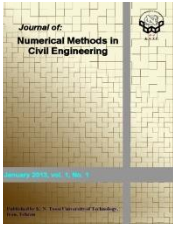


Numerical Methods in Civil Engineering

Predicting traffic flow between bike-sharing system stations: A case study of Chicago

Ali Behroozi*, Ali Edrisi**

ARTICLE INFO

RESEARCH PAPER

Article history:
Received:
April 2024
Revised:
February 2025
Accepted:
April 2025

Keywords:
Bike-sharing
Traffic flow
Machine learning
Gated recurrent unit
Multilayer perceptron

Abstract:

Active transportation systems, such as bike-sharing systems, offer several advantages, notably their integration with public transportation networks, pollution reduction, congestion alleviation, and decreased fuel consumption. However, a major challenge for shared bicycle companies is the efficient redistribution of bikes to ensure balanced availability across stations. Predicting station demand at various times is crucial for achieving this balance. To address this, we propose a framework leveraging data from Chicago's shared bicycle system and employing modern machine-learning techniques to forecast station demand throughout the day. In this research, we utilize features such as weather, accessibility level for each station, and historical transactions for the accurate prediction of traffic flow. Specifically, we compare two parallel multilayer perceptron deep learning models, incorporating matrix factorization and gate recurrent unit (GRU) neural networks. Furthermore, this research compares two performance models for predicting traffic flow. This research not only aids in optimizing bicycle distribution but also lays the groundwork for predicting demand in other public transportation systems, such as subways and buses, utilizing smart card technology.

1. Introduction

The advent and evolution of technology have facilitated the storage and utilization of vast amounts of data, including electronic records from public transportation systems. This wealth of data enables us to analyze travel patterns based on location surveys, providing valuable insights into usage patterns essential for optimizing transportation systems.

The bicycle serves as a key component of active transportation infrastructure. In both medium-sized and large cities, bicycles are utilized for various purposes, including daily commutes, occasional trips, and recreational outings. The benefits of cycling extend across multiple levels: societal, local, and individual.

One crucial aspect of transportation system planning involves forecasting demand and traffic flow. In shared bike-sharing, a significant challenge lies in maintaining a balanced distribution of bicycles across stations, crucial for optimizing resources based on estimated demand variations throughout the day. Presently, numerous cities worldwide provide access to data on their shared bike systems through open platforms on their websites.

One crucial aspect of transportation system planning involves forecasting demand and traffic flow. In shared bike-sharing, a significant challenge lies in maintaining a balanced distribution of bicycles across stations, crucial for optimizing resources based on estimated demand variations throughout the day. Presently, numerous cities worldwide provide access to data on their shared bike systems through open platforms on their websites.

This research aims to leverage recorded data from shared bike-sharing systems, along with weather data. Initially, we will conduct a comprehensive analysis of the data, followed by calculating access levels and employing machine learning models, including deep learning models, to predict demand levels at stations and payment structures. Ultimately, we will compare the predictive capabilities of these models. The primary objective of this study is to utilize and implement novel machine learning models for demand prediction. A practical application of this research is to establish a framework for forecasting travel demand in other public transportation systems, such as subways and buses, which utilize smart (magnetic) cards.

The proliferation of shared bicycle schemes globally underscores their significance in promoting healthier lifestyles and environmentally friendly modes of

---

** M.Sc. in Transportation Engineering in K. N. Toosi University of Technology*
***Corresponding author: Ali Edrisi: Associate Professor in K. N.Toosi University of Technology, Transportation Department, Civil Faculty*

https://doi.org/10.61186/NMCE.2404.1053

transportation, gradually assuming a pivotal role in urban mobility. These systems, bolstered by technologies like geographic information systems, enhance the management and efficiency of urban transportation fleets.

In broad terms, bicycle sharing distribution systems (BSS) fall into two main categories: station-based and dockless (float) systems. Station-based BSS involves renting bikes from designated stations and returning them to another station upon completion of use. Cities like New York City, San Francisco, Chicago, and DC exemplify this model. On the other hand, dockless BSS offers users greater flexibility by allowing them to pick up and drop off bikes at any location within a designated area. While this freedom enhances user experience, it can pose operational challenges such as sidewalk obstruction and the necessity for station repositioning to maintain adequate spacing.

As the shared bicycle system rapidly expands, its global significance in promoting healthier lifestyles and reducing environmental impact vis-à-vis motorized transportation is increasingly recognized. This transformation positions it as an emerging urban intermediary. Leveraging technologies like geographic information systems, shared bike-sharing systems enhance the management and efficiency of urban transportation fleets, contributing to their ongoing evolution and integration within urban landscapes.

One of the biggest challenges in managing bike-sharing at each station is rebalancing the fleet. On the other hand, advancements in technology allow us to store vast amounts of GIS and transaction data. Additionally, a comprehensive model is needed to predict travel demand at various temporal and spatial resolutions.

An ongoing challenge in enhancing bicycle sharing systems' performance lies in accurately predicting travel demand. This entails forecasting usage patterns based on previous sequences at specific times. The objective of this research is to develop a system capable of predicting sequences for each period based on input time characteristics [1-6].

In summary, our research proposes a framework leveraging machine learning models such as Multi-Layer Perceptron (MLP) and Gate Recurrent Unit (GRU) to predict short-term traffic flow in bike-sharing systems. By incorporating weather conditions, access data, and bike-sharing transactions, we capture spatial and temporal influences between stations.

We evaluate our approach using simulated data from Chicago's bike-sharing system, finding that the gate graph convolution model outperforms other models, exhibiting lower RMSE and MSE. Our experiments highlight the framework's capability to forecast short-term traffic flow effectively, surpassing conventional deep learning methods. Furthermore, our study reveals significant insights into how urban form shapes bike-share traffic behavior across various spatial and temporal contexts, underscoring the importance of spatio-temporal prediction in transportation forecasting.

In this research, the initial focus will be on reviewing previous studies related to demand forecasting in public transportation systems. Subsequently, the collected data will be analyzed in Chapter Three, where the method and model utilized in this research will also be elaborated upon, along with the presentation of results. Chapter Four will evaluate the performance of the model.

2. Literature review

The objective of forecasting demand at the city level is to anticipate the volume of bicycle usage across all stations within the city. In 2014, Kaggle, a leading platform in the field of data science, launched a challenge that tasked participants with analyzing and modeling data from Washington City's bike-sharing system to predict total demand within a 24-hour timeframe [7]. This demand forecasting was conducted in 24-hour intervals, utilizing data from the Capital Bike Share system at the city level. Various machine learning algorithms were tested in this challenge, including shrinkage regression, AdaBoost regression, support vector regression, random prediction trees, and gradient boosting regression trees. Results indicated that the AdaBoost algorithm outperformed other algorithms [8]. While forecasting total demand across all shared bicycle stations at a city scale may not directly address the challenge of bicycle rebalancing among stations, it serves to significantly simplify the issue.

Numerous studies have utilized time-related factors such as hours, days, months, days of the week, and holidays, along with meteorological variables like temperature, precipitation, and wind speed, to forecast demand in shared bike-sharing systems. Yang et al. introduced a novel approach rooted in deep learning, termed spatio-temporal graph convolutional neural networks, for analyzing recorded subway data. This method primarily predicts two types of volumes: inbound and outbound trips at each station within a city's subway network. Rather than representing metro stations as nodes in the network and employing conventional neural convolutional networks, this approach transforms the city's metro network into a graph to capture spatial-temporal dependencies and utilizes convolutional neural networks for prediction [10].

Yang Hyun Seo et al. conducted a study aimed at forecasting demand at the station level for both picking up (rental) and dropping off bicycles within the bicycle-sharing system. They utilized recorded trip data from the stations, along with the count of bicycles picked up and dropped off, as well as time and weather data. For prediction, the study employed the random forest machine learning technique, considering variables from 1 to 3 hours prior to the prediction time interval at each station [9].

Jiangu Chen et al. introduced a paper outlining a dynamic bicycle station planning system designed to offer optimal solutions for networked bicycle sharing systems. The system consists of several components: clustering bicycle drop-off locations, modeling bicycle stations, predicting bicycle station locations, and recommending bicycle station capacity. In the section addressing abandoned bicycle locations, the authors utilize large-scale cycling data to spatio-temporally cluster selected stations. The graph modeling segment for bicycle stations involves filtering and clustering stations with lower utility and value in a weighted manner. Subsequently, the authors construct and analyze a graph model as a time series. For bicycle station location prediction, the authors employ a graph convolutional neural network model with a gate mechanism. This model is trained using dynamic time series data to predict stations in the subsequent time period [10].

Wang et al. introduced a novel deep learning approach termed spatio-temporal graph convolutional neural networks for analyzing recorded subway data. This method primarily predicts two types of volumes: inbound and outbound trips at each station within a city subway. Specifically, rather than representing metro stations through a network and employing convolutional neural networks, the authors transform the city's metro network into a graph. This allows them to capture spatial-temporal dependencies and utilize convolutional neural networks for prediction [11].

Yang et al. This study focuses on short-term demand forecasting for bike-sharing systems, crucial for managing infrastructure and balancing fleets amid fluctuating travel patterns. While machine learning has advanced in demand prediction, limited research has explored feature engineering for model selection. The study introduces time-lagged graph-based features (e.g., Out-strength, In-strength, Out-degree, In-degree, and PageRank) derived from real-world bike usage data. These features were tested with machine learning models (XGBoost, MLP, LSTM), showing that graph-based attributes outperform traditional meteorological data. Deep neural networks proved most effective in handling time-lagged sequences, leading to more accurate forecasts [12].

Edrisi et al. employed gated graph convolutional networks and graph attention neural networks to predict travel demand in a bike-sharing system while incorporating accessibility factors such as population and employment access. By integrating weather and transaction data, their model effectively captured the spatial and temporal dependencies influencing bike usage patterns in Chicago's Divvy system. This study highlights the role of accessibility in demand forecasting, providing a relevant perspective on measuring accessibility levels in transportation research [13].

3. Case Study

One crucial aspect of data modeling involves both data analysis and visualization, alongside understanding the features utilized within the model. This chapter aims to delineate the requirements and prerequisites for establishing an effective method and tool for analyzing recorded data from the shared bicycle system. Ultimately, it seeks to furnish a valuable resource for policymaking within the shared bicycle system. Focusing on the city of Chicago as the study area, this chapter delves into the collected data pertaining to the recorded trips of the bicycle sharing system. It examines the most significant and influential parameters within this dataset.

An initial area of investigation entails understanding the behavior of shared bicycle users throughout the day. As depicted in Figure 1, the peak demand occurs during the period from the 7th to the 9th months, corresponding to the warmer months of the year. In other words, this period witnesses the highest number of shared bicycle trips. Additionally, an essential aspect of planning involves prioritizing bike stations according to demand, as illustrated in Figure 2. According to Figure 2, 19 stations had more than 1000 trips per hour on average.

Figure 3 illustrates a negative correlation between wind speed and bike usage. As wind speed increases, the demand for the bike-sharing system declines, indicating reduced user willingness under windy conditions. As depicted in Figure 4, the coastal areas exhibit the highest density of stations, correlating with the highest number of trips occurring in these regions.

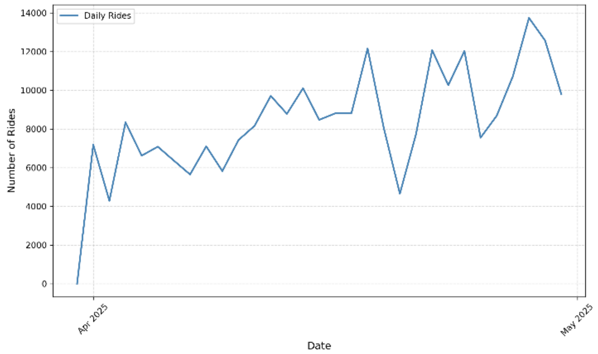


Fig. 1: Demand for Bike-Sharing by Day

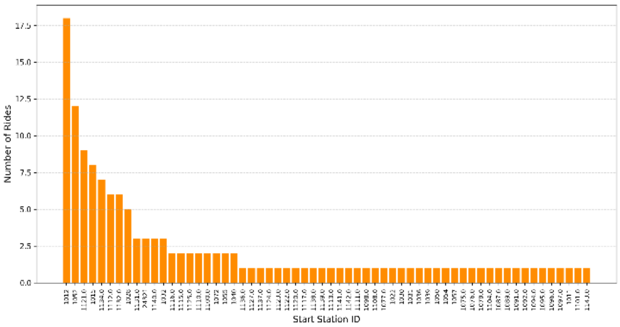


Fig. 2: Average Bicycle Demand per Station

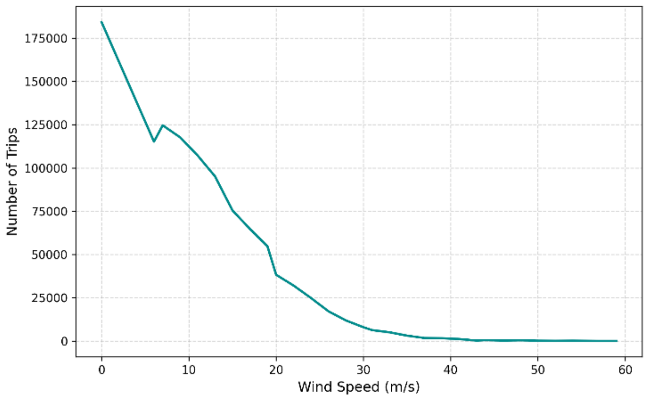


Fig. 3: The Impact of Wind Speed on Demand

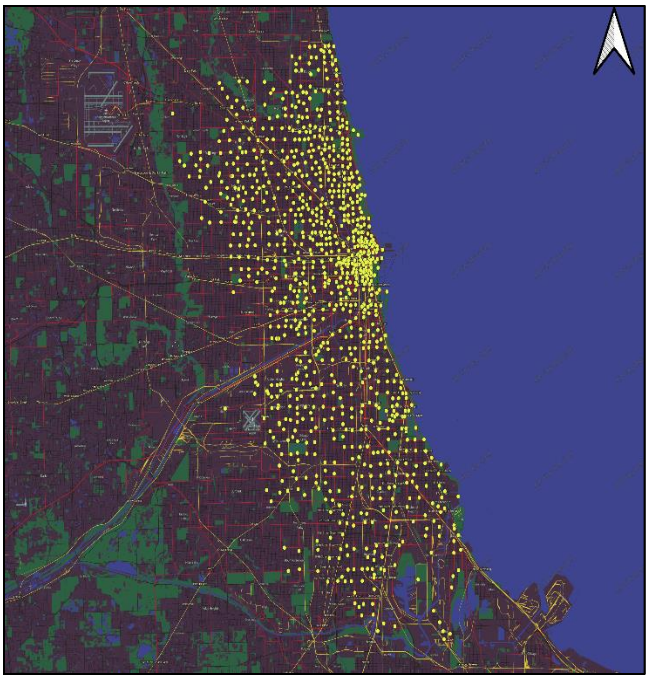
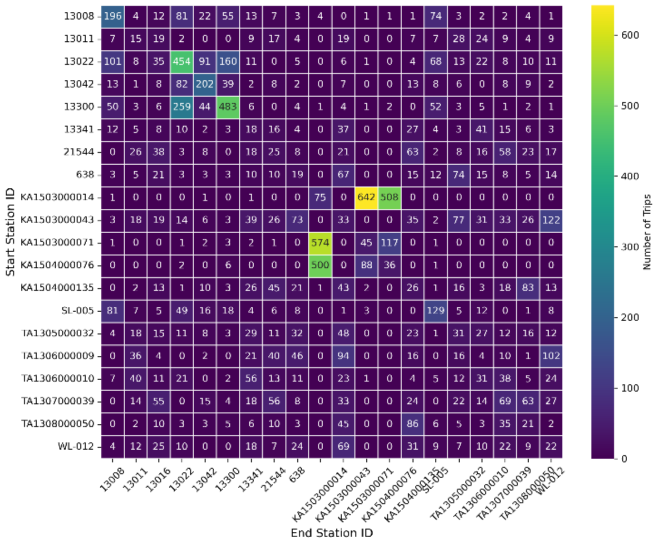


Fig. 4: Dispersion of Bike-Sharing Stations

The analysis of spatial dependence between stations is a critical aspect of demand analysis and forecasting. This examination identifies stations with the highest correlation. The Heatmap illustrates this correlation.

According to Figure 5, the higher the number of trips between two stations, the greater the mutual influence between them. Based on the diagram, stations 76 and 35 exhibit the most significant impact on each other.

Accessibility in urban applications refers to the ease with which individuals can reach various destinations and services within a city. It encompasses not only physical mobility but also factors such as affordability, safety, and convenience. One crucial aspect of urban accessibility is transportation, which includes public transit, cycling infrastructure, pedestrian pathways, and road networks. Improving transportation accessibility entails enhancing the efficiency, reliability, and inclusivity of these systems to ensure that all residents, regardless of their socioeconomic status or physical abilities, can access essential services, employment opportunities, education, and recreational facilities within a reasonable timeframe.



	13008	13011	13016	13022	13042	13300	13341	21544	638	KA1503000014	KA1503000043	KA1503000071	KA1504000076	KA1504000135	SL-005	TA1305000032	TA1306000010	TA1307000039	TA1308000050	WL-012
13008	196	4	12	81	22	55	13	7	3	4	0	1	1	1	74	3	2	2	4	1
13011	7	15	19	2	0	0	9	17	4	0	19	0	0	7	7	28	24	9	4	9
13022	101	8	35	454	91	160	11	0	5	0	6	1	0	4	68	13	22	8	10	11
13042	13	1	8	82	202	39	2	8	2	0	7	0	0	13	8	6	0	8	9	2
13300	50	3	6	259	44	483	6	0	4	1	4	1	2	0	52	3	5	1	2	1
13341	12	5	8	10	2	3	18	16	4	0	37	0	0	27	4	3	41	15	6	3
21544	0	26	38	3	8	0	18	25	8	0	21	0	0	63	2	8	16	58	23	17
638	3	5	21	3	3	3	10	10	19	0	67	0	0	15	12	74	15	8	5	14
KA1503000014	1	0	0	0	1	0	1	0	0	75	0	642	508	0	0	0	0	0	0	0
KA1503000043	3	18	19	14	6	3	39	26	73	0	33	0	0	35	2	77	31	33	26	122
KA1503000071	1	0	0	1	2	3	2	1	0	574	0	45	117	0	1	0	0	0	0	0
KA1504000076	0	0	0	2	0	6	0	0	0	500	0	88	36	0	1	0	0	0	0	0
KA1504000135	0	2	13	1	10	3	26	45	21	1	43	2	0	26	1	16	3	18	83	13
SL-005	81	7	5	49	16	18	4	6	8	0	1	3	0	0	129	5	12	0	1	8
TA1305000032	4	18	15	11	8	3	29	11	32	0	48	0	0	23	1	31	27	12	16	12
TA1306000009	0	36	4	0	2	0	21	40	46	0	94	0	0	16	0	16	4	10	1	102
TA1306000010	7	40	11	21	0	2	56	13	11	0	23	1	0	4	5	12	31	38	5	24
TA1307000039	0	14	55	0	15	4	18	56	8	0	33	0	0	24	0	22	14	69	63	27
TA1308000050	0	2	10	3	3	5	6	10	3	0	45	0	0	86	6	5	3	35	21	2
WL-012	4	12	25	10	0	0	18	7	24	0	69	0	0	31	9	7	10	22	9	22

Fig. 5: Station Correlation Based on Travel

In addition to transportation, urban accessibility also encompasses the availability of essential services and amenities within close proximity to residential areas. This includes access to healthcare facilities, grocery stores, schools, parks, and community centers. Creating accessible urban environments involves urban planning strategies that prioritize mixed land use, compact development, and the provision of amenities within walking distance of residential neighborhoods. Furthermore, technology plays a crucial role in enhancing accessibility through applications such as ride-sharing services, mobility-as-a-service platforms, real-time transit information systems, and smart city initiatives that optimize traffic flow and improve the overall efficiency of urban infrastructure. By prioritizing accessibility in urban planning and leveraging technological advancements, cities can create more equitable and sustainable environments that enhance the quality of life for all residents [17, 18].

One aspect that has received limited attention in previous literature is the influence of each station's access level on its demand. Consequently, statistical data concerning the population and geographic distribution of businesses across the city of Chicago were gathered in the form of a CSV file. Using the OTP open-source software in walking mode over a 15-minute period, clouds were formed at each station, indicating the level of access at each location. This information was derived from Figure 6.

According to Figure 7, there is minimal correlation between the level of access based on the population and the demand volume, with a correlation coefficient of 0.26. According to Figure 8, there is a positive correlation between the level of access to jobs at each station and the demand level at that station. The correlation between the level of access and the level of demand is 0.58.

The use of accessibility parameters for each station is based on the assumption that a 10-minute walking distance defines a station's effective catchment area. Within this area, we assume that both the number of employees and the residential population contribute to the station's accessibility. This implies that the commuting patterns and travel demand at each station are influenced by the density of people living and working within this predefined radius.

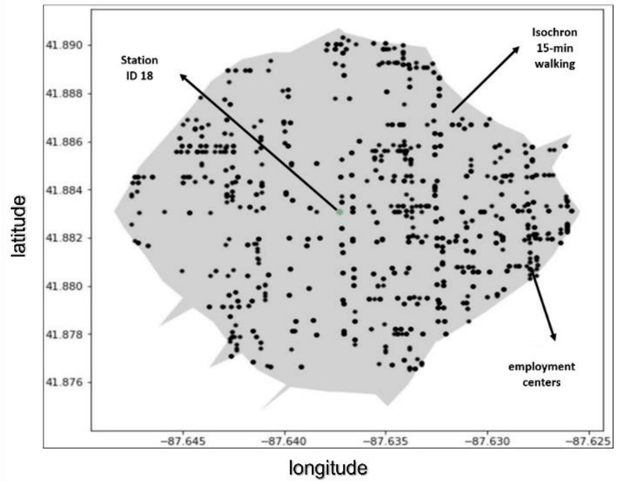


Fig. 6: Level of access to employment at Station 18

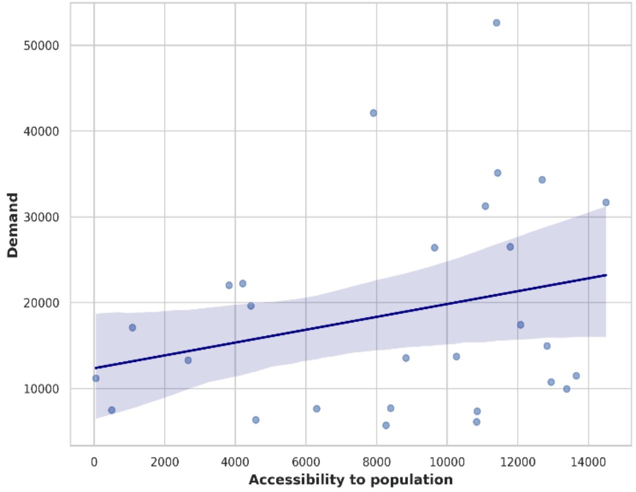


Fig. 7: Communication of the level of access is based on the population and the demand volume.

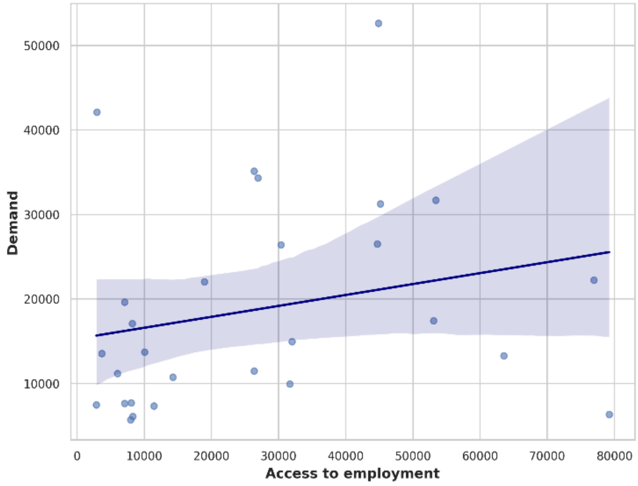


Fig. 8: The relationship between the level of access and the demand volume (accessibility to employment).

4. Methodology

Machine learning, as a subset of artificial intelligence, aims to develop and understand algorithms that learn from data to improve performance across a range of tasks. These algorithms, referred to as "learners," leverage data to enhance their capabilities. There are various approaches within machine learning that cater to different types of data and tasks.

Supervised learning, for instance, utilizes training data with known labels or outcomes. For example, in a transportation context, this could involve data with labeled features such as time of day, weather conditions, and historical demand, along with corresponding labels like the number of trips taken. On the other hand, unsupervised learning tackles unlabeled input data, seeking to identify patterns or structures within the data without explicit guidance. In the transportation domain, this could involve clustering similar patterns of demand or identifying anomalies in travel behavior.

In addition to supervised and unsupervised learning, there is also semi-supervised learning, which operates with a combination of labeled and unlabeled data. This approach is particularly useful when acquiring labeled data is expensive or time-consuming. In the realm of transportation, semi-supervised learning could involve leveraging a small dataset of labeled demand patterns alongside a larger dataset of unlabeled travel data to improve predictive models or identify trends in passenger behavior.

Each of these machine learning paradigms offers unique advantages and applications within the transportation domain, contributing to the development of more efficient, adaptive, and data-driven solutions for urban mobility challenges. Among existing neural network models, the recurrent neural network with a gate mechanism stands out as the most widely used model for sequential data. This model effectively processes sequence data such as time series and speech signal data. By employing mechanisms for incorporating long-term information into the input, the recurrent neural network with a gate mechanism effectively addresses the issue of gradient vanishing that may arise in traditional recurrent neural network models.

Compared to another commonly used recurrent neural network model, the short-term and long-term memory neural network, the recurrent neural network with a gate mechanism boasts fewer parameters, resulting in significantly faster computation while offering comparable performance in prediction [12, 13, 14, 15].

We aim to utilize recurrent neural networks with gating mechanisms to capture time-series patterns while incorporating spatial features. Our approach will then be compared with a multilayer perceptron (MLP) to evaluate its effectiveness.

According to Figure 9, the Gated Recurrent Unit (GRU) employs two key components—an update gate and a reset gate—to manage the flow of information. The reset gate determines how much of the past information should be forgotten, while the update gate controls how much of the previous state should be carried forward to influence the current state. This mechanism enables the GRU to efficiently capture temporal dependencies in sequential data. The structure of a GRU unit is illustrated in the figure below.

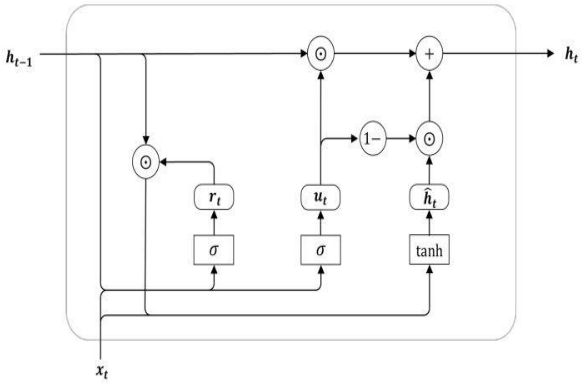


Fig. 9: Gate Recurrent Unit

In the recurrent neural network with a gate mechanism, each unit takes the previous hidden state as input and produces the new hidden state as output. The reset gates and update gates are computed by fully connected layers using an activation function. In this case, the sigmoid function serves as the activation function for the gate mechanism. Mathematically, for a given time step, the reset and update tuning gates are calculated by Equations 1 and 2, respectively [39].

$$r_t = \sigma(x_t W_{xr} + h_{t-1} W_{hr} + b_r) \quad (1)$$

$$u_t = \sigma(x_t W_{xu} + h_{t-1} W_{hu} + b_u) \quad (2)$$

The weight parameters are denoted by(W), and the bias parameters are denoted by(b). After calculating the reset gate and update gate using Equations 1 and 2, respectively, we integrate the reset gate with the hidden state of the previous time step and the input of the current time step to compute the hidden state for the current time step, as described by Equation 3:

$$\widehat{h_t} = tanh(x_t W_{xh} + (r_t W_{xt} + (r_t \odot h_{t-1}) W_{hh} + b_h \quad (3)$$

Subsequently, we utilize the update gate along with the current hidden state and the hidden state from the previous time step to compute the new hidden state. This new hidden state, calculated at the output of the recurrent neural network unit with a gate mechanism, is then passed on to the next time step. The expression for the current hidden state is represented by Equation 4 [39].

$$h_t = u_t \odot h_{t-1} + (1 - u_t) \odot \widehat{h_t} \quad (4)$$

A multi-layer perceptron (MLP) is a form of artificial neural network that comprises several layers of neurons. These neurons often employ nonlinear activation functions, enabling the network to discern intricate patterns within datasets. MLPs hold substantial importance in machine learning as they possess the capability to grasp nonlinear associations within data, rendering them potent models for classification, regression, and pattern recognition tasks. Throughout this tutorial, we will delve further into the fundamentals of MLPs and gain insight into their operational mechanisms [19].

Matrix factorization is a mathematical technique used to decompose a matrix into a product of two or more matrices, which can help in understanding and representing underlying patterns or structures within the data. In the context of recommendation systems, matrix factorization is often employed to analyze user-item interaction data, such as user ratings on items. By decomposing the user-item interaction matrix into two lower-dimensional matrices, one representing users and the other representing items, it becomes possible to approximate the original matrix and predict missing values. This allows for the generation of personalized recommendations for users based on their past interactions with items. Matrix factorization has applications beyond recommendation systems, including collaborative filtering, data compression, and dimensionality reduction.

The Figure 10 shows that the concept of matrix factorization for predicting traffic flow entails using mathematical techniques to decompose a traffic flow matrix into lower-dimensional matrices.

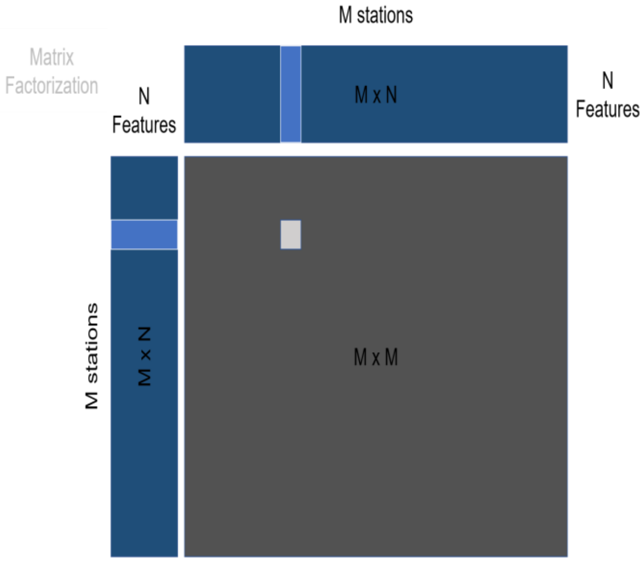


Fig. 10: The concept of matrix factorization is employed for predicting traffic flow

This decomposition allows for the identification of underlying patterns or structures within the traffic data. By representing the traffic flow matrix as a product of these lower-dimensional matrices, it becomes feasible to approximate and predict traffic patterns at different locations and times. This approach aids in comprehending the complex dynamics of traffic flow and facilitates more accurate traffic predictions, thereby contributing to better traffic management and planning.

Matrix factorization is utilized in this context because our goal is to predict traffic flow between stations. By structuring the problem as a matrix factorization task, we can decompose the complex interactions between origin and destination stations into lower-dimensional representations. This allows us to capture latent features that influence traffic flow. Additionally, assigning unique feature representations to both origin and destination stations helps model station-specific characteristics and enhances the predictive capability of our approach.

To enhance the accuracy of predicting traffic flow, we employ a strategy of integrating two branches of MLP. Each branch specializes in capturing distinct sets of features: one focusing on origin features while the other on destination features. Subsequently, we merge these two branches, effectively concatenating their outputs. This integrated approach enables us to provide more precise predictions of traffic flow. Figure 11 illustrates the implemented Multi-Layer Perceptron (MLP).

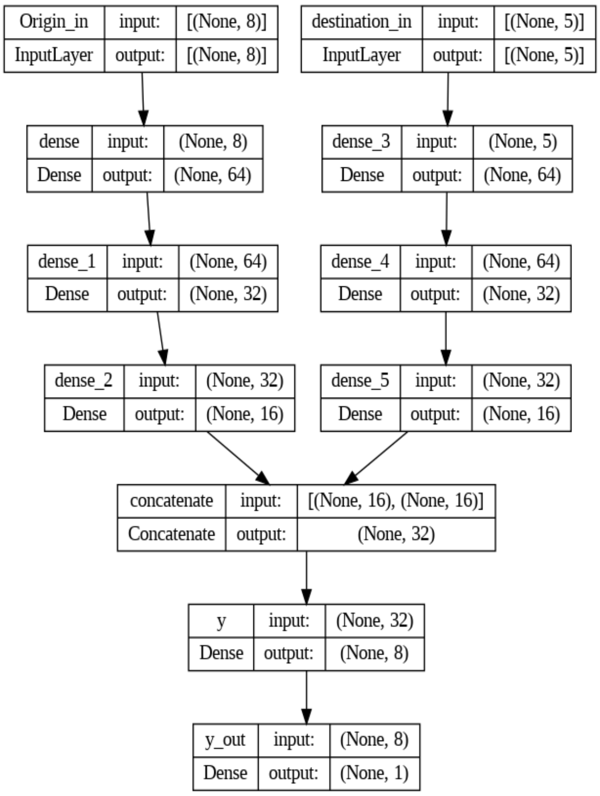


Fig. 11: The architecture implemented utilizes Multilayer Perceptrons (MLPs)

Locational access, which quantifies cumulative opportunities, assesses the reachable locations from a specific point. A map depicting locational access offers valuable insights into the dispersion of access to various locations, identifies areas requiring services, and underscores potential opportunities for development and investment. The number of opportunities, such as residents or employment, within the isochrone of the query point corresponds to the value of locational access. The query point, also referred to as the point of interest, could be a station gate or platform, for instance. An isochrone in a transportation network represents the widest coverage provided by one or more modes from a transit node within a predetermined time window. One measure of opportunity access is the proportion of residents in an isochrone's vicinity who can reach the platform of a train station within a 15-minute walk. Equation 5 presents a representation of location access analysis [17, 18].

$$A_{i,T} = \sum_{j=1}^{J} Oj.f(Cij) \qquad (5)$$

Cumulative opportunity $A_{i,T}$ between station i and every other reachable location within time T is determined by:

- O_j: Number of opportunities (population or jobs) at location j
- C_{ij}: Walking time from station i to location j
- $f(C_{ij})$: A function equal to 1 if $C_{ij} \leq T$ and 0 otherwise.

In order to gauge the performance of the models in this investigation, several evaluation metrics were utilized. These included Root Mean Squared Error (RMS) and Mean Squared Error (MSE). By calculating these metrics, a comprehensive assessment of the accuracy and predictive ability of the model was obtained.

Both continuous and discontinuous spatiotemporal networks can compute isochrones. Isochrones pertaining to public transportation systems exhibit discreteness in both temporal and spatial dimensions. Conversely, walking isochrones manifest in a continuous spatial context, wherein all points remain accessible at any given moment. This study exclusively focuses on the walking isochrones originating from each station of the bike-sharing system.

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

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

In this formula:

- N: represents the total number of samples.
- $\hat{y}_i$: denotes the predicted value for the i-th sample.
- y_i: represents the actual value for the i-th sample.

The datasets underwent a specific procedure to prepare them for demand modeling. Initially, considering a notable portion of trips logged by trip ID and user ID, the count of pickup bikes was established for each hour on every date at each geographic coordinate station, utilizing the Python packages "Pandas and Numpy" for grouping.

Following this, the trajectory trip data was transformed from a detailed level (individual trip records) to a summarized level (travel demand at each station). During this stage, we calculated novel attributes such as the total number of in-trips and out-trips per hour, as well as access at each station. Subsequently, we integrated the historical weather dataset, with a one-hour time step, into the main dataset. Furthermore, we employed the Min-Max scaling technique on the datasets.

Lastly, utilizing this amalgamated dataset, we constructed a spatial-temporal model of bike-sharing demand in Chicago. For data partitioning, 30 percent of the complete dataset was set aside as test data, maintaining uniformity across comparisons between base models and the main model. The remaining 70 percent of the data was allocated for training and evaluation purposes. Figure 12 illustrates the structure of the database.

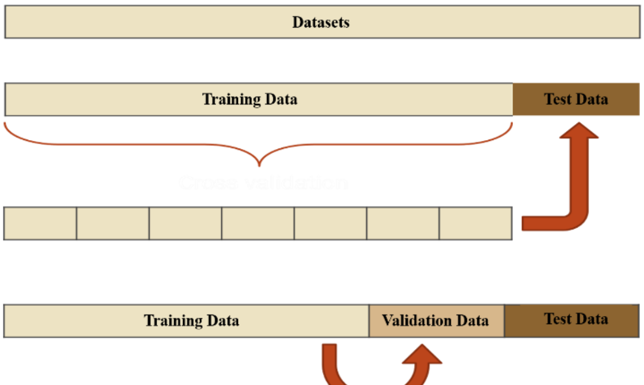


Fig. 12: Preprocessing data

5. Result

In this study, the recurrent gate unit neural network was employed as a comparative benchmark against the multilayer perceptron model. This neural network operates similarly to the long short-term memory neural network.

Table 1 shows hyperparameter settings for GRU and MLP models. Among the pivotal parameters of the network, key considerations include a batch size of 64, a learning rate of 0.001, 100 iterations for learning, and a sequence interval length of 24. Additionally, factors such as weather conditions, station accessibility, and the utilization of the Adam optimizer, LeakyRelu activation function, and mean square error loss function are integral to the model.

The network architecture comprises six layers with 32 and 64 hidden cells. Data is structured as a three-dimensional matrix (batch size, sequences, features), where each feature represents demand at individual time steps. The network receives data from the past 24 time steps to predict the current sequence, maintaining consistency in sequence and category size.

Table 1: Hyperparameter settings for GRU and MLP models.

Hyperparameter	GRU	MLP
Batch Size	64	64
Learning Rate	0.01	0.001
Epochs	100	100

As depicted in Figure 13, the Mean Squared Error (MSE) plots for both the training and validation datasets exhibit strikingly similar patterns, with reductions to nearly identical levels. This suggests the absence of overfitting or underfitting phenomena in the model. Upon implementing the proposed framework and reverting to the original scale, the mean square error was determined to be 0.88, with a corresponding root mean square error of 0.92.

Figures 14 to 17 display the actual and predicted demand across various time steps, demonstrating the model's effectiveness in forecasting the true demand levels accurately.

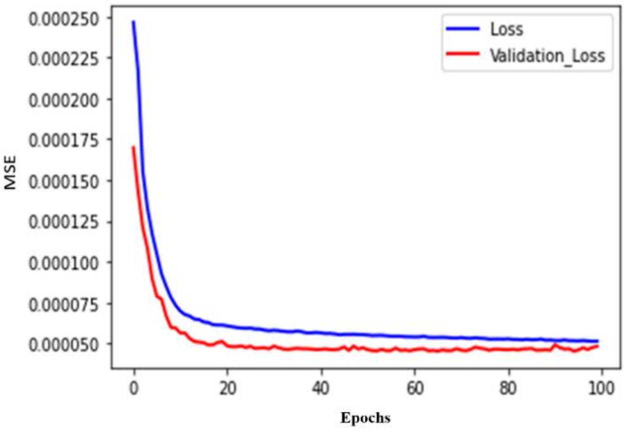


Fig. 13: Training and Validation Loss Curves

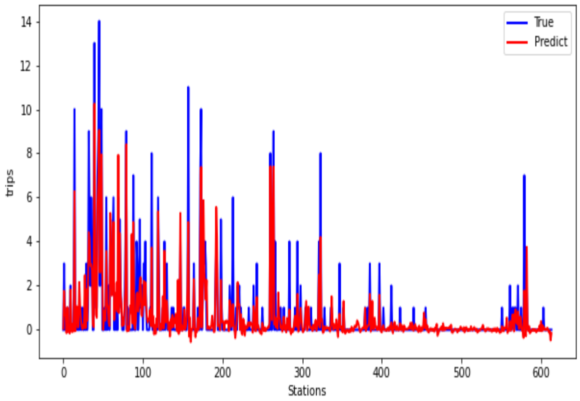


Fig. 14: Actual and predicted values by the GRU neural network

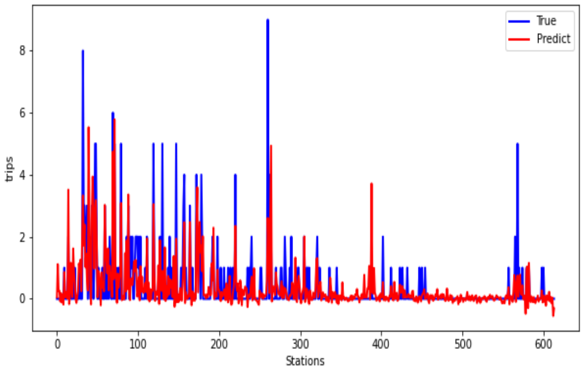


Fig. 15: Actual and predicted values by the GRU neural network

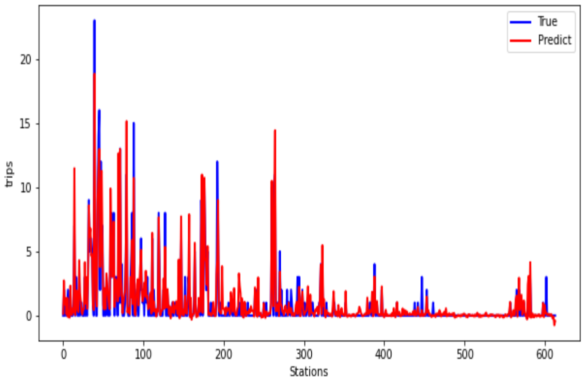


Fig. 16: Actual and predicted values by the GRU neural network

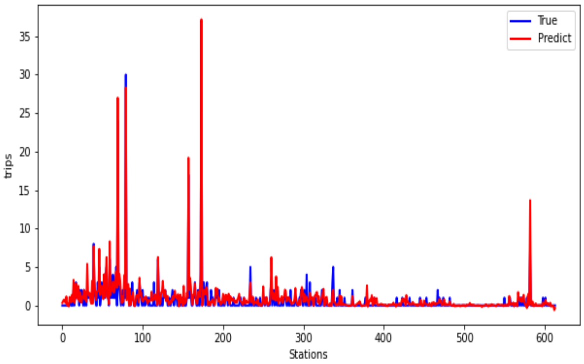


Fig. 17: Actual and predicted values by the GRU neural network

Multilayer perceptrons, abbreviated as MLP, represent a classical type of neural network architecture characterized by one or more layers of interconnected neurons. Data is fed into the input layer, which may comprise multiple hidden layers, and predictions are generated at the output layer, often referred to as the visible layer. MLPs find applicability across prediction and classification tasks, where each input receives a class or label assignment, as well as in regression prediction problems, where real-valued quantities are forecasted based on a set of inputs. Typically, data is presented in the form of a two-dimensional matrix. MLPs exhibit high flexibility, enabling them to discern patterns from input-output relationships.

Among the specified parameters of the MLP neural network utilized in this study, notable considerations include a batch size of 128, a learning rate of 0.0001, and 100 iterations for learning. Additionally, the model incorporates the Adam optimizer, the Relu activation function, and employs the mean square error method for loss computation.

As depicted in Fig. ***13***, the Mean Squared Error (MSE) plots for both the training and validation datasets exhibit strikingly similar patterns, with reductions to nearly identical levels. The input data is structured as a two-dimensional matrix, with features corresponding to variables such as rainfall, humidity, temperature, and access level, while the input label is represented as a one-dimensional matrix denoting the demand per hour post-implementation of the proposed framework. The mean square error (MSE) computed for this model was 0.9, with a corresponding root mean square error (RMSE) of 0.95. Figures 19 to 22 display the actual and predicted demand across various time steps, demonstrating the model's effectiveness in forecasting the true demand levels accurately.

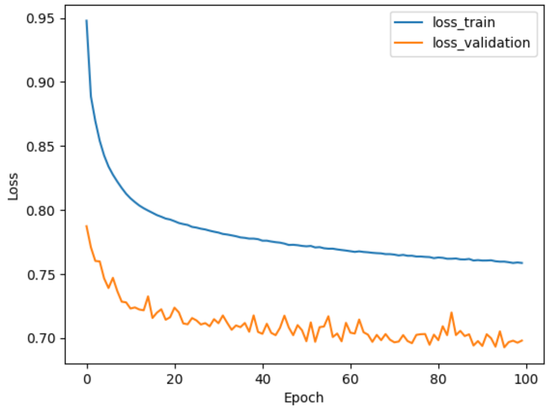


Fig. 18: Training and Validation Loss Curves

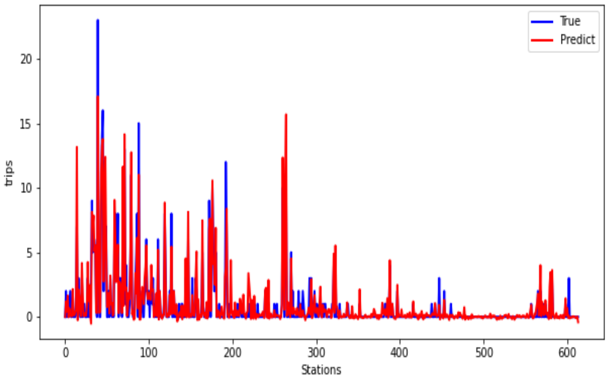


Fig. 19: Actual and predicted values by the MLP neural network

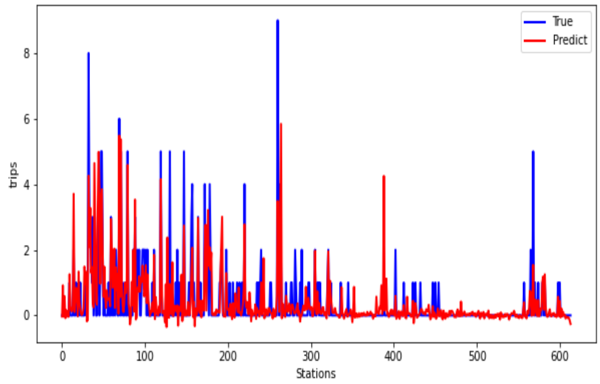


Fig. 20: Actual and predicted values by the MLP neural network

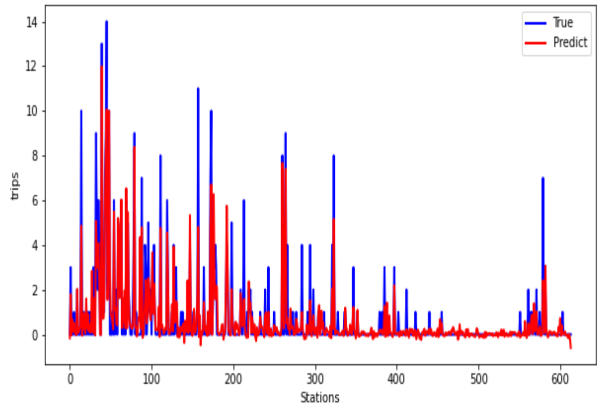


Fig. 21: Actual and predicted values by the MLP neural network

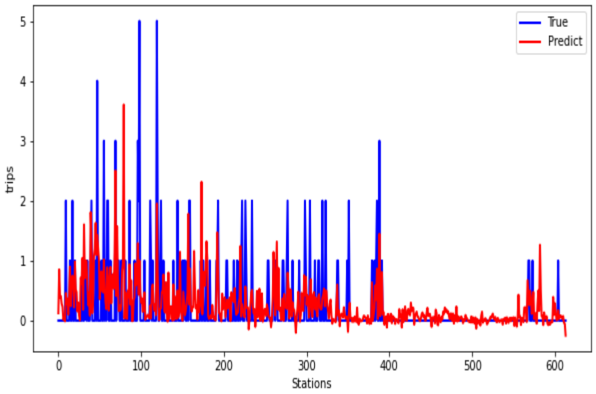


Fig. 22: Actual and predicted values by the MLP neural network

Table 2 indicates that the GRU model outperforms the parallel multilayer perceptron, demonstrating its capability to extract demand patterns based on weather characteristics, access levels of each station, and historical travel data. Another critical concern for transportation planners is accurately forecasting system demand on an hourly basis. One key network for predicting time series data is the short-term and long-term memory neural network (LSTM). This study discusses the implementation of LSTM with a recurrent layer mechanism, which examines sequences from both the beginning and the end.

Table 2: Summary Results

Models	MSE	RMSE
Parallel-MLP	0.9	0.95
GRU	0.88	0.92

Several essential parameters are set for this network, including a batch size of 64, a learning rate of 0.001, 100 epochs for training, and a sequence length of 24. The Adam optimizer, LeakyReLU activation function, and mean square error loss function are employed. The network architecture comprises four layers with 64 and 32 hidden cells, with an output layer of 1. Input data is structured as a three-dimensional matrix (hourly demand, sequence, number of batches), enabling the prediction of system-wide demand every hour.

As illustrated in Figure 23, the Mean Squared Error (MSE) plots for both the training and validation datasets display remarkably similar patterns, converging to nearly identical levels. The input data is organized into a two-dimensional matrix, with features corresponding to variables like rainfall, humidity, temperature, and access level. Meanwhile, the input label is represented as a one-dimensional matrix, indicating the hourly demand for all stations following the implementation of the proposed framework. The computed mean square error (MSE) for this model was 0.8, accompanied by a corresponding root mean square error (RMSE) of 0.89.

In Figure 24, the horizontal axis illustrates the time steps of the clock, while the vertical axis displays the total system demand. A comparative analysis between the predicted and actual values is presented, showcasing the performance of the LSTM model. The model demonstrates its effectiveness by closely aligning with the actual demand values, indicating its capability to accurately forecast system demand over time.

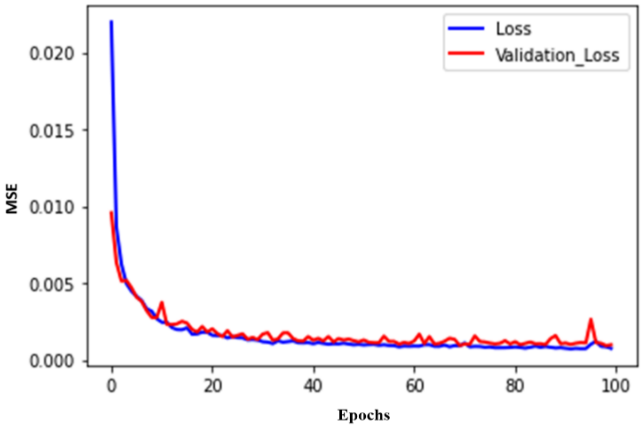


Fig. 23: Training and Validation Loss Curves

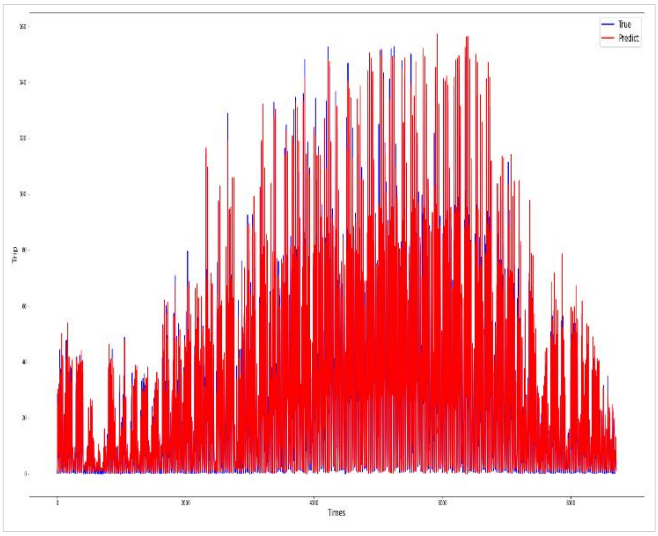


Fig. 24: Actual and predicted values of system demand are generated for each time step (hour) by the LSTM network.

6. Conclusion

Machine learning models offer a distinct advantage in simulating intricate interactions without relying on linear assumptions, setting them apart from classical models. Multi-Layer Perceptron (MLP) and Gate Recurrent Unit (GRU) exemplify this by providing a comprehensive analysis of outcomes, capturing both spatial and temporal influences between stations.

Our research introduces a framework incorporating weather conditions, access data, and bike-sharing transactions to facilitate short-term traffic flow predictions. By modeling the topology of the traffic network through a dynamic traffic flow probability graph, we depict the spatial features of a given area. Utilizing Parallel MLP and GRU, we extract spatial and temporal features, training deep-learning models for accurate short-term predictions of bike-sharing traffic flow between stations.

Validation of our approach using simulated data from Chicago City's bike-sharing system reveals the gate graph convolution model's superior performance, with lower RMSE and MSE. Our experiments underscore the framework's ability to forecast short-term traffic flow, surpassing popular deep learning methods.

Exploration of different models considering temporal or spatial approaches concludes that spatio-temporal prediction proves most effective. Additionally, our study uncovers significant patterns in how urban form influences bike-share traffic behavior across various spatial and temporal contexts. Additionally, the weather data was sourced from OpenWeather [21], while transaction data was obtained from the Divvy bike-sharing system [22].

Future research could benefit from incorporating a wider range of transaction data to enhance the accuracy of travel demand predictions. Additionally, exploring different graph neural network architectures, including various gating mechanisms and attention-based approaches, may improve model performance. Furthermore, evaluating the model's effectiveness across different periods, such as pre-COVID-19 and post-COVID-19, could provide valuable insights into how travel behavior has evolved in response to the pandemic.

References

[1] Kon, F., et al., Abstracting mobility flows from bike-sharing systems. Public Transport, 2021.

[2] Lin, L., Z. He, and S. Peeta, Predicting station-level hourly demand in a large-scale bike-sharing network: A graph convolutional neural network approach .Transportation Research Part C: Emerging Technologies, 2018. 97: p. 258-276.

[3] Van Acker, V., B. Van Wee, and F. Witlox, When transport geography meets social psychology: toward a conceptual model of travel behaviour. Transport Reviews, 2010. 30(2): p .219-240.

[4] Handy, S.L., Y. Xing, and T.J. Buehler, Factors associated with bicycle ownership and use: a study of six small US cities. Transportation, 2010. 37(6): p. 967-985.

[5] Fishman, E., Bikeshare: A review of recent literature. Transport Reviews, 2016 .36)1 :(p. 92-113.

[6] Bushell, M.A., et al., Costs for pedestrian and bicyclist infrastructure improvements. University of North Carolina Highway Safety Research Center, University of North Carolina, Chapel Hill, 2013. 45.

[7] Bike Sharing Demand. Available from: https://www.kaggle.com/c/bike-sharing-demand.

[8] Giot, R., et al., Predicting bikeshare system usage up to one day ahead.

[9] Seo, Y.-H., et al. Predicting demand for a bike-sharing system with station activity based on random forest. in Proceedings of the Institution of Civil Engineers-Municipal Engineer. 2021. Thomas Telford Ltd.

[10] Xu, Y., M. Paliwal, and X. Zhao, Real-time Forecasting of Dockless Scooter-Sharing Demand: A Context-Aware Spatio-Temporal Multi-Graph Convolutional Network Approach. arXiv preprint arXiv:2111.01355, 2021.

[11] Wang, J., et al. Traffic speed prediction and congestion source exploration: A deep learning method. in 2016 IEEE 16th international conference on data mining (ICDM). 2016. IEEE.

[12] Yang, Yuanxuan, Alison Heppenstall, Andy Turner, and Alexis Comber. "Using graph structural information about flows to enhance short-term demand prediction in bike-sharing systems." Computers, Environment and Urban Systems 83 (2020): 101521.

[13] Behroozi, A., Edrisi, A. Predicting travel demand of a bike sharing system using graph convolutional neural networks. *Public Transp* (2024). https://doi.org/10.1007/s12469-024-00371-w

[14] Yang, Z., et al. Mobility modeling and prediction in bike-sharing systems. in Proceedings of the 14th annual international conference on mobile systems, applications, and services. 2016.

[15] Tian, Y. and L. Pan. Predicting short-term traffic flow by long short-term memory recurrent neural network. in 2015 IEEE international conference on smart city/SocialCom/SustainCom (SmartCity). 2015. IEEE.

[16] Yipeng Liu, H.Z., Xinxin Feng, and Zhonghui Chen, Short-Term Traffic Flow Prediction with Conv-LSTM, in IEEE Access. 2017

[17] He, Z., C.-Y. Chow, and J.-D. Zhang, STANN: A spatio–temporal attentive neural network for traffic prediction. IEEE Access, 2018. 7: p. 4795-4806.

[18] Kowsrihawat, K., P. Vateekul, and P. Boonkwan. Predicting judicial decisions of criminal cases from thai supreme court using bi-directional GRU with attention mechanism. in 2018 5th Asian Conference on Defense Technology (ACDT). 2018. IEEE.

[19] Owen, A. and B. Murphy, Access Across America: Transit 2017 Data. 2018.

[20] Owen, A. and D.M. Levinson, Developing a comprehensive US transit accessibility database, in Seeing Cities Through Big Data. 2017, Springer. p. 279-290.

[21] openweathermap.org

[22] divvybikes.com