GCN-Transformer for short-term passenger flow prediction on holidays in urban rail transit systems

Shuxin Zhang, Jinlei Zhang, Lixing Yang, and Ziyou Gao

*State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China
*Correspondence: zhangjinlei@bjtu.edu.cn; lxyang@bjtu.edu.cn

Abstract

The short-term passenger flow prediction of the urban rail transit system is of great significance for traffic operation and management. The emerging deep learning-based models provide effective methods to improve prediction accuracy. However, most of the existing models mainly predict the passenger flow on general weekdays, while few studies focus on predicting the holiday passenger flow, which can provide more significant information for operators because congestions or accidents generally occur on holidays. To this end, we propose a deep learning-based model named GCN-Transformer comprising graph conventional neural network (GCN) and Transformer for short-term passenger flow prediction on holidays. The GCN is applied to extract the spatial features of passenger flows and the Transformer is applied to extract the temporal features of passenger flows. Moreover, in addition to the historical passenger flow data, social media data are also incorporated into the prediction model, which has been proven to have a potential correlation with the fluctuation of passenger flow. The GCN-Transformer is tested on two large-scale real-world datasets from Nanning, China during the New Year holiday and is compared with several conventional prediction models. Results demonstrate its better robustness and advantages among baseline methods, which provides overwhelming support for practical applications of short-term passenger flow prediction on holidays.

Keywords: Urban rail transit; short-term passenger flow prediction on holidays; GCN; Transformer; social media data; deep learning

1 Introduction

The urban rail transit (URT) system has experienced rapid development in recent decades. As an important component in intelligent URT systems, short-term passenger flow prediction has been extensively studied nowadays. It is a critically significant task because leveraging the results of short-term passenger flow prediction, passengers can better schedule their travel plans and operators can take corresponding measures to provide high-level services. However, it is also a critically challenging problem especially on holidays because passenger flows contain complicated temporal and spatial characteristics and are easily affected under specific scenarios. For example, during holidays, passenger flows often vary significantly and the regularities of passenger flow on holidays are extremely different from workdays, as shown in Figure1. Therefore, how to conduct an accurate prediction of passenger flow, especially during holidays remains a necessary and challenging task.
To solve this problem, some scholars have studied the short-term passenger flow prediction in holidays. For instance, Chen and Liang (2015) proposed a machine learning-based method which hybridizes support vector regression model with an adaptive genetic algorithm to forecast holiday daily tourist flow. Additionally, considering that the support vector machine can deal with complex nonlinear characteristics, Liu and Yao (2017) proposed a modified support vector machine to forecast passenger flow on holidays. Moreover, Xie and Sun (2020) analyzed the spatiotemporal characteristics of holiday passenger flow and then established a deep learning model based on modified backpropagation neural network (BPNN) to forecast the holiday passenger flow. These predictive models fill the gap of holiday passenger flow prediction. However, there are also several shortcomings as follows. First, most of these models ignore the fact that there is a limited sample size of holiday passenger flow. A small sample size of passenger flow data in holidays generally increases the difficulty of accurately predicting the short-term passenger flows. Second, most models just utilize conventional data such as present and past passenger flow data and it is not adequate for accurate holiday passenger flow prediction. Only relying on passenger flow data cannot fully capture the temporal and spatial characteristics of passenger flows on holidays. Social media messages under specific scenarios have been proved that can provide reliable contacts for forecasting traffic flow (Roy et al., 2021). Hence, the motivation of this paper is to study how to combine conventional passenger flow data with social media data to accurately predict holiday passenger flow leveraging limited holiday passenger flow data.

In this paper, we propose a deep-learning model based on graph convolutional neural network (GCN) and Transformer architecture, named GCN-Transformer, for accurate network-scale short-term passenger flow prediction on holidays in the URT system. In addition to the conventional passenger flow data on holidays, social media data (microblogs volumes data), are also integrated into the GCN-Transformer to fully capture the fluctuant trends of passenger flows on holiday. We compare the model with other eight prediction models such as ARIMA, BPNN, Convolutional Neural Network (CNN),
and Long-Short Term Memory (LSTM). Experimental results on two real-world datasets from Nanning, China show the superiority and the great robustness of the GCN-Transformer model. Specifically, the main contributions of this paper are as follows:

1. The passenger flow data and social media data (microblogs volumes data) related to holidays are combined to improve the accuracy of short-term passenger flow predictions on holidays.
2. We propose a deep learning architecture called GCN-Transformer based on GCN and Transformer for short-term passenger flow prediction in URT systems. The GCN layer is applied to fully capture the spatial and topological correlations of stations in the URT network. The Transformer architecture is introduced because of its great ability to capture both short- and long-term temporal features.
3. We utilize passenger flow data on holidays for two consecutive years to fully capture holiday characteristics, which solves the problem of insufficient prediction accuracy caused by the limited sample size of holiday passenger flow.
4. The advantages of the GCN-Transformer model are demonstrated by two real-world datasets of Nanning, China around the New Year holiday from 2019 to 2021. Results show its favorable prediction performance in URT passenger flow prediction on holidays and the considerable robustness in different scenarios.

The remainder of this article is organized as follows: Section 2 reviews the literature about passenger flow prediction. Section 3 provides the problem definition. Section 4 describes the details of the proposed GCN-Transformer model. In section 5, we evaluate the prediction performance of our model based on two real-world datasets. Finally, we make a conclusion of our paper in Section 6.

2 Literature review

In this section, short-term passenger flow prediction models, holiday passenger flow forecasting, and the application of social media data are summarized.

2.1 Short-term Passenger flow prediction

In recent years, lots of short-term passenger flow prediction methods have been put forward (Liu, 2017; Zhang and Chen, 2019; Ma, 2021). Generally, the prediction methods can be divided into two categories, one is the regression model based on mathematical statistics, and the other is the machine learning-based model.

The general mathematical statistic methods include the Autoregressive Integrated Average model (ARIMA) and grey prediction model (Liu and Yang, 2016; Li and Xiang, 2020), etc. Zeng and Xu (2008) proposed a hybrid model that combines both ARIMA and multilayer artificial neural networks for short-term traffic flow prediction. Kumar and Vanajakshi (2015) proposed a prediction scheme based on the seasonal ARIMA model for short-term prediction of traffic flow using only limited input data, which could overcome the problem of data availability. Ni et al. (2017) integrated linear regression and seasonal ARIMA to predict passenger flow under event occurrences. Considering that the nature of
passenger flow forecasting is a time series problem, it is believed that the key is to capture the spatiotemporal characteristics of passenger flow series, which is the weakness of statistical regression models like ARIMA. On the contrary, methods based on machine learning can better capture spatiotemporal features and thus outperform these classic models.

Machine learning-based models such as support vector regression (SVR) and random forest learning, which can fully capture nonlinear features and spatiotemporal characteristics of passenger flow data, are increasingly applied in traffic prediction (Hong and Dong, 2011; Leshem, 2007). Castro-Neto and Jeong (2009) presented an online SVR model for the prediction of short-term traffic flow under both typical and atypical conditions. Hu and Yan (2016) proposed a hybrid prediction method based on particle swarm optimization and SVR for short-term traffic flow forecasting. However, most of these models only consider temporal characteristics and might not fully consider spatial correlations in the model formulation. Moreover, these models are generally verified on a single URT station and are inapplicable for the stations of an entire network (Zhang and Chen, 2020).

As a branch of machine learning, deep learning-based model, such as BPNN (Zheng and Lee, 2006), CNN (Zhang and Yu, 2019), LSTM (Ma et al., 2015), GCN (Yu and Lee, 2020), have also received lots of attention both industrial and academic in recent year. For instance, Yang and Chen (2019) proposed an improved model enhancing long-term features based on LSTM to capture long temporal dependence for URT passenger flow prediction. Ren et al. (2019) utilized a residual network (ResNet) based on CNN to model the spatiotemporal dependency in order to improve the traffic flow prediction accuracy. Wu and Tan (2016) presented a hybrid deep learning-based architecture combining CNN and LSTM to predict short-term traffic flow. To capture topological information, Xu et al. (2019) presented a spatiotemporal multi-graph convolution network (ST-MGCN) to conduct traffic demand prediction. Zhao and Song (2020) proposed a temporal graph convolutional network (T-GCN) model combining the GCN and the gated recurrent unit (GRU) for traffic flow prediction. Zhang and Chen (2021) developed a novel OD flow forecasting method that considers the unique characteristics of the URT system, which mainly consists of a channel-wise attention mechanism and split CNN.

These short-term passenger flow prediction models have favorable performance on weekdays. However, due to the significant irregularity and fluctuation of passenger flows on holidays, these models might not fully capture spatial-temporal characteristics and holiday features, resulting in poor prediction performance on holidays. Hence, how to improve the accuracy of passenger flow prediction on holidays remains to be explored.

2.2 Holiday passenger flow forecasting

To forecast the holiday passenger flow accurately, many scholars have carried out in-depth research. For example, Zeng and Sheng (2019) proposed a learning framework based on weighted knowledge transfer for daily peak load forecasting during holidays. Luo and Li (2019) utilized a hybrid prediction model combining discrete Fourier transformer (DFT)
with support vector regression (SVR) to extract common trends in the traffic flow for accurate holiday traffic flow prediction. Zhou et al. (2020) analyzes the characteristics of passenger flow during holidays and constructed a forecast model based on support vector machine to realize the accurate prediction. To find a more reliable model under various conditions such as holidays, Zhang and Yao (2021) presented a hybrid deep spatiotemporal model combining convolutional neural network (CNN), gated recurrent unit (GRU), and convolutional long short-term memory (ConvLSTM) models. Wen and Zhao (2022) proposed a decomposition-based forecasting method with transfer learning, which considered the linear and nonlinear time series, has been proved that it is beneficial to improve the accuracy of passenger flow prediction in holidays.

Most of the forementioned models focus on the regularity of passenger flow during holidays, but the sample size of historical passenger flow during holidays is too small to capture the clear pattern of passenger flow and train the models, resulting in low prediction accuracy. Hence, combining with other data that correlated with passenger flow may be a feasible way to improve the accuracy of prediction.

### 2.3 Social media application

In recent years, with the rapid development of social media, an increasing number of researchers have attempted to integrate social media into the traffic prediction field. Considering that social media (Chaniotakis and Antoniou, 2015) can reflect users’ intentions, it is seen as an effective data source to apply in passenger flow forecasting, especially regarding special events. For instance, Ni et al. (2014) proposed a short-term traffic flow prediction model, incorporated with tweet features to forecast incoming traffic flow before sport game events. A few years later (2017), they found that there existed a moderate positive correlation between passenger flow and the rates of social media posts. And they presented a parametric and convex optimization-based approach to predict subway passenger flow under event occurrences. Essien et al. (2020) proposed a deep learning prediction model based on Bi-directional LSTM and stacked autoencoder (SAE) that combines information extracted from tweet messages with traffic and weather information to improve predictive accuracy. Roy and Hasan (2021) utilized traffic sensor and Twitter data to predict traffic demand during hurricane evacuation. Xue and Liu (2022) firstly studied the multivariate disturbance that affect passenger flow of a station under event occurrences, and then they presented a three-stage deep learning model to model the disturbance of inbound flow from nearby stations and social media post trends.

The above literature interpreted that there is a great potentiality to apply social media to explore the comprehensive features for passenger flow under event occurrences so that improves prediction accuracy. However, there is little research exploring how to apply social media to holiday passenger flow prediction. In this paper, we present an approach combining historical passenger flow data and microblog data to predict passenger flow on holidays.

### 3 Problem definition
In this section, the methodological framework is formulated. We first define several key parameters and then put forward the learning problem of the prediction of subway passenger flow during holidays.

The purpose of this study is to use historical AFC data to predict the passenger flow of URT system at the next time interval. The data extracted from the ID card data can be counted and integrated at different time intervals (e.g., 10 minutes, 30 minutes). The time interval used in this study is 10 minutes.

**Definition 1 (passenger flow matrix):** A subway ID card transaction includes the following information: passenger’s card number, the arrival time of passenger, the arrival station of passenger, the departure time of the passenger, the departure station of the passenger.

Given the passenger arrival information at a time period $t$ at each station, let $p_n(t)$ be the observed volume at station $n$ during the $t^{th}$ time interval. For prediction, we define a passenger flow matrix as follow:

$$P_{t-1} = \begin{pmatrix} p_1(t-1) & p_1(t-2) & \cdots & p_1(t-q) \\ p_2(t-1) & p_2(t-2) & \cdots & p_2(t-q) \\ \vdots & \vdots & \ddots & \vdots \\ p_n(t-1) & p_n(t-2) & \cdots & p_n(t-q) \end{pmatrix}$$

(1)

Where $P_{t-1}$ denotes the observed inbound flow of the entire rail transit network during $t-1$ to $t$, $n$ denotes the number of stations in the whole systems, and $q$ denotes the maximum time steps during $t-1$ to $t$. In this paper, we chose 12 as the maximum time steps to predict passenger flow at the next time step $t$, $Y_t$.

**Definition 2 (traffic network):** To express the spatial topological relationship of each station in the whole subway network, a Graph $G = (S, E, A)$ was proposed, where $S = \{s_1, s_2, \cdots, s_n\}$ is the set of stations, $n$ is the number of stations in the network, and $E$ are the edges between the adjacent stations. We construct the graph by connecting two adjacent sites on the network so that spatial information can be represented through the graph. The formulation of the adjacent matrix definition is given as follows.

$$A_{|i|j|} = \begin{cases} 1, & \text{station } i \text{ and station } j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases}$$

(2)

Thus, the prediction problem passenger flow at the whole subway network during holidays can be formulated as follows:

**Problem:** At time interval $t-1$, given all AFC transactions, network graph $G$, and microblog data related to specific holiday, the historical passenger flow matrix $P_{t-1}$, social media matrix $V_t^{-1}$, are extracted, and used to predict the network passenger flow at next time step $t$, $Y_t$. Thus, the problem can be defined as follow:

$$Y_t = f(P_{t-1}, G, V_t^{-1})$$

(3)

Where $f$ indicates the model to be learned during the training process.

## 4 Methodology

In this section, we describe our proposed GCN-Transformer model in detail. We first depict the framework of GCN-Transformer. Then, one of its components, namely the GCN, is described. Finally, another component, namely the Transformer, is displayed.
4.1 GCN-Transformer framework

The GCN-Transformer framework is shown in Figure 2, which mainly consists of two parts. Specifically, in Part 1, the inbound flow and adjacency matrix are dealt with by GCN layer to capture the spatial and topological information. The outputs of GCN are input into Transformer layers, mainly consisting of multi-head attention structure, which is used to extract the continuous passenger flow trend and temporal correlation. Following the Transformer layers, the output is then input into a $1 \times 1$ convolution layer to aggregate the deep spatial-temporal features. In order to easily optimize the model and prevent overfitting, we insert a shortcut connection in this step, namely residual structure proposed by He (2015). The output after the residual operation is flattened and input into fully connected layers. The fully connected layers are used to capture the nonlinear relationship of the spatiotemporal features and reduce the output dimension to what we adopted. In Part 2, the social medial feature matrix is dealt with by fully connected network to capture nonlinear characteristics of holiday passenger flow. Ultimately, the output of Part 1 is combined with the Part 2’s to finally predict the passenger flow in the next time step, $Y_t$. 
4.2 GCN layer

Since the powerful ability to capture spatial correlation and topological information of the graph, Graph Convolutional Networks (GCN) (Kipf and Welling, 2016) has attracted more
and more attention in recent years. And the architecture of GCN is shown in Figure 3. However, most conventional predictive models regard the traffic network as a grid matrix, which makes the prediction less accurate (Li and Zhang, 2022). In this study, to better capture the internal topological dependence between adjacent stations of the URT network, we applied GCN structure that dealt with the inbound flow and network graph. Owing to the great performance of the GCN layer with the first-order filter, we used the GCN proposed by Kipf et al. [22] as shown in the following equation:

\[ P^{t+1} = f(H^t, A) = \sigma \left( \hat{D}^{-1/2} \hat{A} \hat{D}^{-1/2} P^t W^t + b^t \right), \hat{A} = A + I \]  

(4)

Where \( A \in R^{n \times n} \) denotes the adjacent matrix, \( I \in R^{n \times n} \) is the unit matrix, \( \hat{D} \in R^{n \times n} \) denotes the diagonal node degree matrix of \( \hat{A} \), \( P^l \) is the feature matrix of the \( l^{th} \) layer, which originally represented passenger flow matrix \( P^{t-1} \in R^{n \times q} \), \( W^l \) is the weight matrix of the \( l^{th} \) layer, \( b^l \) is the bias vector, and \( \sigma(\cdot) \) denotes the activation function.

Given the subway network graph \( G = (S, E, A) \) and the inbound flow matrix \( P^{t-1} \), the calculation processes of GCN can be summarized as the following parts. First, to fully capture the spatial topology of nodes, all neighbor nodes (stations) of target station are selected according to the topological information based on \( G = (S, E, A) \). As the feature matrix, inbound flow of the neighbor nodes and target node are aggregated based on the first-order filter operation. Eventually, the feature matrix containing spatial characteristics of each station can be obtained to forecast passenger flow more accurately.

As the inbound flow in holidays is significantly different from the normal weekdays passenger flow, and there is no obvious correlation and regularity with the daily passenger flow, it is not appropriate to use the common three patterns: the real-time pattern, the daily
pattern, and the weekly pattern to predict the holiday passenger flow. Hence, in this paper, we only utilize the passenger flow under the real-time pattern. Supposed the time interval is \( t_i \), the time step is \( t_s \), and the passenger flow of the current time \( t \) is to be predicted. The real-time pattern can be described as \( P_{real} = (P_{t-t_0}, P_{t-t_0+1}, \ldots, P_{t-1}) \), a segment of historical time series adjacent to the predicting period. The passenger flow in neighboring period will directly influence the passenger flow in the next time period.

The outputs of the real-time pattern \( (P_{t-t_0}, P_{t-t_0+1}, \ldots, P_{t-1}) \) is integrated and then input into the Transformer layer.

### 4.3 Transformer layer

Inspired by the success of Transformer architecture in Natural Language Processing (NLP) (Vaswani et al., 2017), we attempt to apply Transformer architecture to capture temporal correlation of passenger flow. The initial Transformer model mainly consists of encoder and decoder structures. In our work, we aim to use historical inbound flow \( P_{real} = (P_{t-t_0}, P_{t-t_0+1}, \ldots, P_{t-1}) \) to predict the future passenger flow \( Y_{t+1} \). Hence, we just utilize the encoder structure to capture the temporal characteristics of holiday passenger flow as much as possible, which mainly consists of multi-head attention mechanism. It has been proved that multi-head attention mechanism can enhance the ability to capture local information and be more suitable for processing time series (Shiyang et al., 2019).

The attention mechanism, also called scaled dot-product attention in the origin paper, can be regarded as a function. For self-attention mechanism, a task-related query vector \( Q \in \mathbb{R}^{N \times C_q} \), a key vector \( K \in \mathbb{R}^{N \times C_k} \) and a value vector \( V \in \mathbb{R}^{N \times C_v} \) are given by linear operation of the model input passenger flow matrix \( P^{t-1} \). In our model, \( Q, K, V \) are set as consistent, and the input to product them is not a sentence, but a time series data \( P^{t-1} \). And this latent subspace process can be formulated as follow:

\[
Q = W_q \cdot P^{t-1} \tag{5}
\]

\[
K = W_k \cdot P^{t-1} \tag{6}
\]

\[
V = W_v \cdot P^{t-1} \tag{7}
\]

Where \( W_q, W_k, W_v \) denote the weight matrices for \( Q, K, V \) respectively. During this process, we can aggregate the vectors of each time into a matrix, and calculate the output of all moments by Dot-product, which will improve the calculation speed and save memory space. After getting \( Q, K, V \), we can calculate the temporal dependencies \( Z \) by dot-product as:

\[
Attention(Q,K,V) = Z = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{8}
\]

Where the Softmax Layer makes non-linear changes to the input so that the output is between \([0, 1]\), and \( d_k \) denotes the columns of the \( Q, K \) matrix, which is used to avoid the softmax making the gradient too small to backpropagate.

As its name suggests, multi-head attention consists of multiple self-attention, which could capture multiple temporal features of passenger flow through multiple heads. The process of the multi-head attention mechanism is shown in Figure 4. Firstly, the passenger flow matrix \( P^{t-1} \) are input into \( h \) different self-attention respectively to calculate \( h \) temporal feature matrices. And then \( h \) temporal feature matrices are concatenated together.
and put into a Linear layer to obtain the final temporal feature matrix $Z$.

$$\text{Temporal Feature matrix } Z$$

$$\text{(Multi-Head Attention)}$$

$\text{Linear}$

$\text{Concat}$

$\text{Scaled Dot-Product Attention}$

$H$

$\text{Linear}$

$\text{Linear}$

$\text{Linear}$

Figure 4 The processing of Multi-Head Attention

5 Evaluation

In this section, we will verify the feasibility of the GCN-Transformer model with two real-world datasets. We firstly introduce the dataset used in our study, and then our model configurations and the evaluation metrics are described. We also chose several conventional models as benchmark models to compare the prediction performance. Finally, the experimental results are analyzed from several perspectives.

5.1 Dataset

Our study is based on Nanning metro network from 2019 to 2021 as shown in Figure 5. We focus on the Nanning URT passenger flow data, which contains 10 weeks of passengers’ data from 6:00 a.m. to 11:00 p.m. around New Year’s Day 2019 to 2021. We collect inbound flow data from the ID card of passengers entering and leaving the station as shown in Table 1. Through pre-processing, the passenger flow data is segmented and counted with appropriate time granularity (10 min is selected here) to obtain the data format we want as shown in Table 2. To better capture the holiday characteristics of New Year’s passenger flow and predict the passenger flow during New Year’s holiday more accurately, we adopt two years of New Year’s passenger flow for training and prediction. Take passenger flow data of five weeks before New Year’s Day for 2019 and 2020 as an example, including a total of ten weeks of passenger flow data. The passenger flow data of the first nine weeks which contains the New Year’s Day passenger flow of 2019 are used to train and validate
model, and the remaining data in the last week, that is, the New Year’s Day of 2020, are used to test the model. To maintain the accuracy of prediction, our paper only considers the same stations in different years. Two datasets were applied in our paper, as shown in Table 3. Moreover, we give a unique station number for all subway stations.

As for the social media data, we firstly crawl microblogs containing keywords “New Year holiday” and “Nanning” during a specific period, which is consistent with the time period of URT passenger flow data. Because the sample size of microblogs we have crawled is not enough, we count the number of microblogs according to the time interval of days, and then expand samples by the time interval of passenger flow.

Here, we briefly analyze the time series data. As shown in Figure 6, in Nanning, especially on weekdays, the URT passenger flow have distinct peak and off-peak periods, which is consistent with the general temporal distribution of passenger flow. In addition, holiday passenger flow is significantly more than the conventional passenger flow. And this is consistent with the distribution of New Year’s Day-related microblog volumes, indicating that there is a potential correlation between URT passenger flow and related microblog volumes.
Table 1 Original ID card data

| Card number | Tap-in time | Tap-in station | Tap-out time | Tap-out station     |
|-------------|-------------|----------------|--------------|---------------------|
| 3099****    | 2018/12/31  | Xiuxiang       | 2018/12/31   | Chaoyang Square     |
|             | 06:26:18    |                | 06:40:35     |                     |
| 3093****    | 2018/12/31  | Xinmin Road    | 2018/12/31   | Macun               |
|             | 19:45:17    |                | 20:00:30     |                     |
| 3150****    | 2018/12/31  | Baicanglin     | 2018/12/31   | Jinhu Square        |
|             | 07:26:33    |                | 07:43:55     |                     |

Table 2 Inbound passenger flow statistics of 2021

| Station index | 06:00-06:10 | 06:10-06:20 | 06:20-06:30 | ... | 22:50-23:00 |
|---------------|-------------|-------------|-------------|-----|-------------|
| 1             | 0           | 12          | 29          | ... | 0           |
| 2             | 0           | 7           | 16          | ... | 2           |
| 3             | 0           | 10          | 16          | ... | 1           |
| ...           | ...         | ...         | ...         | ... | ...         |
| 76            | 1           | 2           | 34          | ... | 0           |

Table 3 Data description

| Description                  | 2019, 2020 New Year’s Day                                      | 2020, 2021 New Year’s Day                                      |
|------------------------------|----------------------------------------------------------------|----------------------------------------------------------------|
| Date                         | December 3, 2018 to January 6, 2019                             | December 2, 2019 to January 5, 2020                             |
|                              | December 2, 2019 to January 5, 2020                             | November 30, 2020 to January 3, 2021                             |
| Time in a day                | 06:00 to 23:00                                                  | 06:00 to 23:00                                                  |
| Line number                  | 2                                                                | 3                                                                |
| Station number               | 41                                                               | 61                                                               |
| Time interval                | 10                                                               | 10                                                               |
| Data record                  | 55 million                                                       | 73 million                                                       |
| Week number                  | 5                                                                | 5                                                                |
| Day number | 35 | 35 |
|-------------|----|----|
| Time slice in one day | 102 | 102 |
| Total time slice | 3570 | 3570 |

### 5.2 Model configurations and evaluation metrics

In this paper, all models are implemented with PyTorch on a desktop computer with Intel® Core™ i9-10900X CPU, 64 GB memory, and an NVIDIA GeForce RTX3050 GPU.

**Hyperparameters**: The same parameters of our GCN-Transformer model are applied for both of the two datasets to evaluate its performance as follows. As mentioned above, our model consists of two parts. In part 1, the input dimension and output dimension of GCN layer are consistent, both are (batch size, number of stations, time steps). The encoder structure consists of three encoder units, and each encoder unit mainly consists of multi-head attention mechanisms, including eight self-attention mechanisms. And the parameters of $1 \times 1$ convolution layer are in_channels = out_channels = 1, kernel_size = 3, stride = 1 and padding = 1, respectively. The fully connected network consists of two hidden layers and one output layer, the unit numbers of the two hidden layers are [2048, 1024] respectively. And we adopt the RELU function as the activation function of the fully connected network. In part 2, the fully connected network consists of one hidden layer and one output layer, of which the unit number of hidden layers is 1024. The batch size is 32. The optimizer is Adam with the learning rate of 0.0001.

**Preprocessing**: Before training, all data were normalized to range (0,1) with Min-Max normalization scalers. The results evaluation was conducted after the predicted results were re-scaled to their original scale range.

**Evaluation Metrics**: For evaluating the performance of the implemented model, we have used Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Weighted Mean Absolute Percentage Error (WMAPE) as performance measures. When the predicted value is exactly the same as the real value, it is equal to 0. The closer the values of the three indicators are to 0, the better the prediction accuracy of the model is. We chose the model considering the best performance overall models with different parameters. And the formula of performance measures is as follows:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}
\]

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |Y_i - \hat{Y}_i|
\]

\[
WMAPE = \frac{1}{\sum_{l=1}^{n} Y_l} \sum_{l=1}^{n} \left| \frac{Y_l - \hat{Y}_l}{Y_l} \right| \times 100\%
\]

Where, $Y_i$ is the actual passenger flow at time $l$, $\hat{Y}_l$ is the predicted passenger flow at time $i$ and $n$ is the number of stations.

### 5.3 Benchmark models

In this section, we compare the prediction performance of our proposed GCN-Transformer model with several benchmark models on Nanning rail transit passenger flow data of 2020 and 2021 New Year holidays. The details of the benchmark models are described as follows:
Autoregressive Integrated Moving Average (ARIMA): The Autoregressive Integrated Moving Average, a parametric technique, is known as a traditional time-series prediction method and is widely used in passenger flow prediction. The three parameters in ARIMA, namely the lag order, the degree of difference, and the order of the moving average, are set as 2, 1, 0, respectively after fine-tuning.

Back Propagation Neural Network (BPNN): As the most basic neural network, BPNN has been proved to be applicable in predicting passenger flow. We applied a BPNN model consisting of two fully connected layers with 512 neural units in each layer. The optimizer is Adgrad with a learning rate of 0.0001. We use passenger flow data of full network stations to train this model. The inputs are the inbound passenger flow of the last 12 steps. The output is the inflow passenger flow for next time step.

Convolutional Neural Network (CNN): Convolutional Neural Network, a typical neural network, is well known for the great performance of capturing spatial correlation, which has already been applied to time-series data processing. We applied a general 2D CNN model with one CNN layer and two fully connected layers. The hidden layer has 8 filters. The kernel size is 3*3 with one padding and one stride. The optimizer is Adam with learning rate of 0.0001. The inputs and outputs are the same as BPNN.

Long Short-Term Memory Neural Network (LSTM): Specifically, we establish an LSTM model with two hidden layers and two fully connected layers. Each LSTM layer consists of 128 neurons. The optimizer is Adam with learning rate of 0.0001. The inputs and outputs of LSTM are the same as BPNN.

ST-GCN: It was proposed by Yan et al. (2018), which can automatically learn both the spatial and temporal features from passenger flow data. We adopt three branches of spatial temporal graph convolution units. The other parameters are similar to Yan et al.

GCN-CNN: A deep learning architecture composed of graph convolutional network (GCN) and 3D convolution neural network (CNN). The parameters are the same as Zhang et al. (2020).

ST-ResNet: Proposed by Zhang et al. (2017), it has been proved to outperform well-known methods in citywide crowd flows prediction. Here, we only adopt three branches of residual convolutional units and do not consider external factors. The other parameters are the same with his paper.

ConvLSTM: Proposed by Shi et al. (2015), ConvLSTM can fully capture the temporal and spatial characteristics of passenger flow. We set up a ConvLSTM model with three hidden layers and two fully connected layers. Other settings are the same as BPNN.

5.4 Experiment results

5.4.1 Network-wide prediction performance

Table 4 and Figure 7 show the performance of our GCN-Transformer model in comparison to other benchmark models for both passenger flow datasets of Nanning Metro in 2020 and 2021 New Year’s Day. As shown in Table 4, deep learning models significantly outperform the mathematical statistics-based model, ARIMA. It performed the worst no matter in which dataset, with RMSE of 181.128 and 167.459, and MAE of 125.011 and 98.095,
respectively. The reason is that ARIMA has a poor ability to capture complex nonlinear relationships of passenger flow, it can only capture limited temporal correlations.

Furthermore, we compare our model with other deep learning time series prediction models. Among these deep learning models, BPNN could only capture limited nonlinear characteristics so that predict the second worst, while CNN can capture more spatial correlations and LSTM can capture more temporal correlations so that they performed better than the conventional models. The complex deep learning architectures like ST-GCN, GCN-CNN, ST-ResNet, and ConvLSTM, which considered both spatial and temporal characteristics, have realized prediction accuracy improvements over the models mentioned above. However, since these models are proposed to predict weekday passenger flow, they do not apply to holiday passenger flow prediction. Thus, we proposed a model for holiday passenger flow that takes into account the impact of social media.

As studies have shown, there exists a moderate positive correlation between event passenger flows and social medial volumes [21], in our model, the social medial volumes were considered to improve holiday passenger flow forecast accuracy. In addition, as we have discussed, Transformer has shown its potential in time series processing and GCN has a strong ability to capture spatial correlation. Based on the advantages mentioned above, we then put forward the GCN-Transformer model combining the GCN with Transformer and considering the impact of social medial volumes to better explore the spatial-temporal characteristics of holiday passenger flow. As we expected, the proposed framework shows great performance and achieves the most accurate prediction among the benchmark models with the lowest RMSE of 26.860 and 29.882, MAE of 16.232 and 15.850, and WMAPE of 0.124 and 0.158 for the two datasets, respectively.

| models          | MetroNewYear2020 | MetroNewYear2021 |
|-----------------|------------------|------------------|
|                 | RMSE  | MAE     | WMAPE | RMSE  | MAE     | WMAPE |
| ARIMA           | 63.291 | 38.929  | 0.287  | 64.452 | 34.678  | 0.337  |
| BPNN            | 46.915 | 25.454  | 0.195  | 43.095 | 21.734  | 0.220  |
| CNN             | 30.030 | 17.562  | 0.134  | 33.587 | 17.410  | 0.175  |
| LSTM            | 29.928 | 17.786  | 0.136  | 32.016 | 17.152  | 0.172  |
| ST-GCN          | 29.593 | 17.378  | 0.133  | 32.040 | 16.990  | 0.170  |
| GCN-CNN         | 29.352 | 17.720  | 0.134  | 32.019 | 16.733  | 0.168  |
| ST-ResNet       | 28.364 | 17.272  | 0.132  | 31.884 | 17.129  | 0.170  |
| ConvLSTM        | 28.247 | 17.155  | 0.131  | 30.968 | 16.872  | 0.169  |
| GCN-Trans       | 26.606 | 16.306  | 0.124  | 28.912 | 15.826  | 0.158  |
During holidays, not all stations of the subway network have obvious characteristics of holiday passenger flow. Those stations adjoining the business district may have apparent characteristics of holiday passenger flow, while those stations that undertake daily commuting and connect urban and suburban areas do not. In this paper, we choose three stations with different passenger flow patterns to show the prediction performance of our GCN-Transformer. The first station is TingHong Road station, which is adjacent to the main business district, and many citizens choose to visit here during holidays. The second station is Guangxi University station, a typical commuter station with many passengers living nearby. The last station is Nanning Railway station, which is an apparent transfer hub that can achieve the various modes of transportation transference. The prediction results of these three stations during New Year’s holiday as shown in Figure 8, and below are the prediction results analysis.

The prediction result of TingHong Road station is shown in Figure 6 (a). It can be seen that the passenger flow during New Year’s Day has presented apparent holiday characteristics: the peak passenger flow is more obvious and much larger than usual. In this case, our proposed model can capture the holiday passenger flow characteristics well, making the forecast result consistent with the actual value.

The prediction result of Guangxi University station is shown in Figure 6 (b). It can be seen that the passenger flow in Guangxi University station has significant commuting characteristics, including obvious morning and evening peak characteristics. In addition,
during New Year’s holiday, the commuter traffic has dropped significantly, which is consistent with the law of commuting. Because of these regular passenger flow features, the forecast performance is preferable no matter during the peak or off-peak period, holidays, or weekdays, which indicates that our model has strong robustness, and performs well in different scenarios.

The prediction result of Nanning railway station is shown in Figure6 (c). As a transfer hub, the passenger flow has no obvious peak and off-peak period, there is no significant morning rush hour, instead, there are two tiny peak hours during afternoon and evening. Besides, during New Year’s holiday, the number of foreign tourists has increased, so has the number of subway passenger. It can be seen from the figure that the predicted value of our model is close to the actual value, meeting the demand of prediction precision, which manifests that the GCN-Transformer model can study the characteristics of various types of passenger flow.

In summary, the GCN-Transformer model can achieve accurate prediction not only for the whole subway network but also for different types of stations.
5.4.3 Prediction performance in different time intervals

To further study the prediction performance of our proposed model in different time intervals of a day, we calculate the average loss at each time interval from 6:00 to 23:00 for both Nanning Metro New Year 2020 and 2021. The prediction performance between our proposed model and other benchmark models at different time intervals is described as Figure 9. And below are several conclusions.

Firstly, we will discuss the correlation between the prediction performance in different time intervals of a day and the overall prediction effect. As figure 5 and figure 6 show, the overall prediction performance of all the models has the same pattern as the performance of different time intervals. The mathematical-statistical model ARIMA has the worst performance during the different time intervals of a day, just as its overall prediction performance is the worst of all the models. Besides, the fluctuations of ARIMA’s performance metrics over time intervals are the most drastic among all models, indicating that the statistics-based model is not available for large-scale passenger flow of urban railway network. Having the lowest evaluate metrics of overall prediction performance, our GCN-Transformer architecture also performed better than the benchmark most of the time intervals of a day. All of above illustrates that GCN-Transformer is suitable for network-wide prediction which has stable predictive performance.

Secondly, the predictive performance of the same model in different time intervals of a day will be analyzed. It can be shown from Figure 9 that the performance metrics of different time intervals during a day has an obvious trend of peak and off-peak, which is consistent with the morning and evening peak features of passenger flow, indicating that the performance metrics will go up when the passenger flow increases sharply, also meaning that the prediction accuracy will be reduced. However, as the characteristics of passenger flow during New Year holiday are quite different with those of normal weekdays, and the fluctuation of passenger flow during off-peak period is large, the result to the fluctuation of evaluation indicators during the off-peak period of a day. In addition, the statistics-based model ARIMA has the most significant peak and off-peak
characteristics of performance metrics, while our GCN-Transformer model has the slightest morning and evening peak characteristics of evaluating metrics during different time intervals, which illustrates that GCN-Transformer can accurately capture peak and off-peak features of passenger flow.

Then, we analyze the performance prediction of different models in the same time intervals of a day. No matter in peak period or off-peak period, the mathematical-statistical model ARIMA performs worst among all models, which has the highest evaluate metrics value. However, the deep learning models outperform the statistics-based model with much lower metrics value, and the performance of these deep learning models are roughly similar. Among these models, our GCN-Transformer performs best in most of the time intervals during a day, which means that our model is also suitable for passenger flow prediction in a single time interval.

Eventually, the prediction performance among all models in different datasets is discussed. It can be inferred that there exists similar prediction performance of all models for both Nanning Metro passenger flow datasets in 2020 New Year and 2021 New Year holiday, suggesting that the predictive performance of all models are generally consistent in different datasets.

To sum up, our proposed GCN-Transformer can obtain considerable prediction performance no matter on the daily or hourly granularity, indicating our model has a wide application prospect.
To further explore the influence of microblog volumes and model architecture on passenger flow prediction accuracy, the control variable method is used to design the experiment. We firstly change part of the structure of the model and keep the rest of structure unchanged to verify the reliability of the model structure, then we compare the prediction performance of the model using different data, such as using one year or two consecutive years of passenger flow, with and without using microblog volumes, finally we compare the RMSE, MAE, and WMAPE of the forecast results.

Table 5 and Figure 10 illustrates the experimental results on different conditions. According to the results, we summarize that GCN-Transformer with Microblog volumes outperforms other conditions with the highest accuracy and lowest average RMSE of 28.371, MAE of 16.041, and WMAPE of 0.141, respectively. When forecasting separately, the GCN-Transformer without ResNet performs the worst, indicating that ResNet can easily optimize the model and prevent overfitting to significantly reduce the RMSE, MAE, and WMAPE. While the prediction performance of our model without CNN layer is similar to the model without Transformer layer or GCN layer, illustrating that these structures can capture the temporal and spatial characteristics of passenger flow and

Figure 9 Comparison of the model performance in different time intervals in MetroBJ2016 and MetroBJ2018

5.4.4 Performance comparison of different conditions

To further explore the influence of microblog volumes and model architecture on passenger flow prediction accuracy, the control variable method is used to design the experiment. We firstly change part of the structure of the model and keep the rest of structure unchanged to verify the reliability of the model structure, then we compare the prediction performance of the model using different data, such as using one year or two consecutive years of passenger flow, with and without using microblog volumes, finally we compare the RMSE, MAE, and WMAPE of the forecast results.

Table 5 and Figure 10 illustrates the experimental results on different conditions. According to the results, we summarize that GCN-Transformer with Microblog volumes outperforms other conditions with the highest accuracy and lowest average RMSE of 28.371, MAE of 16.041, and WMAPE of 0.141, respectively. When forecasting separately, the GCN-Transformer without ResNet performs the worst, indicating that ResNet can easily optimize the model and prevent overfitting to significantly reduce the RMSE, MAE, and WMAPE. While the prediction performance of our model without CNN layer is similar to the model without Transformer layer or GCN layer, illustrating that these structures can capture the temporal and spatial characteristics of passenger flow and
availabley improve the prediction accuracy of model. We also compare the prediction performance of using one year’s or two consecutive years’ passenger flow data. It can be shown from Table 5 and Figure 10 that when using one year’s passenger flow, the prediction accuracy of GCN-Transformer is lower than using two consecutive years, which indicates that using two consecutive years’ passenger flow can better capture the characteristics of holiday passenger flow and solve the problem of limited sample size of holiday passenger flow. As for the influence of microblog volumes, we compare the prediction performance of GCN-Transformer with and without microblog volumes. Table 5 and Figure 10 manifest that utilizing the microblog volumes can capture the potential trend of holiday passenger flow and achieve favorable prediction results. All of the above illustrates that our model architecture can fully capture the spatial and temporal features of holiday passenger flow and the social media features can identify different date attribute well, which ensures the strong prediction performance of holiday passenger flow.

Table 5 Performance comparison of different conditions

| Model               | 2020 New Year | 2021 New Year | Average |
|---------------------|---------------|---------------|---------|
|                     | RMSE  | MAE   | WMA PE | RMSE  | MAE   | WMA PE | RMSE  | MAE   | WMA   |
| Without ResNet      | 29.931 | 17.954 | 0.137  | 33.376 | 17.334 | 0.173  | 31.653 | 17.644 | 0.155  |
| Without CNN         | 28.967 | 17.908 | 0.137  | 33.017 | 17.378 | 0.174  | 30.992 | 17.643 | 0.156  |
| Without Transformer | 28.011 | 17.539 | 0.134  | 33.788 | 17.657 | 0.176  | 30.899 | 17.598 | 0.155  |
| Without GCN         | 29.437 | 17.440 | 0.133  | 32.207 | 17.415 | 0.174  | 30.822 | 17.427 | 0.153  |
| Using One-year data | 29.302 | 16.263 | 0.135  | 30.750 | 16.368 | 0.173  | 30.026 | 16.32  | 0.154  |
| Without Microblog   | 26.950 | 16.420 | 0.125  | 30.070 | 16.322 | 0.163  | 28.51  | 16.371 | 0.144  |
| GCN-Transformer     | 26.606 | 16.306 | 0.124  | 29.912 | 15.826 | 0.158  | 28.259 | 16.066 | 0.141  |
Predicting short-term passenger flow on holidays for URT systems is a significantly challenging task for traffic management because of its suddenness and irregularity. In our study, we develop a deep learning architecture called GCN-Transformer to conduct the short-term passenger flow on holidays. The main conclusions are summarized as follows.

1. The proposed GCN-Transformer has significant advantages to capture spatial-temporal correlations and topological information of the urban rail transit network.
2. The GCN-Transformer utilizing the microblog volumes outperforms other benchmark models and achieves favorable prediction accuracy. The improvements compared with the best (existing) models are RMSE of 5.82%, MAE of 6.74%, and WMAPE of 6.62%, respectively. This indicates that social media data can be regarded as an effective data source to improve the accuracy of passenger flow prediction.
3. The results tested on two real-world datasets reveal that the GCN-Transformer performs well under various settings and has favorable robustness, indicating that our model meets the requirement of prediction accuracy in practical and has great potential to be applied in the real world.

However, there are several limitations to our study. For example, the GCN-Transformer performs well during the New Year holiday. Whether it can be applied to predict the passenger flow during other holidays has not been verified. In addition, we only use a single feature, social media volumes, when capturing the factors affecting passenger flow. In the future, we will consider multi-features, such as emotions of social media users, weather conditions, and date attributes, which may improve the accuracy of
the prediction. Besides, whether the proposed model can be applied to other scenarios, such as speed prediction is also worth studying in future work.

**Conflicts of Interest**

The authors declare no conflict of interest.

**Acknowledgments**

We wish to thank the anonymous reviewers for the valuable comments, suggestions, and discussions. This work was supported by the Fundamental Research Funds for the Central Universities (No. 2021RC270) and the National Natural Science Foundation of China (Nos. 71621001, 71825004).

**References**

Castro-Neto, M., Jeong, Y., Jeong, M. & Han, L. D. (2009), "Online-SVR for short-term traffic flow prediction under typical and atypical traffic conditions", *Expert Systems with Applications*, Vol. 36 No. 3, pp. 6164-6173.

Chaniotakis, E. & Antoniou, C. (2015), "Use of geotagged social media in urban settings: Empirical evidence on its potential from Twitter", in *18th International Conference on Intelligent Transportation Systems*. IEEE, pp. 214-219.

Chen, R. & Liang, C. (2015), "Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm", *Applied Soft Computing*, Vol. 26, pp. 435-443.

Essien, A. & Petrounias, L. (2020), "A deep-learning model for urban traffic flow prediction with traffic events mined from Twitter", *World Wide Web*, Vol. 24 No. 4, pp. 1345-1368.

Gaoxiang, Z. & Jinjin, T. (2020), "Forecast of urban rail transit passenger flow in holidays based on support vector machine model", *5th International Conference on Electromechanical Control Technology and Transportation*. IEEE, pp. 585-589.

Geng, X., Yaguang, L. & Leye, W. (2019), "Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting", *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. No. 01, pp. 3656-3663.

He, K., Zhang, X. & Ren, S. (2015), "Deep residual learning for image recognition", in *Proceeding of the IEEE conference on computer vision and pattern recognition*, Boston, Massachusetts, USA, pp. 770-778.

Hong, W. & Dong, Y. (2011), "Forecasting urban traffic flow by SVR with continuous ACO", *Applied Mathematical Modelling*, Vol. 35 No. 03, pp. 1282-1291.

Hu, W. & Yan, L. (2016), "A short-term traffic flow forecasting method based on the hybrid PSO-SVR", *Neural Processing Letters*, Vol. 43 No. 01, pp. 155-172.

Jinlei, Z., Feng, C. & Yinan, G. (2020), "Multi-graph convolutional network for short-term passenger flow forecasting in urban rail transit", *IET Intelligent Transport Systems*, Vol. 14 No. 10, pp. 1210-1217.

Kipf, T. N. & Welling, M. (2016), "Semi-supervised classification with graph convolutional networks", *arXiv preprint arXiv:1609.02907*.

Kumar, S. V. & Vanajakshi, L. (2015), "Short-term traffic flow prediction using seasonal ARIMA model with limited input data", *European Transport Research Review*, Vol. 7 No. 3, pp. 1-9.

Leshem, G. (2007), "Traffic flow prediction using Adaboost algorithm with the random forest as a weak learner", *Proceedings of world academy of science, engineering and technology*, Vol. 19, pp. 193-198.
Li, H. & Zhang, J. (2022), "Graph-GAN a spatial-temporal nearest network for short-term passenger flow prediction in urban rail transit systems", arXiv e-prints, arXiv: 2202.06727.

Li, R. & Xiang, M. (2020) "Research and comparison of ARIMA and grey prediction models for subway traffic forecasting", International Conference on Intelligent Computing, Automation and Systems (ICICAS), Vol. 01, pp. 63-67.

Liu, L. (2017), "A novel passenger flow prediction model using deep learning methods", Transportation Research Part C: Emerging Technologies, Vol 84, pp. 74-91.

Li, X. & Yang, J. (2016), "Input data selection for daily traffic flow forecasting through contextual mining and intra-day pattern recognition", Expert Systems with Applications, Vol. 176, pp. 114902.

Ma, X., Tao, Z., Wang, Y., Yu, H. & Wang, Y. (2015), "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data", Transportation Research Part C: Emerging Technologies, Vol. 54, pp. 187-197.

Ni, M., He, Q. & Gao, J. (2017), "Forecasting the subway passenger flow under event occurrences with social media", IEEE Transactions on Intelligent Transportation Systems, Vol. 18 No. 06, pp. 1623-1632.

Ni, M., He, Q. & Jing, G. (2014) "Using social media to predict traffic flow under special event conditions", The 93rd annual meeting of transportation research board. pp. 1-10

Ren, Y., Cheng, T. & Zhang, Y. (2019), "Deep spatio-temporal residual neural networks for road network based data modeling", International Journal of Geographical Information Science, Vol. 33, pp. 1894-1912.

Roy, K. C., Hasan, S., Culotta, A. & Eluru, N. (2021), "Predicting traffic demand during hurricane evacuation using real-time data from transportation systems and social media", Transportation Research Part C: Emerging Technologies, Vol. 131, pp. 103339.

Shi, X. & Chen, Z. (2015), "Convolutional LSTM network a machine learning approach for precipitation nowcasting", Advances in neural information Processing Systems, Vol. 28, pp. 11-21.

Shiyang, L., Xiaoyong, J. & Yao, X. (2019), "Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting", Advances in Neural Information Processing Systems, Vol. 32, pp. 15-25.

Vaswani, A., Shazeer, N. & Parmar, N. (2017), "Attention Is all you need", Advances in neural information processing systems, Vol. 30, pp. 12-25.

Wen, K., Zhao, G., He, B., Ma, J. & Zhang, H. (2022), "A decomposition-based forecasting method with transfer learning for railway short-term passenger flow in holidays", Expert Systems with Applications, Vol. 189, pp. 116102.

Wu, Y. & Tan, H. (2016), "Short-term traffic flow forecasting with spatial-temporal correlation in a hybrid deep learning framework", arXiv e-prints, arXiv:1612.01022.

Xie, B. & Sun, Y. (2020), "Travel characteristics analysis and passenger flow prediction of intercity shuttles in the Pearl River delta on holidays", Sustainability, Vol. 7249 No. 12(18), pp. 11-18.

Xue, G., Liu, S., Ren, L., Ma, Y. & Gong, D. (2022), "Forecasting the subway passenger flow under event occurrences with multivariate disturbances", Expert Systems with Applications, Vol. 188, pp. 116057.

Yan, S., Xiong, Y. & Lin, D. (2018) "Spatial-temporal graph convolutional networks for skeleton-
based action recognition. Thirty-second AAAI conference on artificial intelligence.(Early access).

Yang, D., Chen, K. & Yang, M. (2019), "Urban rail transit passenger flow forecast based on LSTM with enhanced long-term", IET Intelligent Transport Systems. Vol. 13, pp.1475-1482.

Yu, B. & Lee, Y. (2020), "Forecasting road traffic speeds by considering area-wide spatiotemporal dependencies based on a graph convolutional neural network (GCN)", Transportation Research Part C: Emerging Technologies, Vol. 114, pp. 189-204.

Zeng, D. & Xu, J. (2008), "Short-term traffic flow prediction using hybrid ARIMA and ANN models", Workshop on Power Electronics and Intelligent Transportation System, Vol.15, pp.621-625.

Zeng, P., Sheng, C. & Jin, M. (2019), "A learning framework based on weighted knowledge transfer for holiday load forecasting", Journal of Modern Power Systems and Clean Energy,

Zhang, J. & Che, H. (2021), "Short-term origin-destination demand prediction in urban rail transit systems: A channel-wise attentive split-convolutional neural network method", Transportation Research Part C: Emerging Technologies, Vol. 124, pp. 102928.

Zhang, J. & Chen, F. (2020), "Deep Learning Architecture for Short-Term Passenger Flow Forecasting in Urban Rail Transit", IEEE Transactions on Intelligent Transportation Systems, Vol. 22, pp. 7004-7014.

Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M. & Li, H. (2020), "T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction", IEEE Transactions on Intelligent Transportation Systems, Vol. 21 No. 9, pp. 3848-3858.

Zheng, W. & Lee, D. (2006), "Short-Term Freeway Traffic Flow Prediction Bayesian Combined Neural Network Approach", Journal of Transportation Engineering, Vol. 132, pp. 114-121.