Use of Deep Learning to Predict Daily Usage of Bike Sharing Systems

Hong Yang, Kun Xie, Kaan Ozbay, Yifang Ma, Zhenyu Wang

This is the author's version of a work that has been accepted for publication in the Transportation Research Record, Journal of the Transportation Research board after being presented at the Transportation Research Board's 97th Annual Meeting, Washington, D.C., 2018. The final version can be found at doi.org/10.1177/0361198118801354.

Use of Deep Learning to Predict Daily Usage of Bike Sharing Systems

Hong Yang [1*], Kun Xie [2], Kaan Ozbay [3], Yifang Ma [4], Zhenyu Wang [5]

[1] Assistant Professor, Ph.D.
Department of Modeling, Simulation & Visualization Engineering
Transportation Research Institute
Old Dominion University
4700 Elkhorn Ave, Norfolk, VA 23529, USA
E-mail: hyang@odu.edu, Phone: (757) 683-4529
[*] Corresponding Author

[2] Lecturer, Ph.D.
Department of Civil and Natural Resources Engineering
University of Canterbury
20 Kirkwood Ave, Christchurch 8041, New Zealand
E-mail: kun.xie@canterbury.ac.nz, Phone: +64-3-369-2707

[3] Professor & Director, C2SMART Center (A Tier 1 USDOT UTC),
Department of Civil and Urban Engineering &
Center for Urban Science and Progress (CUSP),
Tandon School of Engineering, New York University (NYU)
Six MetroTech Center, 4th Floor (RM 404), Brooklyn, NY 11201, USA
Tel: (646) 997-0552, Fax: (646) 997-0560
E-mail: kaan.ozbay@nyu.edu

[4] Graduate Student, MSc.
Department of Civil and Environment Engineering
Old Dominion University
135 Kaufman Hall, Norfolk, VA 23529, USA
E-mail: yma004@odu.edu, Phone: (732) 781-6988

[5] Graduate Research Assistant
Department of Modeling, Simulation & Visualization Engineering
Old Dominion University
4700 Elkhorn Ave, Norfolk, VA 23529, USA
E-mail: zwang002@odu.edu, Phone: (757) 637-8868

Word count: 5,500 Texts + 7 Figures + 1 Table = 7,500 Texts
Abstract: 248
Resubmission Date: November 15, 2017

TRB18-03033
Paper submitted for Publication in the
Transportation Research Record, Journal of Transportation Research Board after being presented at the
Transportation Research Board's 97th Annual Meeting, Washington, D.C., 2018

ABSTRACT

Bike sharing systems are increasingly deployed in many cities worldwide. These systems provide a prominent solution for the first/last-mile problem with their cost-effectiveness and eco-friendliness. However, the use of bikes among stations is often spatiotemporally imbalanced, causing many problems in daily operations (e.g., rebalancing challenges). Thus, predictively knowing how the system demand evolves in advance helps improve the preparedness of operational schemes. This paper aims to present a predictive modeling approach to analyze the use of bicycles in bike sharing systems. Specifically, a deep learning approach using the convolutional neural networks (CNN) was proposed to predict the daily bicycle pickups at both city and station levels. A numerical study using data from the Citi Bike System in New York City was performed to assess the performance of the proposed approach. Other than the historical records, relevant information like weather was also incorporated in the modeling process. The modeling results show that the proposed approach can achieve better predictive performance in both city- and station-level analyses, confirming the merits of the proposed method against other baseline approaches. In addition, including information of neighboring stations into the models can help improve the performance in station-level prediction. The predictive performance of the CNN was also found to be related to parameters such as temporal window, number of neighboring stations, learning ratio, patch size, and inclusion of addition data such as drop-offs. Thus, the implementation of the proposed models requires necessary calibration to determine appropriate parameters for a given bike sharing system.

INTRODUCTION

Public bike sharing is becoming increasingly prevalent in many cities worldwide. Since the first launch in Europe in 1960s (*1*), bike sharing systems (BSSs) hit the streets in over 870 cities globally (*2*) and the number keeps increasing. These BSSs largely fill up the gaps in public transport modes by providing last mile connectivity effectively. Because of their convenience, low price, health and environment benefits, the shared bicycles are preferred by many travelers for short trips in urban areas (*3*). To further satisfy rider needs with a high-quality service, BSS operators often need to be proactively prepared for daily operation schemes in advance. This greatly motivates them to envision the forthcoming demand for the entire system as well as specific stations. Thus, a reliable forecast of their daily bike usage opens promising avenues for effective and economic system planning, operation (e.g., bike rebalancing), facility maintenance, etc.

Many existing BSSs are equipped with automatic rental systems to facilitate accessing and returning bicycles (*4*) . Through the rental systems, real-time records (e.g., start time and end time) of each trip become available. In addition, real-time data on available docks and bicycles at each station is also available. These data not only offer riders a convenient way to query bicycle information, but also facilitate operators to know their system performance responsively. Given the easy access to the system data, researchers have been leveraging the data to help operators understand their BSS usage predictively. Over the years, tremendous efforts have been made to the development of models for bike usage prediction. In general, these models attempted to use a set of determinants to explain BSS usage (e.g., pickups and / or drop-offs). For example, a recent study (*5*) analyzed the BSS in Montreal using meteorological data, temporal characteristics, and built environment attributes. Another one (*6*) predicted the bike demand in rush hours based on a linear regression model that includes taxi usage, weather, and spatial factors. The success of such models largely lies in the simplified assumptions on the statistical causal relationships between the variables. Despite the simplicity, these tractable mathematical models hardly capture the complexity of the changes in daily usage of BSSs. Thus, this motivates us to seek more reliable approaches that can work in the context of unclear causal relationships and limited explanatory variables.

Recently, deep neural networks have shown some impressive results on a variety of challenging tasks such as speech recognition, image classification, and natural language processing. The capability of deep neural networks to automatically learn complex and nonlinear patterns from observed data makes it applicable to a wide range of problems, including those related to transportation systems. For example, it was used to detect signal lights for improved navigation and advanced driver assistance systems (*7*). Meanwhile, given the accessibility of massive BSS records and other limited data (e.g., weather information), deep learning methods such as convolution neural network and deep belief network hold promise for performance improvements in analyzing the complex scenarios of BSS usage that may have nonlinear and heterogeneous patterns.

Thus, this paper aims to examine the potential of using deep learning techniques to forecast the daily usage of BSSs. It is believed that the deep learning (DL) approach which we explore in this paper is a useful and insightful way of fundamentally rethinking the prediction problem of bicycle usage. To be more specific, the predicted "usage" in this paper is considered as the number of pickups. Depending on the scope of analysis, analysts may consider other definitions of usage, for example, the number of trips between a certain pair of stations or the total number of pickups and drop-offs at a station.

The rest of this paper is structured as follows. The next section presents related work in previous studies. This is followed by the description of the proposed deep learning approach. The selected BSS and its operational data sets to test the proposed approach are then presented. Experimental results are described and discussed in the fifth section. Finally, concluding remarks and future research perspectives are provided in the last section.

LITERATURE REVIEW

The research on public bike sharing systems gains increasing popularity around the world. A number of studies have focused on several main research sub-streams, including strategic expansion of bike sharing systems (*8, 9*), demand analysis (*1, 6*), service level analysis (*10, 11*), and rebalance operation and vehicle routing (*12, 13*). The present paper is mainly concerned with the usage of BSSs and therefore relevant studies have been explored to facilitate understanding the modeling practices, achievements, and possible improvements.

There are a few studies that have focused on the prediction of BSS usage. These studies have investigated factors that may affect bike usage. Typical determinants related to weather, socio-demographics, economic attributes, transportation facilities, etc., were frequently identified (*5, 14, 15*). For example, Faghih-Imani et al. (*5*) developed multilevel linear mixed models to predict hourly bicycle pickups and drop-offs at stations of the Montreal BSS. The impact of meteorological data, temporal features, bicycle infrastructure, land use, and built environment attributes were analyzed. Hampshire and Marla (*16*) studied the impact of built environment on hourly pickups and drop-offs at the sub-city district level in two cities of Spain. The panel regression modeling results showed that BSS station density, capacity of stations, and number of points of interest were important variables related to bike usage. Singhvi *et al*. (*6*) used simple linear regression models to predict the pairwise trips of the Citi Bike during morning rush hours at both station and neighborhood levels. Covariates such as taxi usage, temporal, demographic, and weather factors were found to affect the number of bike trips. Likewise, Schneider (*17*) analyzed the Citi Bike trips from July 2013 to November 2015. The study proposed non-linear regression to model the relationship between bike usage and weather factors, including daily max temperature, precipitation, and snow depth. Many studies were conducted using daily (*15, 18-20*) or monthly (*14, 17, 21, 22*) aggregated data in modeling. In contrast, some others focused on the hourly usage of BSSs (*5, 23-25*). A fine-grained temporal resolution facilitates analyzing the variation of BSS usage in a more detailed manner.

The developed models were often focused on the activity (pickups and/or drop-offs) of the stations or more globally the state of the system. For example, Zhao *et al*. (*26*) used partial least squares (PLS) regression to predict the average daily usage of 69 bike sharing systems in China. It did not specifically analyze an individual system but to examine the impact of urban features and system characteristics on bike sharing ridership in general. Such macroscopic analysis will be more meaningful from the perspective of new system planning, but it fails to characterize the unique pattern of each system over time. Alternatively, a few studies modeled the BSS usage at city (system) level (*17, 23*), sub-city district level (*16*), neighborhood level (*6*), and station clusters (*27*). For example, Wang (*19*) predicted the city-level hourly bike rentals of the Citi Bike system. Borgnat *et al*. (*28*) predicted the number of rentals on a daily and hourly basis for the Vélo'v BSS in Lyon. Michau *et al*. (*29*) applied a parsimonious statistical regression model to relate social, demographic, and economic data of different neighborhoods with the bike trips of the Vélo'v BSS. Li *et al*. (*27*) developed a hierarchical prediction model to forecast the hourly number of bikes that will be rented from / returned to each station cluster. The bike sharing stations were grouped based on a bipartite clustering algorithm. The aggregation at a larger spatial scale in these studies can improve the model performance (*6*). However, the bike usage aggregated at these scales was less pertinent than using bike activity data at a station level. The spatial aggregation may group together stations with significant demand profiles but fail to capture fine-grained spatial effects, which in turn results in less reliable models for prediction (*24*).

To facilitate operational applications, thus, a few studies have attempted to statistically predict the bike usage at a station level. For example, Buck and Buehler (*18*) built the multivariate regression model to identify determinants of the average daily checkouts per station for the Capital BSS. Maurer (*21*) presented a regression model to identify determinants of monthly rentals by station for the Nice Ride BSS in Minnesota. Daddio (*22*) established a regression model to predict monthly checkouts by station of the Capital BSS. Likewise, Rixey (*14*) used a multivariate linear regression model to predict the monthly checkouts per station in three BSSs (Capital Bikeshare, Denver B-Cycle and Nice Ride). Other than the linear regression models, several studies also explored the possibilities of using more advanced statistical approaches to model the BSS usage. In study (*20*), it was found that regression models based on ordinary least squares estimators are not appropriate because the bicycle usage is usually not normally distributed. Alternatively, previous studies have developed negative binomial (NB) models to forecast BSS usage (*15,* *20*). In regard to excessive zero counts, Rudloff and Lackner (*24*) introduced hurdle models to predict the hourly bike usage at a station level with data from the Citybike Wien in Vienna. Despite its promising performance, the hurdle model is relatively complex and difficult to implement without automated variable selection procedure.

Other than the statistical modeling, several studies also sought to predict BSS usage using data mining techniques. For example, Li *et al*. (*27*) introduced the gradient boosting regression tree (GBRT) to estimate the total number of rented bikes of the Citi Bike and Capital Bikeshare. Liu *et al*. (*25*) proposed a meteorology similarity weighted k-nearest-neighbor (MSWK) models to predict hourly pickups. It computed the similarity measurements based on the weather information associated with each station. Then the weighted average trips of the days with top ranks were used to predict the pickups of the target day. They also developed an inter station bike transition (ISBT) model to predict the drop-offs. These models were found to outperform conventional approaches (e.g., using historical average as prediction). Zeng *et al*. (*30*) used support vector regression (SVR), decision tree (DT), and random forest (RF) to predict the usage of the Citi Bike. Considering that some stations may have limited training data, they also improved the models by including additional global features extracted by the gradient boosting decision tree (GBDT) and neural network (NN) algorithms. With a two-month test data set, the predictive performance of the augmented models was found to outperform the baseline models. Using similar data from the same BSS, Wang (*19*) tested the RF model against linear regression, DT and NN, and found that the RF model improved the predictive performance.

Despite the success in modeling bike usage with identified determinants, there were still many issues associated with current modeling practices. One important issue is associated with the subjective extraction of explanatory variables. Many studies used different spatial criteria to obtain the independent variables. For example, Buck and Buehler (*18*) linked each station to independent variables (e.g., population and bike lane supply) measured within a 0.5-mile buffer zone. In contrast, others used smaller buffer zones such as 400-meter ones (*21*, *22*) and 300-meter ones (*31*) to obtain similar variables (e.g., bus stops and length of bike infrastructure). However, as shown in (*31*), modeling results will be often sensitive to the range of the buffer zone. Unfortunately, there was no clear guidance to define the buffer range for data extraction with better quality. Another major issue is temporal inconsistency between the dependent variable and the explanatory variables. For example, bike usage statistics in 2014 were modeled in (*6*). However, taxi usage in 2013, population, and housing data from the 2010 US Census were used as covariates. Likewise, variables such as socio-economic attributes used in (*31*) were collected in 2013 and 2014 whereas the bike flow data were calculated in 2011. The temporal inconsistency makes it difficult to explain bike usage by covariates collected from different periods. Meanwhile, the need to synthesize extensive data sets often makes it difficult to collect them over years consistently (*20*). For example, many variables such as population are dynamically changing over time. From operational perspective, including historical survey data (e.g., historical population and/or jobs) cannot well reflect actual temporal variations at present. Last but not least, modeling issues related to multicollinearity, variable (feature) selection, determination of clusters, model transferability, etc., also limit the use of existing approaches.

The present paper builds upon the lessons learned from early empirical studies and contributes to the research community by developing a new approach for predicting bike usage of BSSs. The proposed approach mainly utilizes the considerably rich data from BSSs with limited external attributes to improve the prediction performance.

PROPOSED METHODOLOGY

Early studies made great efforts to establish statistical models for predicting BSS usage. However, many issues (e.g., challenges to obtain various variables) discussed above have limited their use, especially for operational applications. Rather, this study sought to take advantage of recent development in machine learning techniques to tackle the prediction problem. Specifically, a deep learning approach based on convolutional neural networks is proposed to predict the usage of BSSs. Prior to the description of convolutional neural networks, the approach based on neural network is also introduced as the basis.

Neural Network Model

As discussed in previous section, many statistical models cannot well quantify the bike usage of BSSs. One notable issue is that bike usage data are not normally distributed. In addition, the assumption on the linearity between bike usage and relevant variables is often violated. On the other hand, neural networks will be able to avoid such issues even when some data are skewed and nonlinear relationship exists. Neural networks consist of multiple layers of interconnected nodes (neurons) for learning tasks. These nodes are called perceptrons and mathematically resemble different linear regression (LR) models. The difference between a LR model and a perceptron is that the latter feeds the signal obtained from a LR model into an activation function that may or may not be linear. In general, given a neural network with N layers, the final output p of layer could be represented as follow:

$$p = f(W_n x_{n-1} + b_n) \tag{1}$$

where f is the activation function (e.g., sigmoidal function); W_n is the weighted coefficients of layer ; x_{n-1} is the output of layer $n-1$ and serves as the input of layer ; and b_n is the bias for layer . By quantifying the prediction error using an error measure e, the parameters of all layers of the network are jointly optimized to make the estimates approximate the desired output within the predefined error threshold ε or the maximum iteration I. The gradient descent algorithm can be used to determine how the parameters should be updated to reduce the prediction error. Assume there are I iterations, the parameters of each layer in i^{th} iteration are modified:

$$\mathrm{W}_n^i = \mathrm{W}_n^{i-1} - \eta \frac{\partial e(p^i, y)}{\partial \mathrm{W}_n^{i-1}} \tag{2}$$

$$\mathrm{b}_n^i = \mathrm{b}_n^{i-1} - \eta \frac{\partial e(p^i, y)}{\partial \mathrm{b}_n^{i-1}} \tag{3}$$

where η is the learning ratio that controls the step size and $i = 1, 2, ..., I$. By repeatedly taking small steps in the direction opposite to the gradient, the hidden layers will learn to capture the relationship between the input data and the target output data, and the output layer will predict the desired results based on the learned relationship.

However, when the dimension of the input is large, just using the activation function to approximate the relationship between the raw input and output is time-consuming and inefficient. To better capture the nonlinearity characteristics, the idea of deep learning is initiated to build models that represent data at multiple levels of abstraction and can discover accurate representations autonomously from the data themselves via various methods such as auto encoder, belief propagation, and convolution.

Convolutional Neural Network Model

Lately, the deep learning (DL) concept that originated from artificial neural network research has received increasing attention. Many research efforts have promoted its application in various areas such signal processing, artificial intelligence, etc., due to its achievable high performance. As such, various DL methods such as Deep Belief Network (DBN), Deep Boltzmann Machine (DBM), and Deep Neural Networks (DNN) have been introduced. Among them, the convolution neural network (CNN) under the DNN category has achieved the state-of-the-art performance in various applications (*32*). It is a subclass of neural networks with constrained connectivity patterns between some layers (*33*). Typically, they are very useful if input data exhibit certain patterns, for example, the ordering of image pixels in a grid and the temporal structures of an audio signal. Thus, considering the complex spatiotemporal patterns of bike usage of a BSS, this paper explores the possibility to use the CNNs for its predictive analysis.

As shown in FIGURE 1, a CNN contains three types of layers: convolutional layers, pooling layers, and hidden layers. The convolutional layer takes a stack of feature maps as input and convolves each of these with a set of learnable filters to produce a stack of output feature maps to form a deep model. In i^{th} iteration, this is efficiently implemented by replacing the matrix–vector product $W_n^{(i,j)} \otimes X_{n-1}^{(i)}$ in equation (4) with a sum of convolutions. The output feature map $X_n^{(i,j)}$ is represented as follows:

(4)

where represents convolution operation; the matrix represents the filters of layer ; and represents the bias for the feature map. Note that a feature map is calculated by computing a sum of convolutions with the feature maps of the previous layer that has a predefined patch size .

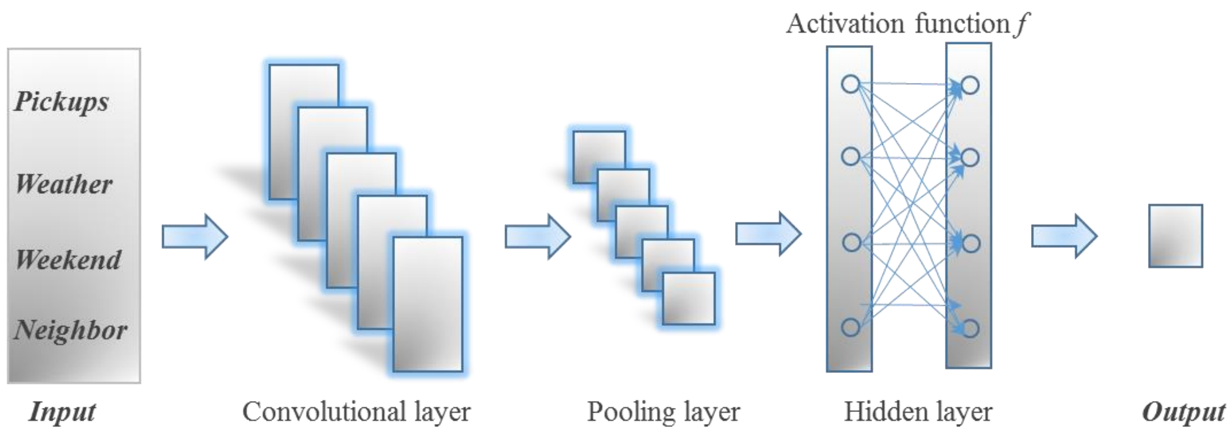

FIGURE 1. A schematic overview of a convolutional neural network.

The convolution operation takes advantage of the input structure and reduce the number of parameters needed during the training process. During the convolution, each perceptron is connected to the patched subset of perceptrons in previous layer using the following equation.

$$W_n^{(i,j)} \otimes X_{n-1}^{(i)} = \sum_{k=1}^{K} W_n^{(i,j,k)} \times X_{n-1}^{(i)} \quad \textbf{(5)}$$

where *k* denotes the number of convolution patches. This means that each perceptron detects feature across the input. With predefined patch size , applying feature detectors across the entire input enables to capture the pattern of the input data, for example, the archived bike usage data of a bike sharing system.

After modeling local relationships in input using convolutional layers, the dimensionality of the feature maps is further reduced by inserting a pooling layer. This allows the hidden layer to model correlations across a larger part of the input with a lower resolution. The pooling layer reduces the dimensionality of a feature map by averaging its outcomes across local regions of the input (*34*). Thus, the model is more invariant to small translations of the input, which provides the robustness to outliers. Unlike convolutional layers, pooling layers typically do not have any trainable parameters. The outputs of pooling layers are used as the input for hidden layers consisted of neural perceptrons, relevant connections, and activation functions. By applying the updating mechanism in equations (2) and (3), parameters are trained to fit the model.

Convolutional and pooling layers enable higher layers in the network to obtain a coarser representation of the input. Thus, the pattern of the relationships between variables such as weather and historical data and current daily usage can be easily modeled and learned. For implementation, the pseudo code of the CNN model for predicting the daily bike usage of a BSS is provided as follows:

Algorithm 1. Pseudo Code for Convolution Neural Network Prediction

Input: *Historical bike usage data* X *(pickups & weather information), learning ratio* η, *patch size ,error threshold* ε, *Maximum iteration* I, and *network layer number* N
Initialization*:* $i=1$*; layer coefficients*
While ($i<I$ ***) & (*** $e>\varepsilon$ ***)***
    For $j=1:J$
        Calculate convolutional output with given patch size
    End
    Calculate final output $p^i = f(W_n^i x_{n-1}^i + b_n^i)$
    For $n=1:N$
        Update using equation $\mathrm{W}_n^i = \mathrm{W}_n^{i-1} - \eta \frac{\partial e(p^i, y)}{\partial \mathrm{W}_n^{i-1}}$ *and* $\mathrm{b}_n^i = \mathrm{b}_n^{i-1} - \eta \frac{\partial e(p^i, y)}{\partial \mathrm{b}_n^{i-1}}$
    End
    Calculate *error* e
    $i{+}{+}$
End

NUMERICAL STUDY

This paper uses the data from the Citi Bike System in New York City (NYC) as a case study to test the performance of the proposed approach. The data sources and relevant analyses are provided below to facilitate understanding of the system as well as the predictive analysis in later sections.

Data Sources

As a privately-owned bike-share program in NYC, Citi Bike was launched in May 2013 with 6,000 bikes at 330 docking stations in Manhattan and parts of Brooklyn, and will be expanded to over 700 stations and 12,000 bicycles by the end of 2017[1]. It is one of the largest BSSs in the world. This paper uses it as a case study and tests the proposed prediction approach.

The trip data archived from 01/01/2015 to 09/30/2017 were used in following analysis. The historical trip data were downloaded from the official webpage of Citi Bike[2]. The total number of bike pickups in this period is 36,030,073. These pickups were from 767 stations operated during the analysis period. Like many BSSs, the archived trip data provide detailed information on Trip Duration (seconds), Start Time and Date, Stop Time and Date, Start Station Name, End Station Name, Station ID, Station Latitude and Longitude, Bike ID, User Type and Gender. Unlike existing studies that need to exhaustively obtain many variables related to the built environment, socio-economics, etc., the only external data sought in this study include weather and holiday information because they have been found to tightly relate to bike usage (*6, 17, 23, 35*). These external data are separately obtained from the Weather Underground website for environmental information[3] and the time and date website[4]. The daily maximum and minimum temperature, precipitation, snow, and rainfall measured by the weather station in New York, and weekend information are assembled with the daily pickups of each station as well as the entire system to test the model performance.

Descriptive Analysis

The collected data were analyzed to explore the characteristics of the system usage. The daily pickups of the study period are shown in FIGURE 2. FIGURE 2(a) illustrates the patterns of bike pickups for the

[1] http://a841-tfpweb.nyc.gov/dotpress/2014/10/citi-bike-program-in-new-york-city/#more-339
[2] https://www.citibikenyc.com/system-data
[3] https://www.wunderground.com/
[4] https://www.timeanddate.com/calendar/?year=2016&country=1

entire system. On average, the daily usage was 36,174 times. It should be noted that there was no data from January 23 to 26, 2016 due to the deadly Winter Storm Jonas passing over NYC. The black line in FIGURE 2(a) denotes the total daily pickup, and the red and green ones are rentals by male and female riders, respectively. The blue line in FIGURE 2(a) shows the usage of riders without gender information. These curves indicate that there was more bike usage in summer and fall but much less in winter.

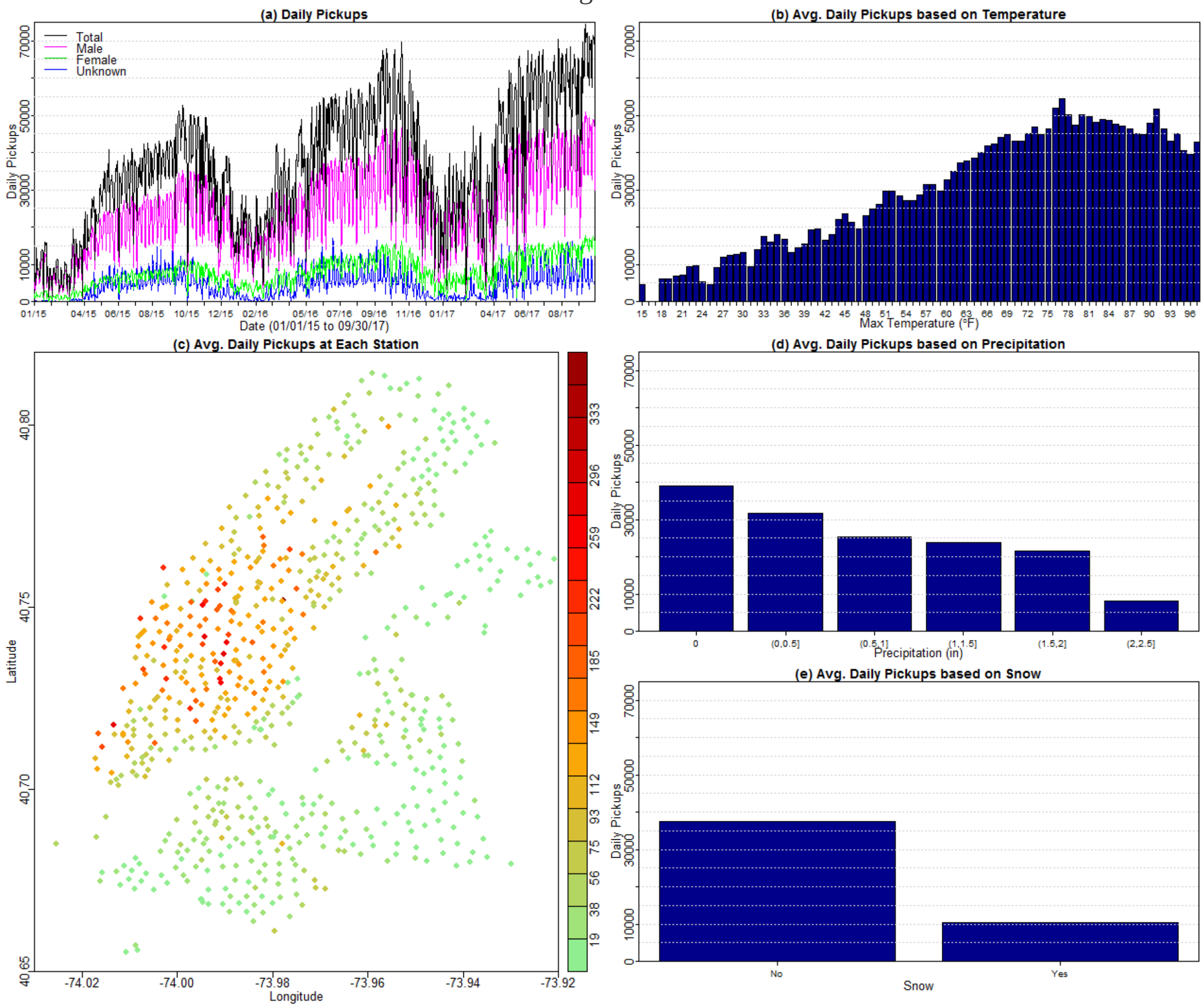

FIGURE 2. Descriptive analysis of Citi Bike pickups (01/01/2015 to 09/30/2017).

FIGURE 2(b) shows daily pickups with respective to the daily maximum temperature. It can be seen that pickups did not linearly change with the change of temperature. Overall, the bike usage shows an increasing trend when temperature gradually increases up to 77 degrees. The peak value of 54,399 times occur at 77 degrees. However, when the temperature is relatively high, the daily pickups tend to decrease. Such trend suggests that a linear regression model using temperature as a regressor will not work appropriately. FIGURE 2(c) shows the spatial distribution of the average daily pickups at different stations. It should be noticed that the distribution of pickups among stations is not uniform (e.g., darker (red) dots denote more frequently used stations). Heavily used stations are concentrated in the downtown and midtown of Manhattan. The most frequently used station (ID=519) on average had 370 pickups daily. The imbalanced distribution of the bike usages among stations should be attributable to the specific location, built environment, and social activities in the area. However, many of these factors (e.g., number of employees near each station; number of tourists) that quantify these attributes are not easy to (continuously) obtain. On the other hand, including some outdated survey data in analysis will lead to a fundamentally flawed prediction model. Thus, it would be more practical to use models with less data needs and can capture spatial variations of BSS usage day to day.

Rather than relying on the untraceable factors (e.g., daily visitors/population in a city), we examined bike usage under different weather conditions. FIGURE 2(d) shows the change of the average daily trips with respect to precipitation. The usage was noticeably higher for the days without rain (39,128 times). Nevertheless, there were still many pickups in the presence of high precipitation, i.e., (2, 2.5] inches. The average pickups with precipitation of (2, 2.5] inches greatly decreased to less than 10,000 times a day. Likewise, FIGURE 2(e) shows a clear relationship between the average daily pickups and the snowfall. The pickups dropped drastically for the days with snow. Specifically, the average pickups of snow days were less than one third of that of days without snow. Both the rainfall and snowfall have shown notable impact on the bike usage of the Citi Bike. Thus, it would be helpful to include these objective measurements in predicting the bike usage of the system.

RESULTS AND DISCUSSION

In this paper, we assembled the Citi Bike trip data along with the weather and holiday data to create the final data set for analysis. By applying the proposed CNN method, the data set was used to train the models and predict the usage of the bike sharing system. Specifically, data from January 1, 2015 to December 31, 2016 were used to train the CNN models so that the information of different periods and factors can be incorporated. Then the remaining 9-month data were used for testing. The data structure can be represented as $(\boldsymbol{x}_1, x_2, x_3, x_4, x_5, x_6, x_7, y_d, p_d)$, where $\boldsymbol{x}_1$ represents historical daily trips and it is a vector consisting of the daily pickups observed in a given time window (e.g., 1-week); x_2 and x_3 are maximum and minimum temperature (^oF) of each day, respectively; x_4 is daily precipitation (inch); x_5 denotes whether it was a rainy day; x_6 denotes whether it was a snowing day; x_7 is whether it is a holiday (weekend) or not; y_d is the actual pickups of d^{th} target day; and p_d is predicted pickups of d^{th} target day. Both city- and station-level models were developed, which will be detailed below. When developing the city-level models, the daily pickups of each station were aggregated to obtain $\boldsymbol{x}_1$ and y_d for the entire system.

In order to evaluate the performance of the proposed models, we choose three metrics that are frequently used in studies on BSS usage prediction (*6, 25, 30*). These metrics include mean absolute error (MAE), root-mean-squared-error (RMSE), and R^2 defined below:

$$MAE = \frac{1}{D}\sum_{i=1}^{D} | p_i - y_i | \quad \textbf{(6)}$$

$$RMSE = \sqrt{\frac{1}{D}\sum_{i=1}^{D} (p_i - y_i)^2} \quad \textbf{(7)}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{D} (y_i - p_i)^2}{\sum_{i=1}^{D} (y_i - \bar{y})^2} \quad \textbf{(8)}$$

where D is the total number of observations in the test data set; p_i is the predicted value; and y_i is the actual observed number of pickups. Different metrics focus on different aspects: (a) MAE directly explains by how many pickups the predictions are off; (b) RMSE focuses on the penalty of predictions far away from the actual observations; and (c) R^2 is the coefficient of determination.

City-level Prediction

The city-level CNN model was developed with an epoch of 40, patch size of 2, and a learning ratio of 1×10^{-5}. A detailed discussion on the choices of the parameters is presented in a later section. For comparison, the NN and ARIMA time series models were also developed. As mentioned earlier, the data of 2015 and 2016 were used to train each of the models and the 2017 data were used for testing the models. The modeling results based on the three approaches are shown in FIGURE 3(a): (i) red line is based on NN model; (ii) green line is based on the ARIMA model; (iii) blue line represents CNN modeling result; and (iv) black line denotes actual pickups for reference. The proposed models were implemented in Python3.5 with Intel(R) Core(TM) i5-6300U CPU @ 2.40 GHz. The training process using up to 2-year data needs less than five minutes for each epoch, depending on data size and model parameters. Once a model is trained, the prediction for the testing data only takes a few seconds.

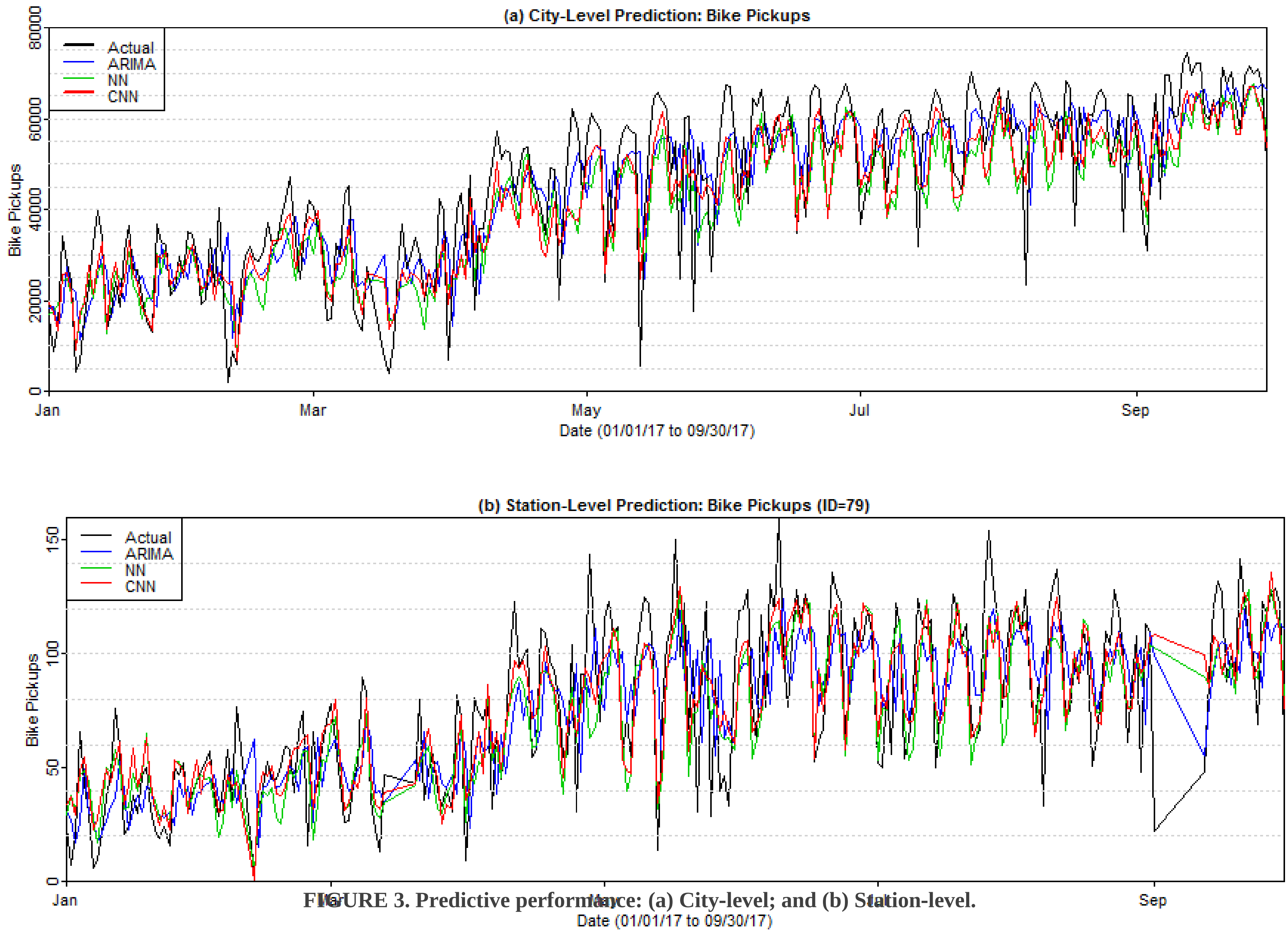

FIGURE 3. Predictive performance: (a) City-level; and (b) Station-level.

The modeling results show that the CNN model outperforms the other base models, with the time series model being worse. The corresponding MAE, RMSE, and R^2 of the NN model are 7,717.614 271.599 (mean sd.), 9,394.204 209.640, and 0.726 0.012, respectively. Likewise, the three performance measures for the ARIMA model are 8414.234, 11031.790, and 0.622, respectively. In contrast, the performance measures of the CNN-based model are 6613.445 246.165, 8168.797 163.519, and 0.793 0.008, respectively. The small MAE and RMSE, and large R^2 performance measures associated with the CNN model confirm that the proposed model improved prediction of the daily pickups at the city level.

Station-level Prediction

Based on the same modeling structure, the station-level modeling results for a station (ID=79) are illustrated in FIGURE 3(b). As does in city-level prediction, CNN performs better than the other models. Its MAE, RMSE, and R^2 are 14.264 0.201, 18.995 0.274, and 0.721 0.008, respectively. The baseline neural network method resulted in slightly worse results for this station, with MAE=14.887 0.231, RMSE=19.784 0.403, and R^2=0.698 0.012, respectively. Once again, the performance of the ARIMA model was less preferred: MAE=20.710, RMSE=26.604, and R^2=0.453.

Compared with the city-level data, the pickups associated with individual stations can have larger variations. For example, some stations may have fewer pickups but the others may have too many pickups. As mentioned in the literature review, many other attributes such as the employment and land use have been considered to account for the location-wise variations when predicting the usage of individual stations (*20*, *25*, *30*). However, some of these variables may dynamically change or cannot be accurately obtained day to day. Thus, it is difficult to incorporate their reliable measurements in the day-to-day prediction models. Alternatively, to further improve the performance of the station-level prediction, this paper proposes to include the spatial relationship between the target station and its neighbors.

Thus, for station-level prediction, we further assemble the information of neighboring stations to the data set of the target station. As an example, we have used five stations (ID = 545, 2021, 430, 270, and 408) as the target stations to test the proposed approach. For each station, we have tested the effect of including different number of nearest neighboring stations in the model. As shown in FIGURE 4, including additional neighbors in the model improves its performance in terms of the MAE reduction. Nevertheless, the marginal benefit is not obvious when three or more neighbors were included. This may be attributed to the fact that the entropy does not increase much with more redundant information of neighboring stations. Besides, adding too many neighbors has the high risk of including irrelevant information in the model. Thus, for further analysis, we include three nearest neighbors for station-level prediction. With their information, the R^2 values for the five target stations are between 0.5 and 0.8.

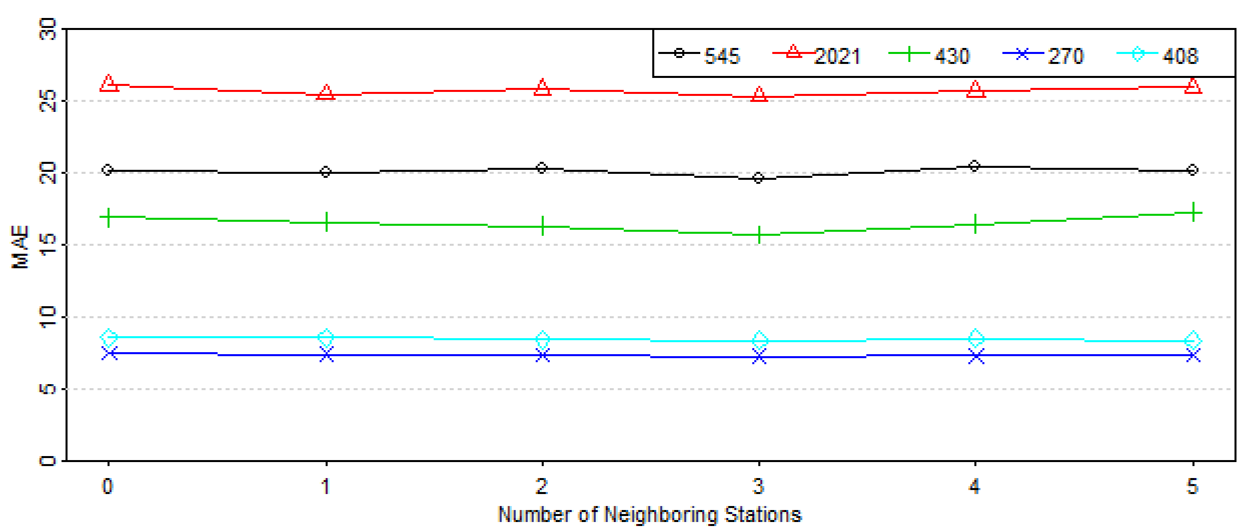

FIGURE 4. Model performance with different number of nearest neighboring stations.

Other than the five stations tested above, we also assessed the proposed CNN approach with additional stations that are distributed across different districts in NYC. The five districts include Downtown, Midtown and Uptown of Manhattan, Brooklyn, and Queens. In total, two stations were randomly selected from each district to cover both high and low usage scenarios. TABLE 1 presents the results when applying the CNN approach to each station with and without (three) neighboring stations. The results were based on the average of 5 simulations. Overall, it can be seen that the MAE, RMSE, and R^2 (mean ± sd.) for model with the nearest neighboring stations are consistently lower than those without neighbors. Therefore, it can be argued that the CNN approach including neighboring stations helps improve the predictive performance. In addition, the consistent findings among various stations suggest that proposed method is applicable to different stations.

TABLE 1. Station-level Predictive Performance

District	Station ID	Include Neighbors	MAE	RMSE	R^2
Downtown	79 (High)	Yes	14.015 0.228	18.876 0.213	0.725 0.006
		No	14.264 0.201	18.995 0.274	0.721 0.008
	408 (Low)	Yes	8.411 0.028	10.849 0.044	0.692 0.002
		No	8.604 0.118	10.977 0.177	0.685 0.010
Midtown	545 (High)	Yes	19.564 0.284	25.234 0.345	0.788 0.006
		No	20.200 0.455	25.672 0.531	0.781 0.009
	2021 (Low)	Yes	25.323 0.282	32.265 0.313	0.649 0.005
		No	26.097 0.534	32.769 0.227	0.638 0.005
Uptown	3164 (High)	Yes	28.587 0.702	35.932 0.866	0.609 0.019
		No	28.975 0.710	36.627 0.888	0.594 0.020
	305 (Low)	Yes	18.419 0.157	23.931 0.261	0.725 0.006
		No	19.137 0.224	24.092 0.232	0.722 0.005
Brooklyn	430 (High)	Yes	15.709 0.180	20.359 0.140	0.708 0.004
		No	16.924 0.549	21.686 0.505	0.669 0.015
	270 (Low)	Yes	7.304 0.123	9.309 0.142	0.526 0.015
		No	7.501 0.119	9.506 0.112	0.505 0.012
Queens	3124 (High)	Yes	8.022 0.068	11.222 0.115	0.560 0.009
		No	8.141 0.112	11.426 0.099	0.544 0.008
	3121 (Low)	Yes	5.329 0.080	8.660 0.142	0.418 0.019
		No	5.570 0.344	8.962 0.375	0.376 0.054

SENSITIVITY ANALYSIS

The performance of the proposed approach has been tested using in aforementioned case studies. In order to further verify its performance under different scenarios, we conducted sensitivity analysis on the impacts of different conditions, including scenarios with different temporal windows, model parameters, and additional input data.

Impact of adding drop-off data in models

This study examined the impact including drop-off data in the models. Other than the use of historical pickup data, historical drop-offs were also included in the input to predict the daily pickups of each bike stations. The corresponding MAE, RMSE, and R^2 of the extended models are shown in FIGURE 5. Each symbol in the figures represents the average value of 5 simulation runs for a selected station. Overall, they suggest that adding additional information on the drop-off records can slightly improve the performance of the proposed CNN models since most points in each figure are close to the reference line. Despite the slight improvement, this will bring extra burdens on data acquisition and processing, and require more computation efforts.

±

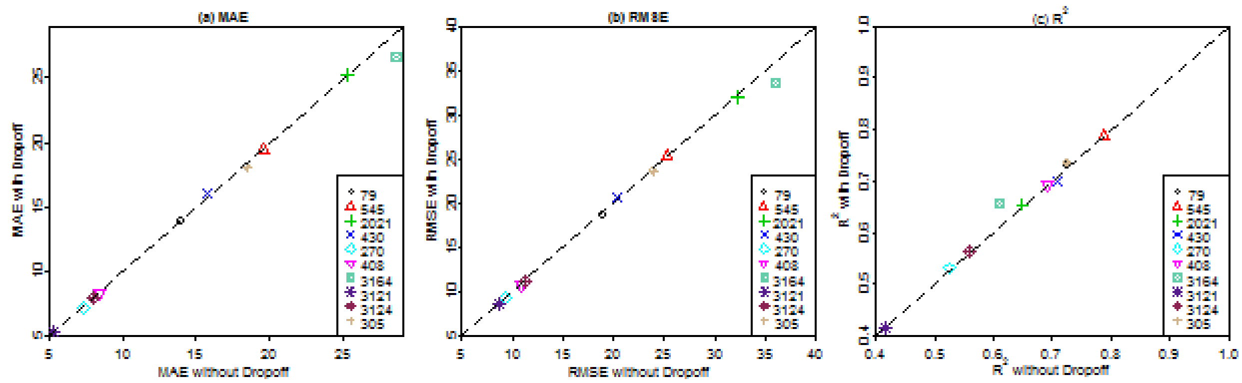

FIGURE 5. Impact of adding drop-off data as input in model development.

Impact of historical data

The analysis presented above used one-week historical data as the input for $\boldsymbol{x}_1$. To confirm this is an appropriate choice and analyze the impact of temporal window on the modeling results, we test five stations (ID = 545, 2021, 430, 270, and 408) considering different temporal windows. Herein $\boldsymbol{x}_1$ was varied from one week to four weeks. In other words, it used trips of the latest 7 to 28 days as the historical input to predict the target day's pickups. As shown in FIGURE 6, the predictive performance in terms of MAE, RMSE, and R^2 show that increasing temporal window led to degraded performance: MAE and RMSE increased with more weeks whereas R^2 decreased. Thus, this suggests that a temporal window of one week is sufficient to achieve relatively good performance.

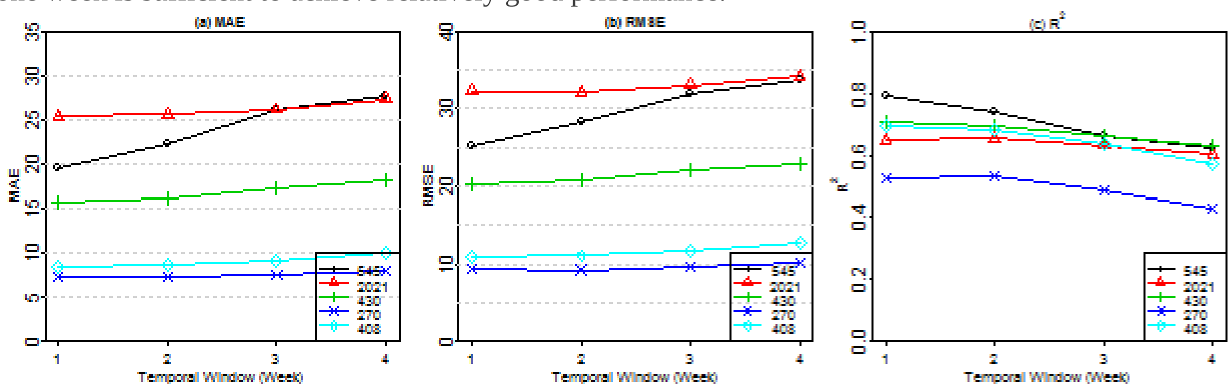

FIGURE 6. Impact of different temporal windows.

CNN parameter analysis

There are two key parameters that may affect the performance of CNN models: learning ratio and patch size. A high learning ratio may lead to quick approximation but degraded performance. A low learning ratio may achieve high performance but sacrifice computational resources. As to the patch size, it determines to which level the pattern is transformed through convolution. A large patch size means that more abstract feature representation of the neural network will be captured from the input. Thus, these two parameters need to be carefully considered to balance the performance and computation complexity.

As shown in FIGURE 7, different combinations of learning ratio and patch size were tested. The tested learning ratios include 1.0×10^{-5}, 5.0×10^{-5}, 1.0×10^{-4}, 1.5×10^{-4}, and 2.0×10^{-4}. The tested patch sizes include 2, 3, and 4. Note that we used one-week historical data, thus the patch size cannot be very large. FIGURE 7(a) shows the performance of city-level prediction. When the patch size was two and the learning ratio was 1.0×10^{-5}, the MAE reached the lowest value. Overall, increasing the patch size increased the MAE but reduced the training time due to high-level abstraction. The increase in the learning ratio led to higher MAE. Similar findings were found when testing the station-level prediction for different scenarios shown in FIGURE 7(b)-(d). The change of modeling performance with respect to these parameters suggests the necessity for calibration when implementing the proposed approach in other BSSs.

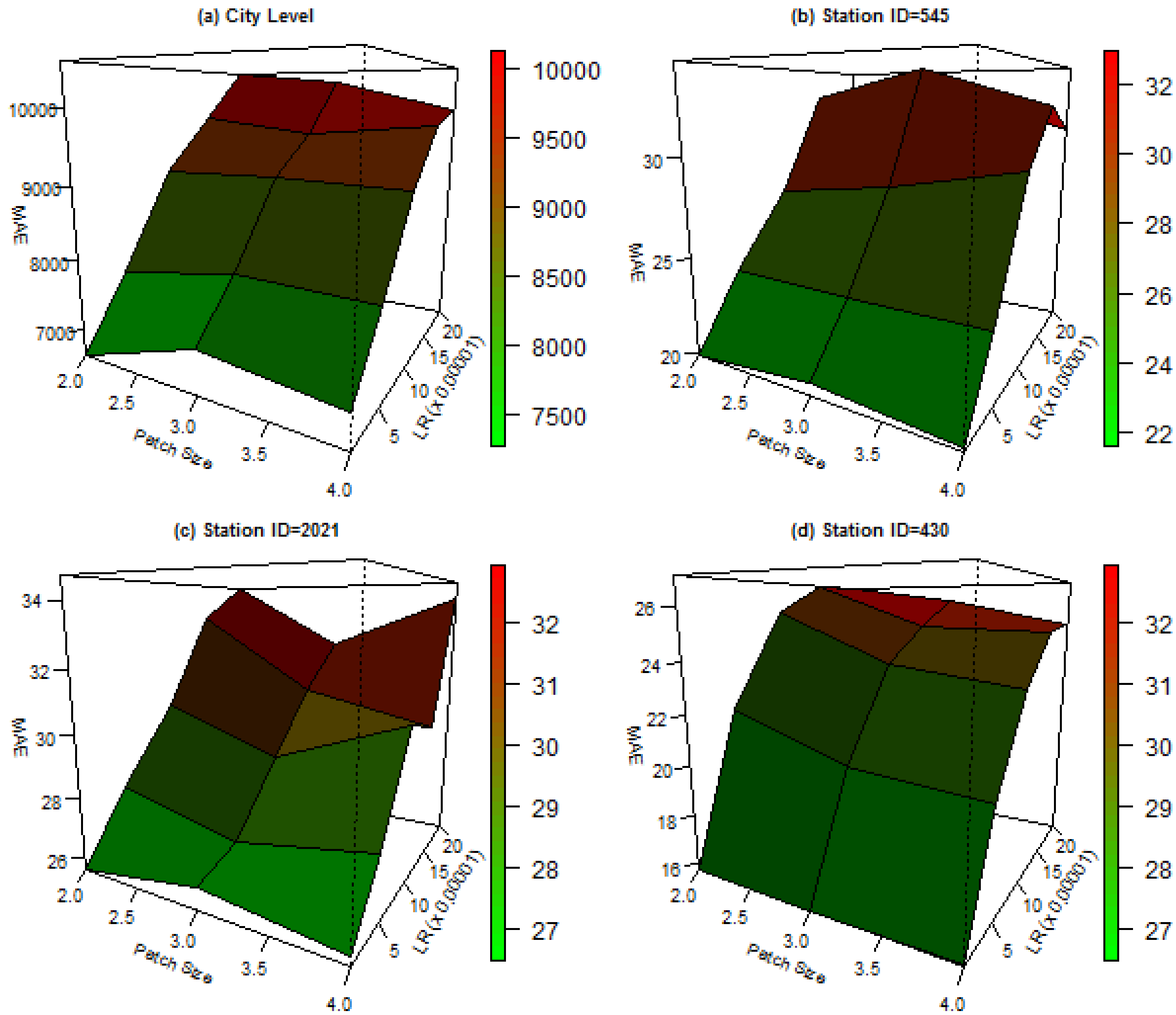

FIGURE 7. Performance changes with respect to different model parameters.

CONCLUSIONS

The prediction of BSS usage continues to attract growing interest. However, many existing models that use a large number of variables to explain the change of bike usage are not flexible enough to envision the real-world data due to the need for restrictive statistical modeling assumptions and exhaustive data acquisition among others. This paper proposed a deep learning approach based on the convolutional neural network model. The proposed approach applied historical BSS data and very limited external variables that can be measured objectively to predict daily bike usage of bike sharing systems. We obtained promising results based on the NYC Citi Bike system by demonstrating that both the city- and station-level predictive analyses can achieve better performance if the CNN model was used. The improved performance for models applied to multiple stations confirmed that the proposed approach is also applicable for different stations. The performance of the station-level models can be further improved if information about the nearest neighboring stations was included. In addition, the sensitivity analysis is necessary for the prediction of any BSS because it can help determine most practical and appropriate model parameters. Overall, the benefit of using CNN method lies in that it uses the convolution transformation and activation functions in neural networks to account for the complex nonlinear relationship between the target variable and the explanatory variables. Thus, it helps improve the predictive performance for modeling the unknown casual relationships between bike usage and various factors.

Despite the promising performance, this paper has used one typical BSS in a high-density urban area as the case study to evaluate the proposed prediction approach. A useful extension to this problem would be to verify it with data from additional BSSs, especially those of smaller systems. In addition, the comparison of the proposed approach with other deep learning algorithms is also suggested as future endeavors.

ACKNOWLEDGEMENTS
The authors appreciate the anonymous reviewers for their valuable comments that helped improve this paper. The contents of this paper only reflect views of the authors who are responsible for the facts and accuracy of the data and results presented herein. The contents of the paper do not necessarily reflect the official views or policies of any sponsoring agencies.

REFERENCES

1. Shaheen, S., H. Zhang, E. Martin, and S. Guzman, 2011. China's Hangzhou public bicycle: understanding early adoption and behavioral response to bikesharing. *Transportation Research Record: Journal of the Transportation Research Board*(2247), pp. 33-41.
2. Policy, E., 2015. Bike Sharing Systems around the world.
3. Shaheen, S., S. Guzman, and H. Zhang, 2010. Bikesharing in Europe, the Americas, and Asia: past, present, and future. *Transportation Research Record: Journal of the Transportation Research Board*(2143), pp. 159-167.
4. Raviv, T. and O. Kolka, 2013. Optimal inventory management of a bike-sharing station. *IIE Transactions* 45(10), pp. 1077-1093.
5. Faghih-Imani, A., N. Eluru, A.M. El-Geneidy, M. Rabbat, and U. Haq, 2014. How land-use and urban form impact bicycle flows: evidence from the bicycle-sharing system (BIXI) in Montreal. *Journal of Transport Geography* 41, pp. 306-314.
6. Singhvi, D., S. Singhvi, P.I. Frazier, S.G. Henderson, E. O'Mahony, D.B. Shmoys, and D.B. Woodard, 2015. Predicting Bike Usage for New York City's Bike Sharing System. In: Proceedings of the AAAI Workshop: Computational Sustainability.
7. John, V., K. Yoneda, B. Qi, Z. Liu, and S. Mita, 2014. Traffic light recognition in varying illumination using deep learning and saliency map. In: Proceedings of the Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on, pp. 2286-2291.
8. García-Palomares, J.C., J. Gutiérrez, and M. Latorre, 2012. Optimizing the location of stations in bike-sharing programs: a GIS approach. *Applied Geography* 35(1), pp. 235-246.
9. Parkes, S.D., G. Marsden, S.A. Shaheen, and A.P. Cohen, 2013. Understanding the diffusion of public bikesharing systems: evidence from Europe and North America. *Journal of Transport Geography* 31, pp. 94-103.
10. Raviv, T., M. Tzur, and I.A. Forma, 2013. Static repositioning in a bike-sharing system: models and solution approaches. *EURO Journal on Transportation and Logistics* 2(3), pp. 187-229.
11. Massink, M. and R. Pakauskas, 2015. Model-based assessment of aspects of user-satisfaction in bicycle sharing systems. In: Proceedings of the Intelligent Transportation Systems (ITSC), 2015 IEEE 18th International Conference on, pp. 1363-1370.
12. Chemla, D., F. Meunier, and R.W. Calvo, 2013. Bike sharing systems: Solving the static rebalancing problem. *Discrete Optimization* 10(2), pp. 120-146.
13. Schuijbroek, J., R.C. Hampshire, and W.-J. Van Hoeve, 2017. Inventory rebalancing and vehicle routing in bike sharing systems. *European Journal of Operational Research* 257(3), pp. 992-1004.
14. Rixey, R., 2013. Station-level forecasting of bikesharing ridership: Station Network Effects in Three US Systems. *Transportation Research Record: Journal of the Transportation Research Board*(2387), pp. 46-55.

15. Wang, X., G. Lindsey, J.E. Schoner, and A. Harrison, 2016. Modeling bike share station activity: Effects of nearby businesses and jobs on trips to and from stations. *Journal of Urban Planning and Development* 142(1), pp. 04015001.
16. Hampshire, R.C. and L. Marla, 2012. An analysis of bike sharing usage: Explaining trip generation and attraction from observed demand. In: Proceedings of the 91st Annual meeting of the transportation research board, Washington, DC, pp. 12-2099.
17. Schneider, T.W., 2017. A Tale of Twenty-Two Million Citi Bike Rides: Analyzing the NYC Bike Share System, http://toddwschneider.com/, Access Junly 20, 2017.
18. Buck, D. and R. Buehler, 2012. Bike lanes and other determinants of capital bikeshare trips. In: Proceedings of the 91st Transportation research board annual meeting.
19. Wang, W., 2016. Forecasting Bike Rental Demand Using New York Citi Bike Data. Dublin Institute of Technology.
20. Morency, C., M. Trépanier, A. Páez, H. Verreault, and J. Faucher, 2017. Modelling Bikesharing Usage in Montreal over 6 years. Working Paper, CIRRELT-2017-33. Interuniversity Reserach Centre on Enterprise Networks Logistics and Transportation (CIRRELT).
21. Maurer, L.K., 2012. Feasibility study for a bicycle sharing program in Sacramento, California. 91st Annual Meeting of the Transportation Research Board, Washington DC.
22. Daddio, D.W., 2012. Maximizing bicycle sharing: an empirical analysis of capital bikeshare usage. University of North Carolina at Chapel Hill.
23. Gebhart, K. and R.B. Noland, 2013. The impact of weather conditions on capital bikeshare trips. In: Proceedings of the Transportation Research Board 92nd Annual Meeting.
24. Rudloff, C. and B. Lackner, 2014. Modeling demand for bikesharing systems: neighboring stations as source for demand and reason for structural breaks. *Transportation Research Record: Journal of the Transportation Research Board*(2430), pp. 1-11.
25. Liu, J., L. Sun, W. Chen, and H. Xiong, 2016. Rebalancing Bike Sharing Systems: A Multi-source Data Smart Optimization. In: Proceedings of the Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1005-1014.
26. Zhao, J., W. Deng, and Y. Song, 2014. Ridership and effectiveness of bikesharing: The effects of urban features and system characteristics on daily use and turnover rate of public bikes in China. *Transport Policy* 35, pp. 253-264.
27. Li, Y., Y. Zheng, H. Zhang, and L. Chen, 2015. Traffic prediction in a bike-sharing system. In: Proceedings of the Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 33.
28. Borgnat, P., P. Abry, P. Flandrin, C. Robardet, J.-B. Rouquier, and E. Fleury, 2011. Shared bicycles in a city: A signal processing and data analysis perspective. *Advances in Complex Systems* 14(03), pp. 415-438.
29. Michau, G., C. Robardet, L. Merchez, P. Jensen, P. Abry, P. Flandrin, and P. Borgnat, 2011. Peut-on attraper les utilisateurs de vélo'v au lasso. In: Proceedings of the Proceedings of the 23e Colloque sur le Traitement du Signal et des Images. GRETSI-201, pp. 46-50.
30. Zeng, M., T. Yu, X. Wang, V. Su, and L.T. Nguyen, 2016. Improving Demand Prediction in Bike Sharing System by Learning Global Features.
31. Tran, T.D., N. Ovtracht, and B.F. d'Arcier, 2015. Modeling bike sharing system using built environment factors. *Procedia CIRP* 30, pp. 293-298.
32. Yan, Y., M. Chen, M.-L. Shyu, and S.-C. Chen, 2015. Deep learning for imbalanced multimedia data classification. In: Proceedings of the 2015 IEEE International Symposium on Multimedia (ISM), pp. 483-488.
33. LeCun, Y., L. Bottou, Y. Bengio, and P. Haffner, 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11), pp. 2278-2324.
34. Dieleman, S., K.W. Willett, and J. Dambre, 2015. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *Monthly notices of the royal astronomical society* 450(2), pp. 1441-1459.

35. Noland, R.B., M.J. Smart, and Z. Guo, 2016. Bikeshare trip generation in New York city. *Transportation Research Part A: Policy and Practice* 94, pp. 164-181.