRU algorithm can be considered as a variation of the LSTM algorithm because both are designed similarly and, in some cases, produce equally excellent results. We select the GRU algorithm to be another forecasting model in this paper. We design our GRU model for multivariate and multi-step time series forecasting. The architecture of the model is illustrated in Fig. 5.

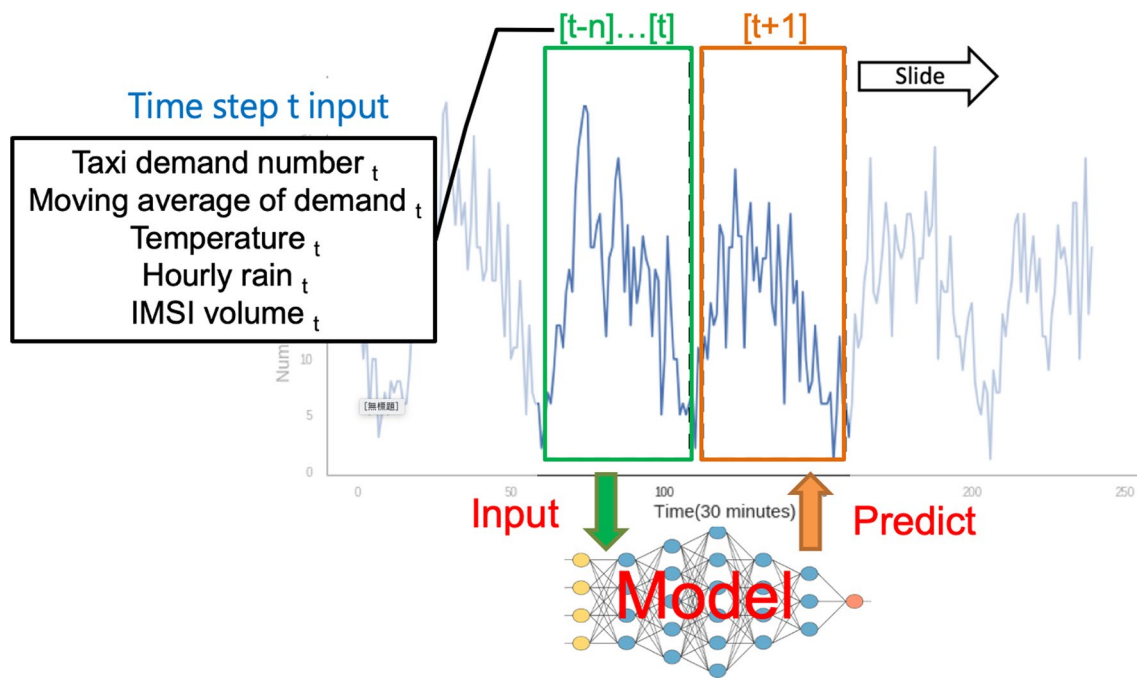
##### 3.1.3 CNN

CNN was introduced in [17]. The authors of [11] mentioned that time-series data with multiple windows can be applied as the input of the convolutional hidden layer. CNN was used for time-series prediction of speech in [10]. Many types of CNN models can be used for specific purposes when solving the time-series prediction problem.

We design a 1-D CNN model for taxi demand forecasting and the model has a convolutional hidden layer that operates over a 1-D sequence. Moreover, it includes a pooling

**Fig. 4** A LSTM network structure





**Fig. 5** Build a predictive model that inputted with multi-step historical data and each step has multiple inputs

layer that reduces the parameters and extracts the output of the convolutional layer to the primary elements. The convolutional and pooling layers are followed by a dense layer, which is also called the fully connected layer. This layer interprets the features extracted by the convolutional part of the model. A flatten layer is used between the convolutional layers and the dense layer to reduce the feature maps to a single 1-D vector. The CNN structure is displayed in Fig. 6.

### 3.1.4 CNN-LSTM

The combination of the CNN and LSTM model can be used in time-series forecasting problems [12]. Liu, Lingbo, et al [18] propose a hybrid model that using CNN as a front-end layer for feature extraction of input data and followed with LSTM as the second layer to support sequence prediction. We extend the CNN-LSTM model for multivariate time-series forecasting problems, where the input data are multiple features and time-steps. We use CNN as an "encoder" for feature extraction that transforming inputs into an internal matrix or vector representation. In addition, LSTM is adopted as a "decoder" for learning the relevant information in a sequence to predict particular outcomes in the future. The architecture of CNN and LSTM is similar to those illustrated in sections 3.1.1 and 3.1.3 separately.

Developing a deep network allows the hidden state at each level to operate at different time steps [19]. Figure 7 illustrates the architecture of our CNN-LSTM model, this class of model that is both spatially and temporally deep that

can be applied to more complicated time-series forecasting problems and optimize the time complexity due to the reduction of feature engineering. In our experiments, the model convergence of CNN-LSTM can faster than the recurrent neural network more than 2 times.

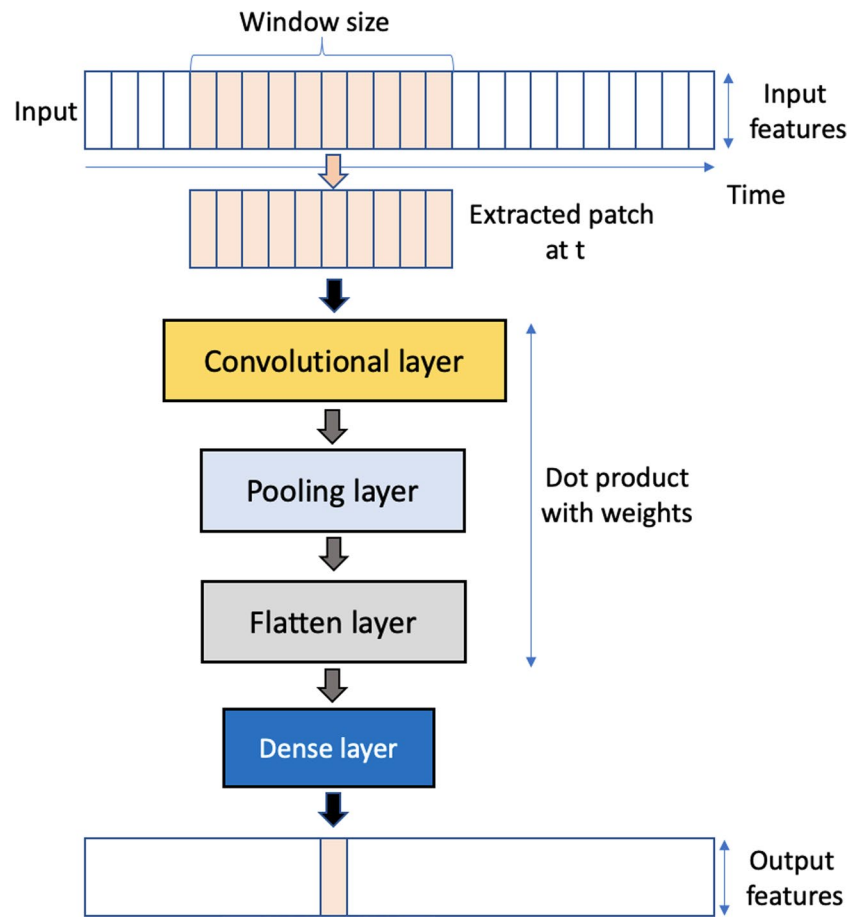
### 3.2 Each Time Step with Multiple Features

In this paper, we build the LSTM model for multivariate time series forecasting in each time step. Figure 8 displays the sequential patterns of each input feature in night market areas as an example. The effect of each feature on the prediction results is discussed in the experimental section. The details of the information are as follows:

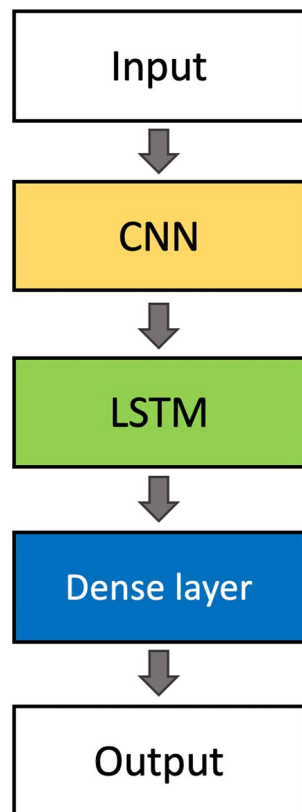
- 1) Taxi demand: the taxi demand over the last half hour
- 2) Day of the week: Day of the week represents which day of the week that time-step is and the range is from 1 to 7.
- 3) Temperature: in Celsius
- 4) Hourly rain: in mm
- 5) Moving average of taxi demand:

Moving average is commonly used with the time-series data to smooth out short-term fluctuations and highlight longer-term trends or cycles. As mentioned in [6], we can add the moving average of the taxi demand as an input feature. In this paper, the moving average is the unweighted

**Fig. 6** A CNN layer architecture. The dense layer is also called the fully connected layer



**Fig. 7** The architecture of the CNN-LSTM model. The dense layer is also called the fully connected layer



arithmetic mean of the values obtained in the first five-time steps.

6) IMSI volume:

The IMSI volume we use in this paper implies the number of population in outdoor, transformed from the International Mobile Subscriber Identity (IMSI) that collected at cell sites by Chunghwa Telecom Company. In the experimental section, we discuss if the predicted results can be improved by using the IMSI volume as one training feature.

**4 Experiments and Discussion**

In this experimental section, the dataset in downtown Kaohsiung during 2019 is obtained from Chunghwa Telecom and the Central Weather Bureau of Taiwan. In the aforementioned section 3.1.1, time-series data must be converted to a supervised learning format. We re-construct the dataset into the input (X) and output (Y) with different sliding windows lengths, that is, the input (X) is a certain range of historical data and output (Y) is the data of next time step. Therefore, the forecasting model predicted output (Y) according



to input (X). We use 80% of the supervised learning format data for training and keep the remaining 20% for validation. Moreover, we fit the models over 1000 training epochs and repeat 20 times, to randomize the network weights in the initialization. To prevent the problem of overfitting in the training progress, we also adopted the EarlyStopping function of Keras.

The predictive model can be trained using arbitrary sequence lengths. However, constrained by the model convergence, we use every one-week data as a sequence and split it into time-steps with different lengths. For instance, if the time-step length is 10 min, the sequence length would be  $24 \times 7 \times 6$ . If the time-step length is 30 min, the sequence length would be  $24 \times 7 \times 2$ . For the 30 min case, the training input data shape is (13939, 336, n) in which 13939 is the total number of sequences in the training dataset, 336 is the sequence length:  $24 \times 7 \times 2$ , and n is the number of input features. Table 1 shows the range of the experimental parameters outlined.

To examine the performance of the predictive models, we adopt the widely used prediction error metrics: Symmetric Mean Absolute Percentage Error (sMAPE) [20]. The formulation of the prediction error metric is given as follows:

$$sMAPE_k = \frac{1}{T} \sum_{t=1}^T \frac{|Y_{k,t} - \hat{Y}_{k,t}|}{Y_{k,t} + \hat{Y}_{k,t} + c} \quad (1)$$

$Y_{k,t}$  and  $\hat{Y}_{k,t}$  represent the real and the predicted values of the actual taxi demand for dispatch area  $k$  at time-step  $t$ , respectively.  $T$  is the length of training or testing data. We also include constant  $c$  in the denominator to avoid division by zero where some dispatch areas the number of taxi demand might be zero at some time-steps.

## 4.1 Experimental Results

We systematically examine the performance of four prediction algorithms in the aforementioned section with sMAPE as prediction error metrics. In addition, we evaluate the prediction performance of the multivariate inputs in different models. For each experiment case, we train forecasting models 20 times in order to randomize the network weights in the initialization.

### 4.1.1 Performance Over Dispatch Area of the Kaohsiung

We select 7 dispatch areas with various geographical positions to illustrate the prediction error of different models and separate them into two groups, where one has the regular demand pattern and there are rare events frequently in another. Theoretically, the prediction model can learn the sequence correlation efficiently based on regular pattern

data. Therefore, we conduct the experiments separately of each group to demonstrate the prediction performance.

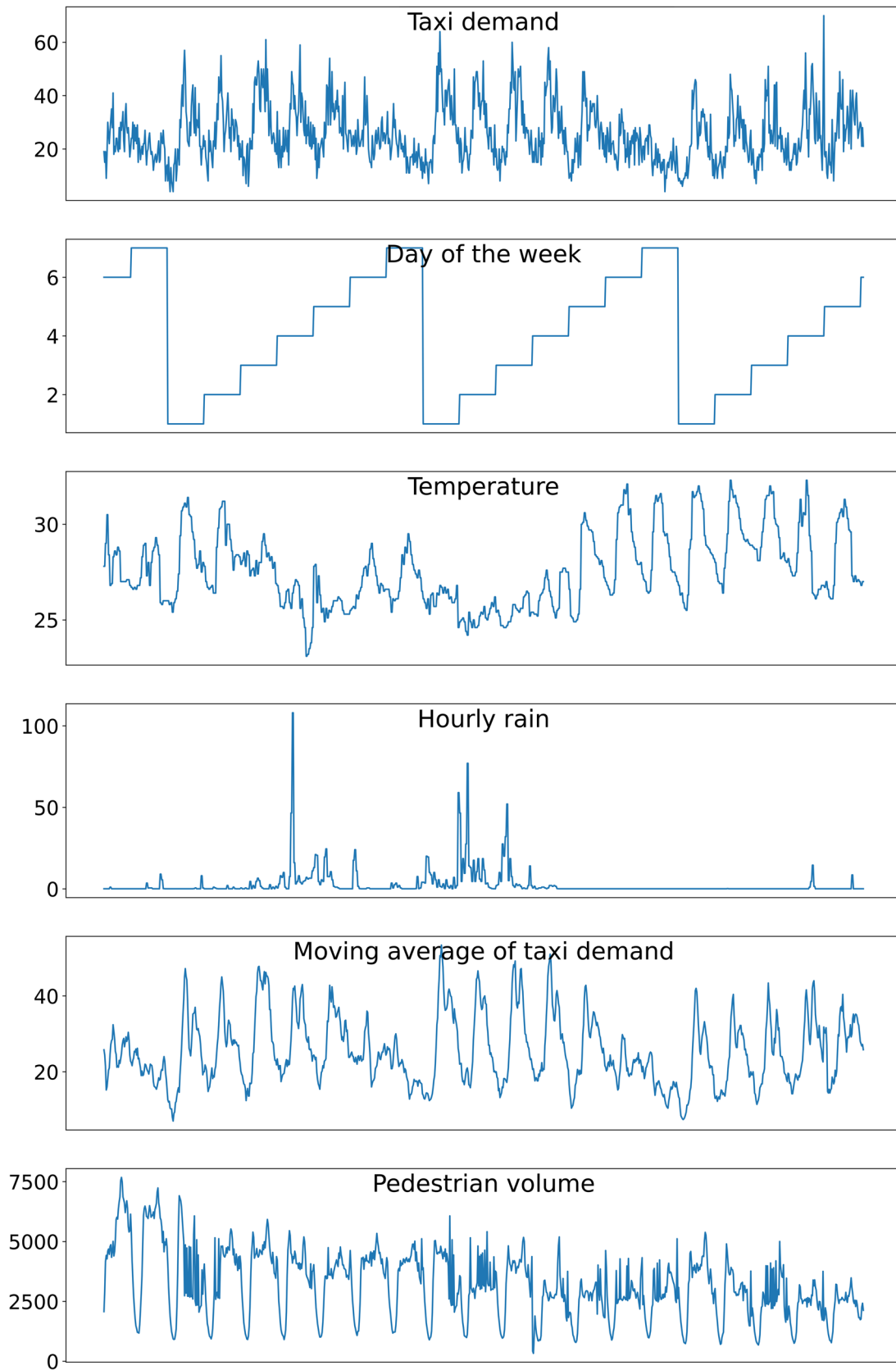
Before the experiment, we need to select the length of historical data for predictive models. We slightly evaluate the performance through the CNN-LSTM model for 5 selection: 6 hours, 12 hours, 24 hours, 36hours, and 48 hours. Fig. 9(a) and Fig. 9(b) show the sMAPE over different reference lengths of historical data. As we can see, the model has the minimum sMAPE when adopted 36 hours of historical data. Therefore, we set the length of the reference data as 36 hours in the following experiments.

We conduct the experiments separately of each group to demonstrate the prediction performance. First, we select areas with regular patterns such as night market, department store, MRT transfer station, and software park. Table 2 (a) displays the prediction result of all models over four dispatch areas. The peak time of the night market usually occurs at night and the crowds in this lifestyle have generated a unique trend of taxi demand. At the weekend, the large volume of IMSI near the department store also suggest a large population, and the demand for taxis will gradually go up along with this phenomenon. For the MRT transfer station, demand events are concentrated during commuting peaks accompanied by variations in the local IMSI volume. The prediction result in the software park is slightly underperformed. Although the demand pattern generally changes along with the commuting hours, due to crowd gathering in some tourist spots such as Kaohsiung 85 Building, there are some short-term fluctuations of taxi demand may affect the prediction results. The prediction result in the hospital areas is slightly worse than those in other areas, with the demand pattern that taxi demand is usually concentrated at the weekend.

Table 2(b) shows the results of another group with an unregular demand pattern. In our analysis, the trend of taxi demand in these areas is susceptible to festivals or occasional activities, resulting in that the neural network algorithm is more difficult to learn based on only the historical data. In dispatch areas such as the art centres and the tourist attractions, there are regular taxi demands from visitors. However, the demand events in these areas prone to be affected by commuting peaks or holidays.

Compare to two groups of prediction results, CNN-LSTM, LSTM and GRU can outperform the CNN model. These three forecasting models are close to each other but the result shows that CNN-LSTM still provides better prediction. The average sMAPE in the areas with a regular pattern is 23.23% and in another group is 36.81%, and it can be seen that CNN-LSTM is the best independent predictive model for taxi demand prediction.

CNN model underperforms in most areas because the CNN model considers spatial relation rather than the temporal sequential relation which makes it difficult to remember



**Fig. 8** The sequential patterns of each feature in the night market, where a time step is 30 min

the correlation in the sequence and learning long-term dependencies. However, the procedure of training time is faster 10 times than LSTM and 2 times than CNN-LSTM. By taking this advantage of faster prediction computation, we can determine previously the performance through different input features.

In the next experiment, we demonstrate the use of Analysis of Variance (ANOVA) for analyzing the significance of performance over prediction sMAPE and prediction time of different forecasting models.

#### 4.1.2 ANOVA and Post-Hoc Test for Forecasting Models

In this paper, four forecasting models are built by using CNN, GRU, LSTM, and CNN-LSTM algorithms at different dispatch areas. We build the predictive model of each algorithm 20 times to randomize the network weighting, therefore there are 20 sets of estimation values for each algorithm at one area.

To verify the statistically significant differences of prediction error and prediction time for the four models respectively shown in Tables 1 and 2, the one-way analysis of variance test [21] method is used, or ANOVA for short. The one-way ANOVA is a statistical test on two or more independent groups to see if the group means are significantly different from each other. A statistically significant result, when a probability (p-value) is less than a pre-specified threshold (significance level  $\alpha$ ), justifies the rejection of the null hypothesis, but only if the a priori probability of the null hypothesis is not high.

Table 3 shows the means, standard derivations and the ANOVA tests of prediction error of the four forecasting models. The results of the ANOVA test for prediction error show that the prediction error of the four algorithms forecasting models is statistically significantly different at those areas based on the evidence that the p-value is lower than a chosen significance level  $\alpha$  ( $\alpha$  set as 0.05).

Prediction computation is crucial for a method to be used in a real-world setting like the large scale of models deployed to forecast in an instant. We compare four predictive models and the mean and standard derivation of prediction time generated from predictive models are shown in Table 4. We also perform the ANOVA tests for the prediction time. The p-value of each dispatch area is also less than the significance level 0.05, as a result, the test verifies the prediction time of the four algorithms forecasting models is statistically significantly different.

When the result is significant from ANOVA, it illustrates that at least one group differs from the others. However, the omnibus test does not recognize the pattern of differences between the means. Therefore, the post-hoc test is executed and Tukey Honestly Significant Difference (HSD) [22] was chosen. We use the data groups of Night Market to illustrate as an example and other areas follow a similar trend. The

boxplot of the sMAPE and the boxplot of the prediction time are shown independently as Fig. 10 and Fig. 11. We adopt the post-hoc test and the results of each comparison combination are shown in Table 5.

Table 5 indicates the significant differences of prediction error between CNN and other group. And according to the means of prediction error in Table 3, the recurrent network algorithms such as GRU, LSTM, and stacked model CNN-LSTM significantly outperform the CNN at Night Market. Furthermore, the p-value between the recurrent networks algorithms reveal a non-significant difference.

For the aspect of the prediction time, Table 4 shows that the CNN model is almost 7 times faster than the LSTM model and 2 times faster than the stacked model. According to Table 6, the CNN model has a significant difference from other models at Night Market. Meanwhile, the stacked model CNN-LSTM reveals a significant difference in prediction time and faster than other RNN models. Therefore, in the comprehensive consideration both of prediction error and time, the experiment shows that CNN-LSTM is more suitable for predicting taxi demand in the future.

#### 4.1.3 Performance Over Specific Dispatch Areas

The comparison of prediction results over neural network algorithms is presented in the aforementioned experiment. In this experiment, we select the night market and the department store and look further into these specific dispatch areas. The night market is unlike common food courts, the peak time of the night market usually occurs at night. In the department store, the IMSI volume is concentrated at the weekend. The higher the IMSI volume in the dispatch area, the higher is the taxi demand.

Figures 9 and 10 displays the comparison of one-day prediction and one-week prediction of each model and illustrate that which forecasting models in different cases can provide a better prediction. The time-step length used here is 30 min in both areas. Figures 12(a) and 13(a) are the comparisons of one-week prediction of each model with the actual demand in selected areas while Figs. 12(b) and 13(b) are the comparisons of one-day prediction. It can be seen that the CNN-LSTM model follows the trend in different areas that provides a better prediction overall.

#### 4.1.4 Importance of the Input Feature

The features considered in this paper include the historical taxi demand, moving average of the taxi demand, day of the week, weather information, and volume of the local outdoor IMSI. The weather information includes temperature and hourly rain. We obtain official weather information for Kaohsiung from the Central Meteorological Bureau of Taiwan. The IMSI volume indicated the population outdoor

**Table 1** Experimental parameters

| Type                          | Range         |
|-------------------------------|---------------|
| Map grid size                 | 500 m × 500 m |
| Data of each sequence         | 36 hours      |
| Time-step length              | 30 min        |
| Batch size                    | 128           |
| Number of dispatch areas      | 7             |
| Number of hidden layers       | 1-2           |
| Number of hidden neurons      | 200           |
| Number of convolution filters | 200           |
| Number of pooling layer       | 1             |

is obtained from the big data department of Chunghwa Telecom. To evaluate the prediction performance of the aforementioned features, we conduct two experiments with the one-week data. To show the impacts of these features on forecasting performance, we conduct two experiments. Both experiments are based on the CNN-LSTM model since we have evaluated the forecasting performance in the aforementioned section which indicated that the CNN-LSTM model can achieve better prediction.

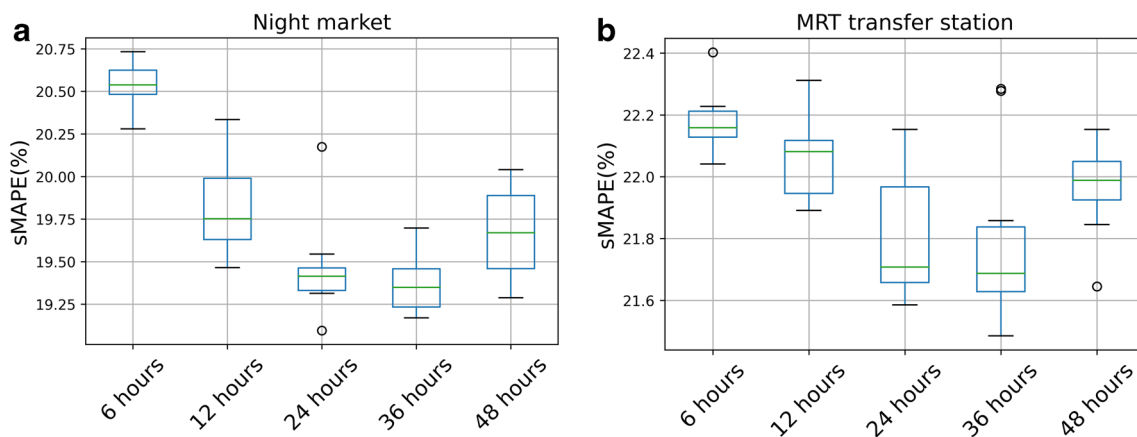
In experiment 1, we illustrate the importance of each impacting feature inputted with the model. We design six models in Table 7 to verify the performance that is inputted to. All the models are expected to output the taxi demand in the city at the next time step.

We conduct the experiments on all dispatch areas and select four specific areas to demonstrate the prediction errors, as depicted in Fig. 14. Model 1 predicts the number of future demand based on the past period of historical data. Theoretically, the historical data of taxi demand is a valuable factor when forecasting future taxi demand. Based on our prediction results

of all areas, taxi demand is truly the most valuable information in conducting the prediction. For model 2, although there is no past taxi demand information provided, the model tries to learn the mapping between the input feature to the real taxi demand at each time-step. It turns out that Model 2 is sometimes outperforming than other models, which indicates that the demand pattern varies with the day of the week. Model 3 and 4 are underperformed in the forecasting. The climate of Kaohsiung is one factor behind this underperformance that rainfall is concentrated in summer and scarce in other seasons. Another problem is that climate information can be obtained only from the major local-based observatories, which are insufficient to represent the city. Prediction of the taxi demand is difficult for models based on weather history. For model 5, we use the moving average of taxi demand in each area as input. A moving average is commonly

**Table 2** Average sMAPE of each model in different areas

|                    | CNN    | GRU    | LSTM   | CNN-LSTM |
|--------------------|--------|--------|--------|----------|
| (a)                |        |        |        |          |
| Night market       | 21.41% | 19.65% | 19.48% | 19.35%   |
| MRT station        | 22.68% | 22.33% | 22.03% | 21.78%   |
| Department store   | 24.26% | 24.67% | 23.67% | 22.77%   |
| Hospital area      | 27.49% | 25.92% | 26.09% | 25.91%   |
| Software park      | 28.15% | 28.05% | 27.07% | 26.35%   |
| All places         | 24.80% | 24.12% | 23.67% | 23.23%   |
| (b)                |        |        |        |          |
| Tourist attraction | 41.29% | 36.75% | 36.07% | 36.81%   |
| Art centre         | 43.64% | 37.94% | 37.42% | 37.56%   |
| All places         | 41.29% | 36.75% | 36.07% | 36.81%   |

**Fig. 9** The sMAPE results of dispatch areas over different lengths of reference data

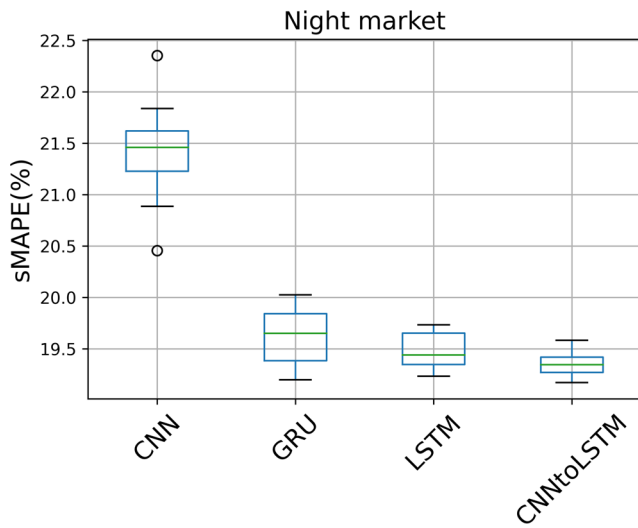


**Table 3** ANOVA test for sMAPE of forecasting models

| Area               | Mean/Std     |              |              |              | p-value   |
|--------------------|--------------|--------------|--------------|--------------|-----------|
|                    | CNN          | GRU          | LSTM         | CNN-LSTM     |           |
| Night market       | 0.214/0.0051 | 0.196/0.003  | 0.194/0.0017 | 0.193/0.0013 | 5.97e-17  |
| MRT station        | 0.226/0.004  | 0.223/0.0011 | 0.22/0.0009  | 0.217/0.0024 | 0.000018  |
| Department store   | 0.242/0.013  | 0.246/0.003  | 0.236/0.005  | 0.227/0.004  | 9.011e-09 |
| Hospital area      | 0.274/0.015  | 0.259/0.001  | 0.26/0.001   | 0.259/0.003  | 0.000171  |
| Software park      | 0.281/0.0105 | 0.28/0.002   | 0.27/0.004   | 0.263/0.0029 | 1.044e-07 |
| Tourist attraction | 0.459/0.034  | 0.391/0.003  | 0.387/0.002  | 0.383/0.004  | 0.000001  |
| Art centre         | 0.412/0.036  | 0.367/0.004  | 0.36/0.003   | 0.368/0.01   | 4.865e-12 |

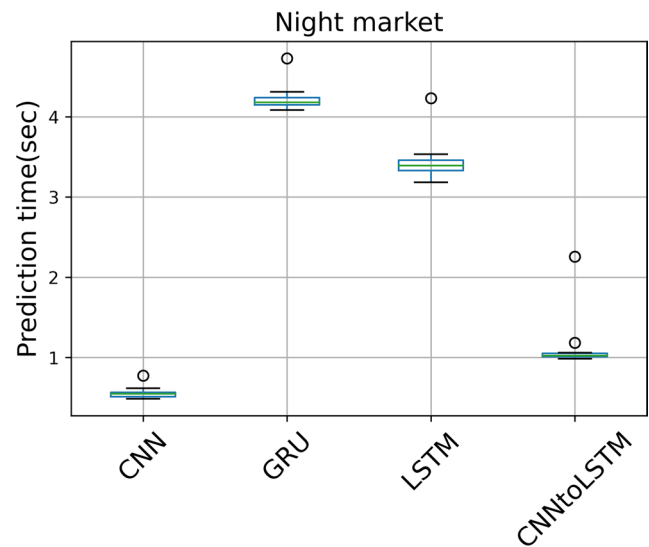
**Table 4** ANOVA test for prediction time of forecasting models

| Area               | Mean/Std     |             |             |             | p-value   |
|--------------------|--------------|-------------|-------------|-------------|-----------|
|                    | CNN          | GRU         | LSTM        | CNN-LSTM    |           |
| Night market       | 0.564/0.082  | 4.235/0.184 | 3.458/0.29  | 1.16/0.389  | 1.292e-28 |
| MRT station        | 0.342/0.024  | 3.128/0.183 | 3.414/0.284 | 0.77/0.0484 | 1.759e-33 |
| Department store   | 0.397/0.059  | 3.098/0.186 | 2.483/0.208 | 0.826/0.052 | 6.734e-33 |
| Hospital area      | 0.525/0.0855 | 4.337/0.145 | 3.387/0.113 | 1.073/0.063 | 4.33e-43  |
| Software park      | 0.561/0.083  | 4.321/0.171 | 3.329/0.106 | 1.181/0.366 | 7.811e-32 |
| Tourist attraction | 0.508/0.043  | 4.388/0.41  | 3.539/0.161 | 1.198/0.43  | 3.301e-32 |
| Art centre         | 0.472/0.039  | 4.144/0.365 | 2.426/0.087 | 1.035/0.072 | 7.299e-27 |



**Fig. 10** The sMAPE results of dispatch areas over different forecasting models

used with time-series data to highlight longer-term trends and smooth out short-term fluctuations. It is interesting to find out it can perform a prediction close to model 1, which means that the moving average pattern has a relevant relationship to the demand pattern. Model 6 considers the volume of IMSI, which is acquired using the distribution of mobile customers in Kaohsiung. In



**Fig. 11** The prediction time of dispatch areas over different forecasting models

our analysis, the taxi demand pattern during commuting peaks or holidays is usually accompanied by variations in the local outdoor IMSI volume. However, the predictive results are worse in some areas due to the reason that the

**Table 5** Post-hoc test for sMAPE of forecasting models at Night Market area

| Group 1  | Group 2  | <i>p</i> -value |
|----------|----------|-----------------|
| CNN      | CNN-LSTM | 0.001           |
| CNN      | GRU      | 0.001           |
| CNN      | LSTM     | 0.001           |
| CNN-LSTM | GRU      | 0.2469          |
| CNN-LSTM | LSTM     | 0.7904          |
| GRU      | LSTM     | 0.7213          |

**Table 6** Post-hoc test for prediction time of forecasting models at Night Market area

| Group 1  | Group 2  | <i>p</i> -value |
|----------|----------|-----------------|
| CNN      | CNN-LSTM | 0.001           |
| CNN      | GRU      | 0.001           |
| CNN      | LSTM     | 0.001           |
| CNN-LSTM | GRU      | 0.001           |
| CNN-LSTM | LSTM     | 0.001           |
| GRU      | LSTM     | 0.001           |

ISMI volume from the big data department of Chunghwa Telecom is not stable yet, makes the data in some periods may be incomplete.

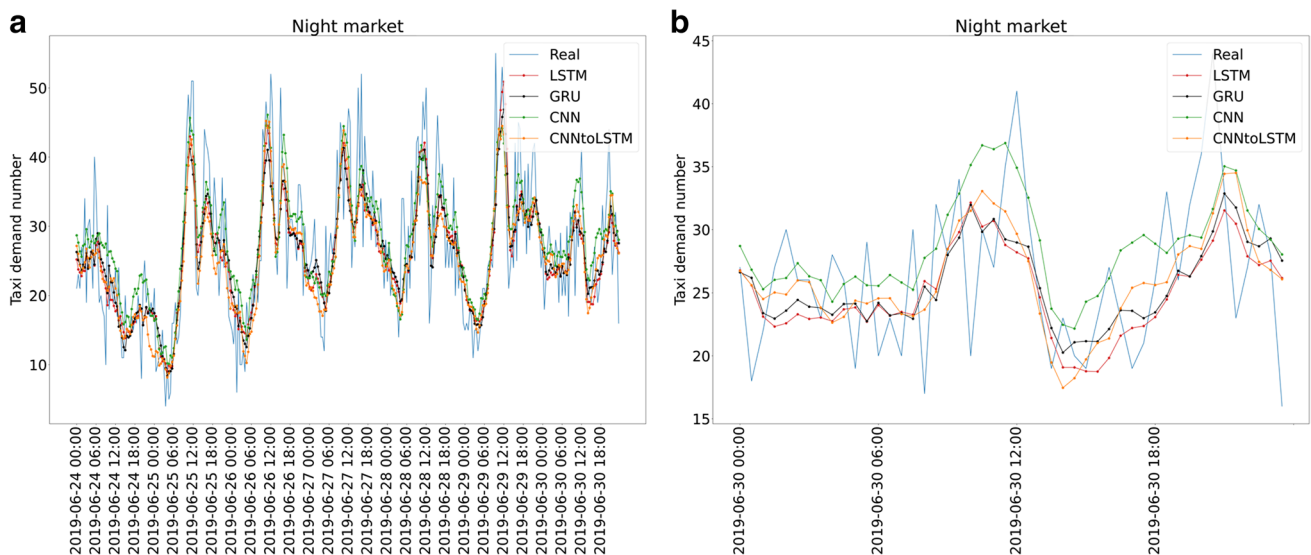
In experiment 2, we set taxi demand as the default input feature, which is based on the prediction results in experiment 1, and combined with one selected impacting feature as multivariate inputs. We also design five models and Table 8 shows the input to each model. As

depicted in Fig. 15, the results indicated taxi demand combined with the feature "day of the week" as input features show an outstanding prediction performance in most areas, which implied that the trend of taxi demand is followed by cycle regularly. Meanwhile, taxi demand combined with temperature, moving average, and IMSI volume respectively, can improve the model forecasting performance compared to the model fed with a single feature. In this paper, the current IMSI volume only indicated the population outdoor. We could extract more crowd information related to the taxi demand in the future and verify the prediction performance.

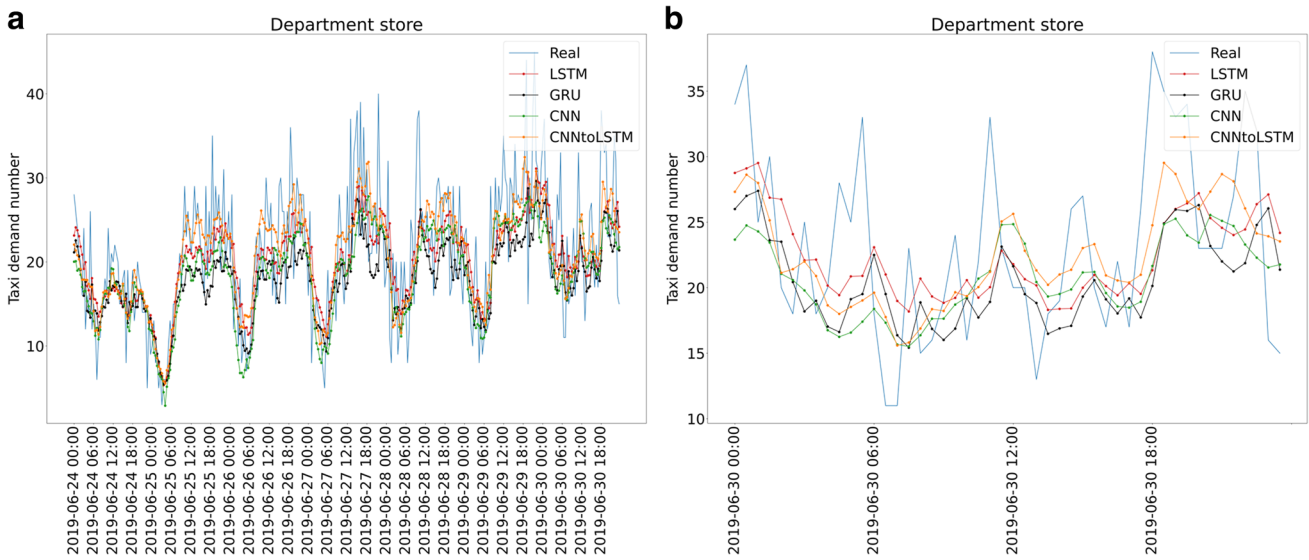
#### 4.1.5 Performance Metrics in Practical Environment

Currently, we start to conduct the proof of concept (PoC) of our proposed method and builds the predictive models in the selected dispatch area. For the evaluation of PoC, we adopt the performance metric proposed in [23] by the NTT DOCOMO, Japan’s largest mobile service provider, where evaluates a demand forecasting service for each block or street. The predicted values are defined as a correct prediction if the error between the predicted and actual value is within 50 percent. The accuracy is defined as the number of correct predictions divided by the number of all forecasts. The best accuracy of our model achieves 92% at located MRT transfer station and the average accuracy in areas with the regular demand pattern is 89%.

To demonstrate the actual prediction results, we use the well-fitted CNN-LSTM model to forecast the taxi demand of the four selected dispatch areas. Figure 16



**Fig. 12** Prediction results of models in the night market area **a** One-week prediction, **b** One-day prediction



**Fig. 13** Prediction results of models in the department store area **a** One-week prediction, **b** One-day prediction

**Table 7** Model with different input features (1)

|         | Model inputs                  | Model Output               |
|---------|-------------------------------|----------------------------|
| Model 1 | Taxi demand                   | Next half-hour taxi demand |
| Model 2 | Day of the week               | Next half-hour taxi demand |
| Model 3 | Temperature                   | Next half-hour taxi demand |
| Model 4 | Hourly rain                   | Next half-hour taxi demand |
| Model 5 | Moving average of taxi demand | Next half-hour taxi demand |
| Model 6 | IMSI volume                   | Next half-hour taxi demand |

compares the trend of real with predicted taxi demands in one week over four dispatch areas. The results indicate the model can identify the demand patterns such as the periods during the commuting hours or the specific peak hours at night, although, the model is slightly underperformed for forecasting in certain large-variation periods.

Overall, the predictive model based on the area-based data can perform well for identifying the demand patterns. One of the differences between our work and the previous works is that our model can consider multivariate feature in each time step and refer to multiple historical intervals to capture long term dependencies in a sequence. Another advantage of our model is that we

consider different characteristics of dispatch areas with a local pattern such as taxi demand, IMSI volume. This approach gives a more realistic prediction as it takes into account the uncertainty while predicting.

For better prediction, we can further improve our model by considering rare events such as public holidays and special activities. The demand pattern in rare events is not likely than general patterns and taxi demand in dispatch areas is ups and downs sharply based on different regional characteristics. Prediction of the taxi demand is difficult for a model based on only a certain range of historical data. Previous work proposed by Uber laboratory [15] investigates the importance of pre-treatment for special holidays, which is the concrete way for the improvement of predictive accuracy.

### 5 Conclusion and Future Work

This paper proposes a taxi-demand prediction approach based on deep-learning models, and the predictions on specific-areas in Kaohsiung, Taiwan is compared. The training data with a duration of 12 months include concerned GPS trajectory of the taxi fleet, weather information from Central Weather Bureau of Taiwan, and IMSI from Chunghwa Telecom. The training data are allocated to the corresponding dispatch areas with a partitioning

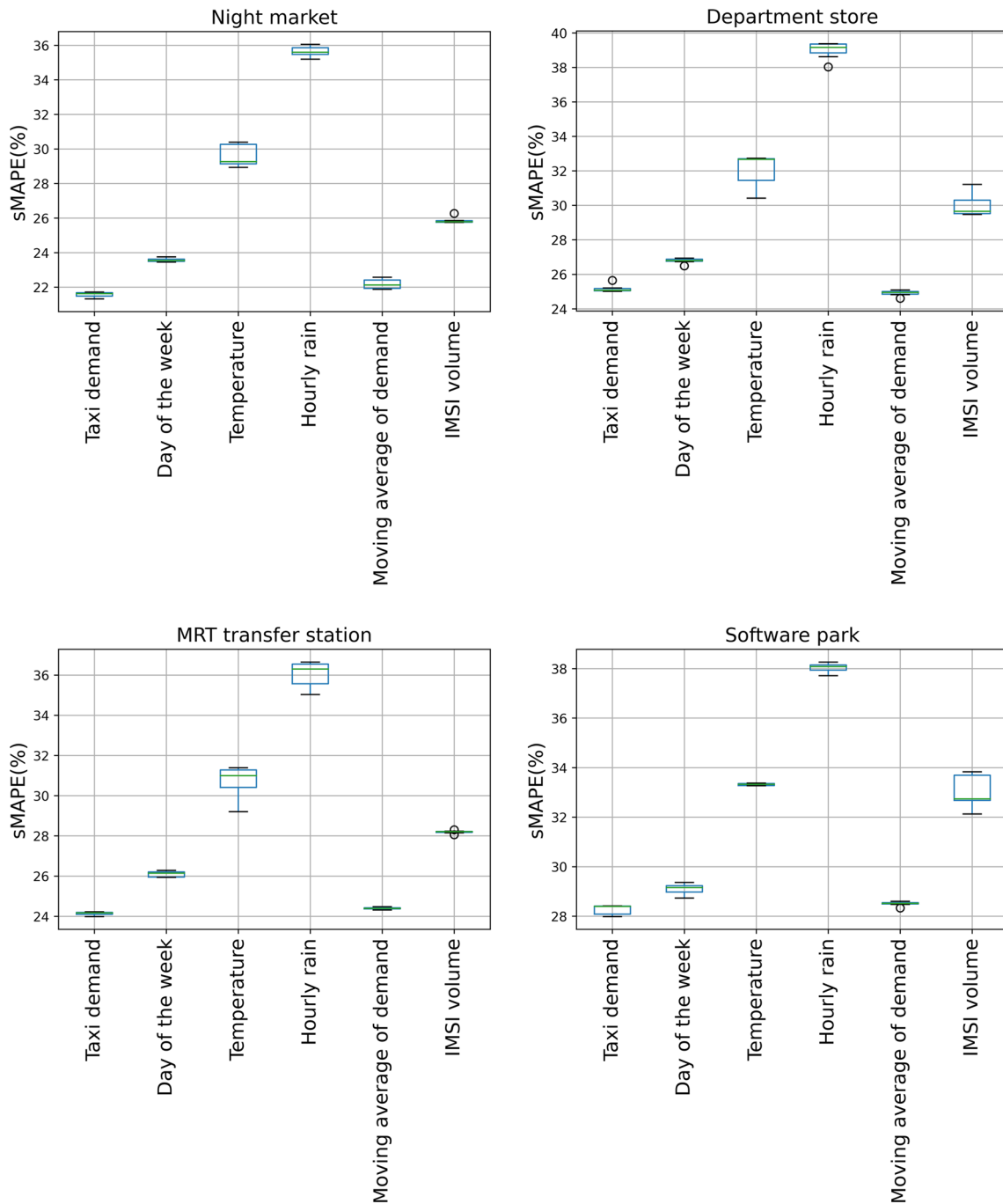


Fig. 14 Prediction performance on different single input features

and grouping process. Because the characteristic of dispatch areas has an influence on the prediction, 7 specific dispatch-areas with various geographical positions are analyzed in our experiments.

In experiments, we build and evaluate four neural network algorithms to find the best predictive model over the dispatch

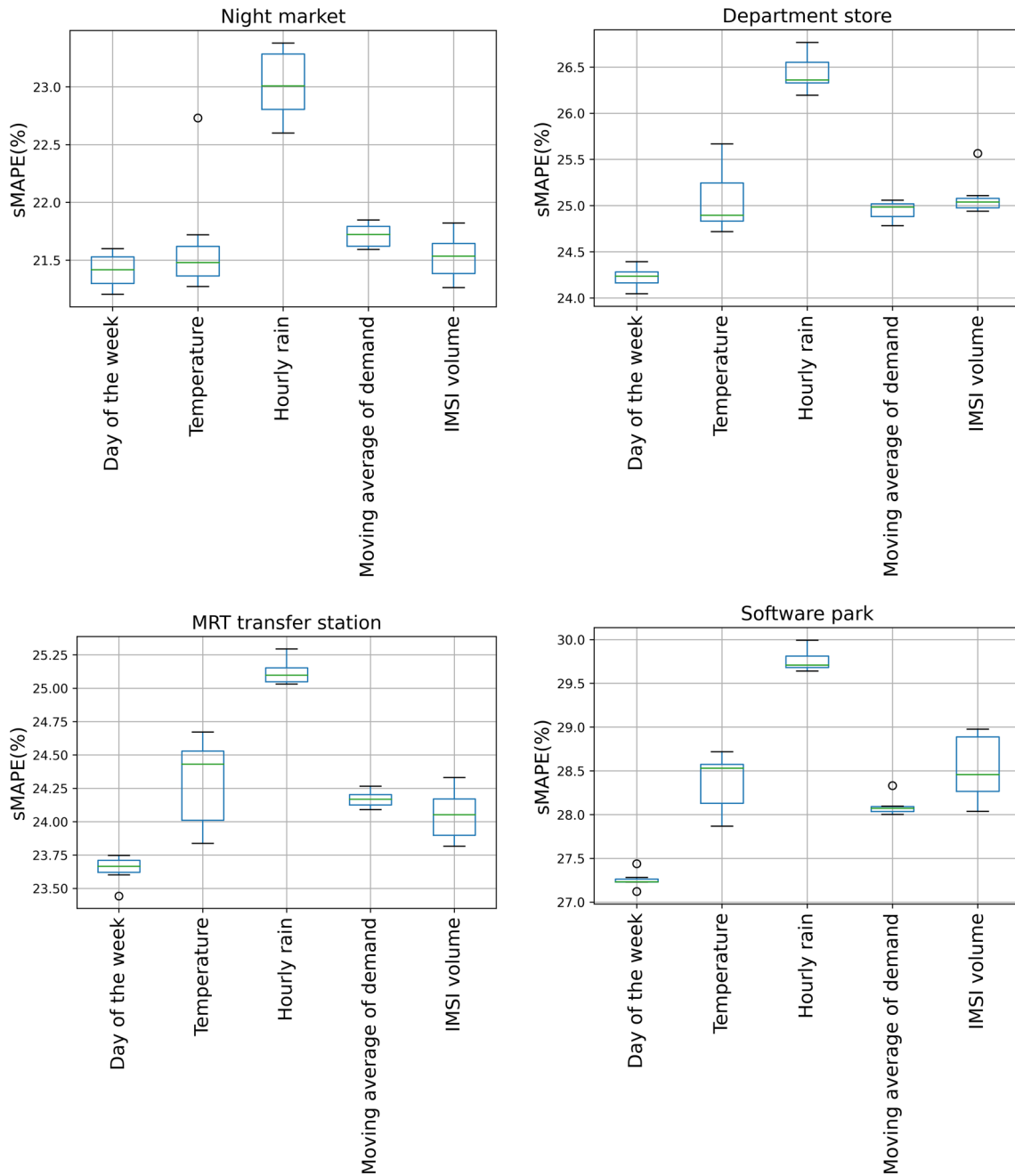
area. The results show that CNN-LSTM is the most suitable forecasting algorithm when considering both performance and time comprehensively. Additionally, referring to the real-world performance metric on taxi demand forecasting by NTT DOCOMO, the proposed area-based prediction for taxi demand is functional. Among the concerned features, how to locate the



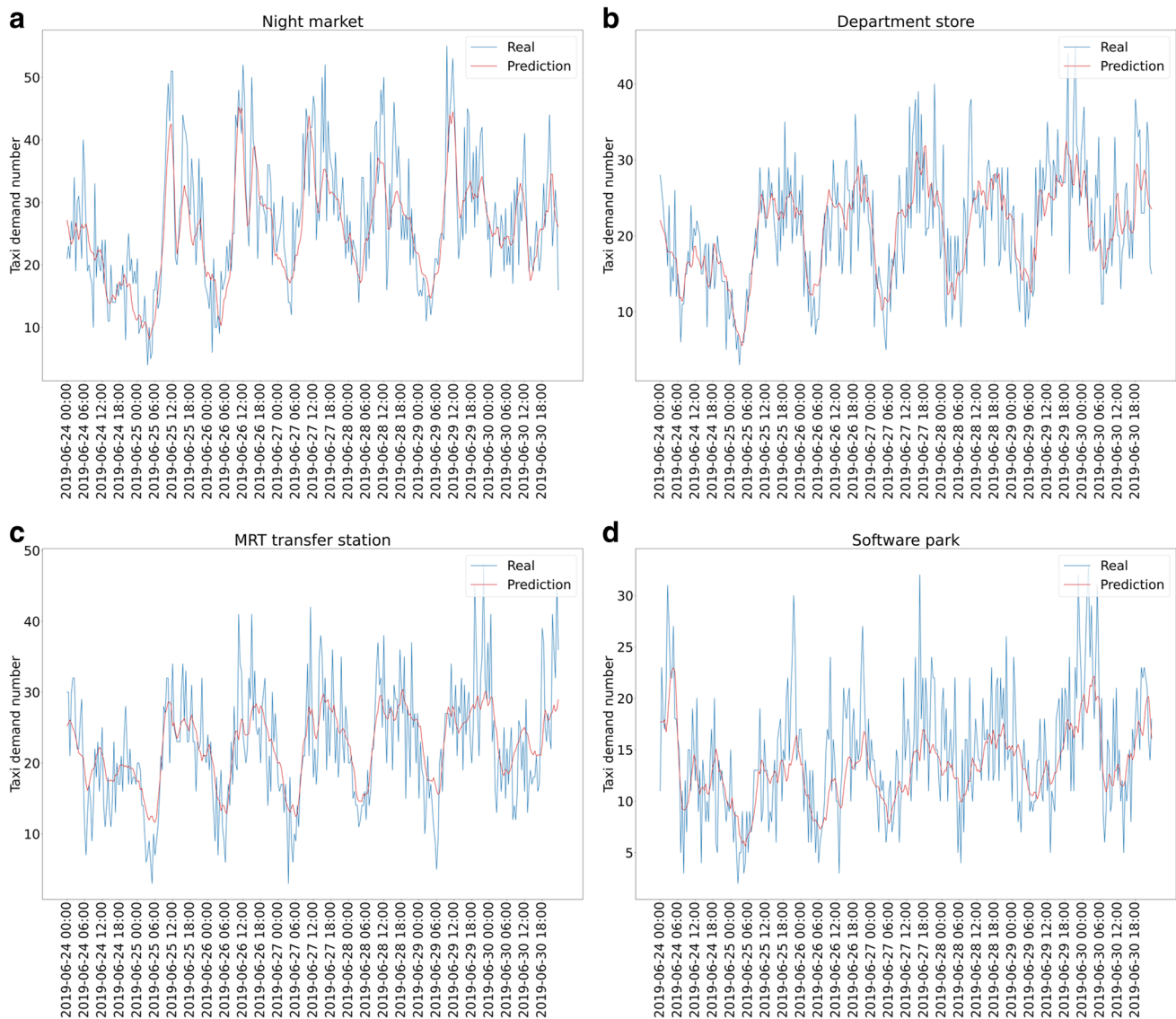
**Table 8** Model with different input features (2)

|         | Model inputs                                | Model Output               |
|---------|---|----------------------------|
| Model 1 | Taxi demand & Day of the week               | Next half-hour taxi demand |
| Model 2 | Taxi demand & Temperature                   | Next half-hour taxi demand |
| Model 3 | Taxi demand & Hourly rain                   | Next half-hour taxi demand |
| Model 4 | Taxi demand & Moving average of taxi demand | Next half-hour taxi demand |
| Model 5 | Taxi demand & IMSI volume                   | Next half-hour taxi demand |

passenger population is critical, and this paper utilizes outdoor IMSI. However, the IMSI data contains more information related to personal behavior and movement; therefore, the IMSI will be further studied. Furthermore, rare events which may cause short-term fluctuations, such as peak traffic demand traffic jam severity, etc.



**Fig. 15** Prediction performance on different input features



**Fig. 16** The comparison of real value and prediction in given dispatch areas. **a** Night market, **b** MRT transfer station, **c** Department store, **b** Software park

## References

- Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., Damas, L.: Predicting taxi–passenger demand using streaming data. *IEEE Trans. Intell. Transp. Syst.* **14**(3), 1393–1402 (2013)
- Xu, J., Rahmatizadeh, R., Bölöni, L., Turgut, D.: Real-time prediction of taxi demand using recurrent neural networks. *IEEE Trans. Intell. Transp. Syst.* **19**(8), 2572–2581 (2017)
- Zhang, D., Sun, L., Li, B., Chen, C., Pan, G., Li, S., Wu, Z.: Understanding taxi service strategies from taxi GPS traces. *IEEE Trans. Intell. Transp. Syst.* **16**(1), 123–135 (2014)
- Rahaman, M.S., Ren, Y., Hamilton, M., Salim, F.D.: Wait time prediction for airport taxis using weighted nearest neighbor regression. *IEEE Access* **6**, 74660–74672 (2018)
- Ke, J., Zheng, H., Yang, H., Chen, X.M.: Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. *Transp. Res. Part C Emerg. Technol.* **85**, 591–608 (2017)
- Yu, H., Chen, X., Li, Z., Zhang, G., Liu, P., Yang, J., Yang, Y.: Taxi-Based Mobility Demand Formulation and Prediction Using Conditional Generative Adversarial Network-Driven Learning Approaches. *IEEE Trans. Intell. Transp. Syst.* **20**(10), 3888–3899 (2019)
- De Brébisson, Alexandre, et al. Artificial neural networks applied to taxi destination prediction. *arXiv preprint arXiv:1508.00021* (2015).
- Hochreiter, S., Schmidhuber, J.: Long short-term memory. *Neural Comput.* **9**(8), 1735–1780 (1997)
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078* (2014)

10. Abdel-Hamid, Ossama, et al. Deep segmental neural networks for speech recognition. *Interspeech*. Vol. 36. (2013)
11. Mittelman, R.: Time-series modeling with undecimated fully convolutional neural networks. *arXiv preprint arXiv:1508.00317* (2015)
12. Shi, Xingjian, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *arXiv preprint arXiv:1506.04214* (2015)
13. Miao, F., Han, S., Lin, S., Stankovic, J.A., Zhang, D., Munir, S., et al.: Taxi dispatch with real-time sensing data in metropolitan areas: A receding horizon control approach. *IEEE Trans. Autom. Sci. Eng.* **13**(2), 463–478 (2016)
14. Duan, Z., Zhang, K., Chen, Z., Liu, Z., Tang, L., Yang, Y., Ni, Y.: Prediction of City-Scale Dynamic Taxi Origin-Destination Flows Using a Hybrid Deep Neural Network Combined With Travel Time. *IEEE Access.* **7**, 127816–127832 (2019)
15. Laptev, N., Yosinski, J., Li, L. E., & Smyl, S.: Time-series extreme event forecasting with neural networks at uber. In *International Conference on Machine Learning* (Vol. 34, pp. 1–5) (2017)
16. Chung, J., Gulcehre, C., Cho, K., Bengio, Y.: Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555* (2014)
17. LeCun, Yann, Yoshua Bengio: Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* **3361**(10), 1995 (1995)
18. Liu, Lingbo, et al. Contextualized spatial–temporal network for taxi origin-destination demand prediction. *IEEE Transactions on Intelligent Transportation Systems* **20**.10: 3875–3887 (2019)
19. Pascanu, R., Gulcehre, C., Cho, K., Bengio, Y.: How to construct deep recurrent neural networks. *arXiv preprint arXiv:1312.6026* (2013)
20. Lopez-Garcia, P., Onieva, E., Osaba, E., Masegosa, A.D., Perallos, A.: A hybrid method for short-term traffic congestion forecasting using genetic algorithms and cross entropy. *IEEE Trans. Intell. Transp. Syst.* **17**(2), 557–569 (2015)
21. Fisher, R.A.: XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth Environ. Sci. Trans. R. Soc. Edinb.* **52**(2), 399–433 (1919)
22. Abdi, H., Williams, L.J.: "Tukey's honestly significant difference (HSD) test." *Encyclopedia of Research Design*, pp. 1–5. Sage, Thousand Oaks (2010)
23. Business, IoT, and Masanori Fujita Ryohei Suzuki Akihito Makishima. *AI Taxi— Taxi Passenger Demand Prediction Technology for Optimizing Traffic*——.(2018)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Shen-Lung Tung** received the Ph.D. degree from the Department of Electrical Engineering, National Central University, Taiwan. His research interests include machine learning, fuzzy theory, and the application of intelligent transport systems. He joined the Chunghwa Telecom Laboratories in 1998, currently working as a researcher at the Transportation and Safety Lab, responsible for the development of intelligent transportation systems.



**Po-Wen Lu** received the Ph.D. degree from the Department of Mechanical Engineering Power, National Tsing Hua University, Taiwan. He joined the Chunghwa Telecom Laboratories in 1999, currently working as a researcher at the Transportation and Safety Lab, responsible for the development of intelligent transportation systems.



**Tzu-Cheng Liu** received the M.S. degree from the Department of Information Management, National Taiwan University of Science and Technology, Taiwan. He joined the Chunghwa Telecom Laboratories in 1999, currently working as an assistant manager at the Transportation and Safety Lab, responsible for the development of intelligent transportation systems.



**Chung-Yi Lin** received the M.S. degree from the Department of Engineering Science, National Taiwan University, Taiwan. He is currently pursuing the Ph.D. degree in computer science from the Graduate Institute of Networking and Multimedia, National Taiwan University, Taiwan. His research interests include computer vision, graph neural network, and the application of intelligent transportation systems.