Multimedia Tools and Applications
https://doi.org/10.1007/s11042-023-14388-z

1229: MULTIMEDIA DATA ANALYSIS FOR SMART CITY ENVIRONMENT SAFETY

# A method for short-term passenger flow prediction in urban rail transit based on deep learning

Ningning Dong[1] · Tiezhu Li[1] · Tianhao Liu[1] · Ran Tu[1] · Fei Lin[2] · Hui Liu[3] · Yiyong Bo[3]

Received: 20 May 2022 / Revised: 22 August 2022 / Accepted: 6 January 2023


**Abstract**
Short-term passenger flow prediction is a critical component of urban rail transit operations. However, predictions of passenger flow are mostly focused on one station, and land use, which has a substantial impact on passenger flow variation, has not been taken into account. A model termed the temporal-spatial network long short-term memory model (TNS-LSTM) is developed to solve the forecasting gap for the metro inbound/outbound passenger flow. The model introduces the spatial characteristics of the land use by extracting the point of interest (POI) data instead of merely considering temporal characteristics and network characteristics. The spatial-temporal network matrix is designed through the K-Means clustering model, extraction for temporal characteristics analysis for land use, and establishment of an origin-destination station matrix. Furthermore, the prediction of short-term passenger flow is implemented for multiple stations in the metro network. Finally, a case study based on actual data from the Nanjing metro is carried out, and the results demonstrate that the proposed model can not only avoid the complexity of constructing the numerous models for each station in urban rail transit but also improve the prediction accuracy and save a substantial amount of time.

**Keywords** Urban rail transit · Passenger flow · A deep learning · Prediction · TNS-LSTM

## 1 Introduction

The urban railway plays an important role in modern cities and has attracted increasing attention due to its reliable and efficient service [39]. However, with the continuous increase in passenger flow, the urban railway suffers from a series of problems, such as metro congestion and insecurity, particularly during peak hours [20]. Accordingly, an accurate and robust short-term passenger flow prediction (STPFP) becomes a prime component of the operation of the urban railway transport system to solve the problems. It can not only help

✉ Tiezhu Li
litiezhu@seu.edu.cn

Extended author information available on the last page of the article

Published online: 17 January 2023

managers arrange operational organization, train schedules, security deployment, and timetable optimization, but it can also help travelers rationally choose a travel route and avoid traffic congestion [36]. In addition, machine learning, deep learning techniques, and big data are becoming increasingly mature and applied to many fields [2, 3, 7], which contribute to good predictions in urban rail transit.

The concept of short-term passenger flow has different interpretations in different research. However, the time granularity of the short-term is mainly 1 to 30 minutes [37, 38]. STPFP has been extensively researched over the past few years. The autoregressive integrated moving average model (ARIMA) and its advanced models, like the seasonal ARIMA (SARIMA), are used as the representatives of regressive time series models for the prediction [26, 35]. The algorithm of ARIMA is simple and maneuverable, but stationary time series are required [34]. Simultaneously, the Kalman filtering model was revised for STPFP, such as the introduction of historical prediction error or nonparametric regression into the traditional Kalman filtering method [19]. The Kalman filtering model and its revised models can be used to process stable or non-stationary data. However, the above-mentioned models perform well and robustly in capturing linear relationships but are disadvantageous when processing data with the nonlinear feature [27].

Machine learning models [1], for instance, K-nearest neighbor [10], the Bayesian method, the random forest algorithm, and support vector machine (SVM), have been developed to overcome the limitations of linear forecasting models. Amiri et al. [4] evaluated the effectiveness of machine learning algorithms like decision trees and the random forest for prediction, and the results showed that machine learning algorithms have higher accuracy due to their ability to accommodate non-linearity. Roos et al. [28] proposed a method that was effective with incomplete data based on dynamic Bayesian networks to forecast short-term passenger flow. Furthermore, Feng et al. [8] presented a novel method for STPFP based on an adaptive multi-kernel SVM with spatial-temporal correlation. The hybrid kernel's weight was adjusted adaptively according to the changing tendency of real-time traffic flow, which was a feature of this model. Sun et al. [31] combined Wavelet and SVM methods to predict various types of passenger flows in the urban railway system. The support vector regression model (SVR) has also been extensively employed in prediction. Jeong et al. [17] developed a prediction model, online learning weighted support-vector regression, which remedied the deficiency of offline methods.

Long short-term memory (LSTM) can effectively overcome the problems of the vanishing gradient problem and exploding gradient, which may cause null value of the model loss and failure in prediction [14]. In addition, LSTM can capture temporal dependencies well [13]. Recently, LSTM has been extensively applied to forecasts for higher accuracy. Li et al. [22] forecasted holiday travel demand based on LSTM. The inputs of the model were a two-dimensional matrix containing the spatial and temporal characteristics of the data on holiday travel demand. Liu et al. [23] presented an end-to-end deep learning architecture based on LSTM to forecast passenger flow. The architecture integrated the effects of temporal characteristics, spatial characteristics between stations and metro operational properties for STPFP.

Shahid et al. [29] utilized a genetic algorithm to optimize the parameters such as the number of neurons in LSTM layers for the prediction. The attention mechanism was introduced to LSTM to improve the performance of modeling sequential data within multiple time-steps [11, 21]. Apart from the optimizing of LSTM, hybrid models, such as the combination of LSTM with a certain method, have also been developed for prediction. Zhang et al. [38] clustered the stations that have similar passenger flow variation trends and passenger flow volume

characteristics using a two-step K-Means clustering model. The prediction model based on the clustering results and LSTM improved the prediction performance.

Despite, researchers implemented the prediction of the passenger flow considering different factors. For example, the temporalities [5] and network information between stations [24] have been considered, as well as the influence of external factors, such as weather conditions [18]. However, it should be noted that the factors mentioned above vary with time. Furthermore, the land use around the subway station affects the transportation occurrence quantity and attraction quantity in the station and leads to the different variation trends of the metro passenger flow. For example, the high proportion of commercial buildings around the station attracts people for work and shopping. Therefore, the passenger flow characteristic presents high outbound passenger flow and high inbound passenger flow during the morning and evening, respectively. However, different from the factors changing with time, the land use conditions are different at different stations. The two kinds of factors are not easily integrated to serve STPFP and are rarely explored by scholars because of the complexity and difficulty in modeling and processing big data to predict at the network level. The factors that change with time or not, which are important for accurate prediction, should be considered for STPFP. Furthermore, the object of STPFP is mainly for one station in previous research, rather than in the perspective of the network of metro stations.

Therefore, an improved LSTM (i.e., TSN-LSTM) is proposed to fill in the gaps. The TSN-LSTM introduces the spatial characteristics of land use, besides temporal and network characteristics between stations. In addition, short-term passenger volume prediction for network subway stations rather than one station is accomplished. The proposed model improves the accuracy and efficiency of passenger flow forecasting.

The detailed contributions of this paper are as follows: First, a deep-learning architecture named TSN-LSTM is proposed. Herein, the architecture has multiple components, including not only the temporal characteristics and network characteristics between stations but also spatial characteristics such as the land use of the subway stations. Second, considering the phenomenon that exchanges working days and holidays in China, data for workdays and non-workdays on the calendar are extracted for exploration of temporal dependencies. Coupled with the OD matrix, the network characteristics are analyzed and introduced into TSN-LSTM. Third, in order to explore the character of land use and prediction for multiple stations, subway stations are clustered by employing the K-Means clustering method, and the data of POI is introduced. The regularity of the location, variation trends, and land use of the stations in every cluster have been studied. Finally, the different contribution degrees of temporal dependencies, network characteristics between stations, and spatial characteristics to the prediction accuracy and the temporal-spatial feature of every cluster for stations are analyzed.

The remainder of this paper is organized as follows: Section 2 provides the data source and preprocessing. Section 3 presents a deep-learning-based metro passenger flow architecture, TSN-LSTM. This case is carried out based on real metro data along with the analysis of the results in Section 4. The conclusion is summarized in section 5.

## 2 Data description and preprocessing

There were ten metro lines and 159 stations in Nanjing in 2019. The automatic fare collection (AFC) data from March 4, 2019 to June 2, 2019 was collected from Nanjing metro and was composed of 270 million data records. They contain a dataset of 91 consecutive days from 05:00 to 24:00 with weekdays, weekends, and holidays. The useful relevant elements extracted

from the original dataset are transaction type, entering station ID, entering date, entering time, exiting station ID, exiting date, exiting time, line ID, and device type as shown in Table 1. Numbers are employed instead of field information for the extraction of the passenger flow time series. For example, 50 in Table 1 represents one of the transaction types, mainly including single tickets, metro cards, and mobile payments. Furthermore, the data cleaning is conducted, such as the incorrect entering/exiting time like '2015-05-18'.

## 3 Methodology

Based on the LSTM and K-Means algorithms, we propose a deep learning-based architecture. The framework of the proposed methodology for predicting the inbound/outbound passenger flow is indicated in Fig. 1. The collected data includes AFC data, calendar data, and POI data. Subsequently, the spatial-temporal network matrix is designed where temporal characteristics, spatial characteristics, and network characteristics are extracted and integrated. The K-Means clustering model is used to classify the stations based on passenger flow trends. The models are developed for every cluster to train the data and predict the passenger flow. Eighty percent of the processed data is the training set, while the remaining samples are the testing set. The Min-Max normalization method, supervised learning backpropagation, and Adam algorithm are utilized to train the model. Finally, the predicted values of the inbound/outbound passenger flow are obtained. The results of the TSN-LSTM model are compared with other predicted models such as SVR, ARIMA, and LSTM for accuracy analysis. Furthermore, the different influences of temporal characteristics, spatial characteristics, and network characteristics on the prediction are explored by changing the input variables.

### 3.1 The LSTM

LSTM is composed of a cell, an input gate, forget gate, and an output gate, which are expressed as $c_t$, $i_t$, $f_t$, and $o_t$:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i), \tag{1}$$

$$f_t = \sigma\left(W_f[h_{t-1}, x_t] + b_f\right), \tag{2}$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o), \tag{3}$$

$$\widetilde{c}_t = tanh(W_c[h_{t-1}, x_t] + b_c), \tag{4}$$

**Table 1** Example of Processed AFC Data for April 15, 2019

| Transaction type | Entering station ID | Entering date | Entering time | Exiting station ID | Exiting date | Exiting time | Line ID | Device type |
|---|---|---|---|---|---|---|---|---|
| 50 | 75 | 2019-04-15 | 8:22:52 | 7 | 2019-04-15 | 9:12:24 | 1 | 2 |
| 50 | 93 | 2019-04-15 | 8:37:22 | 7 | 2019-04-15 | 9:12:38 | 1 | 2 |
| 50 | 93 | 2019-04-15 | 8:34:08 | 7 | 2019-04-15 | 9:12:44 | 1 | 2 |

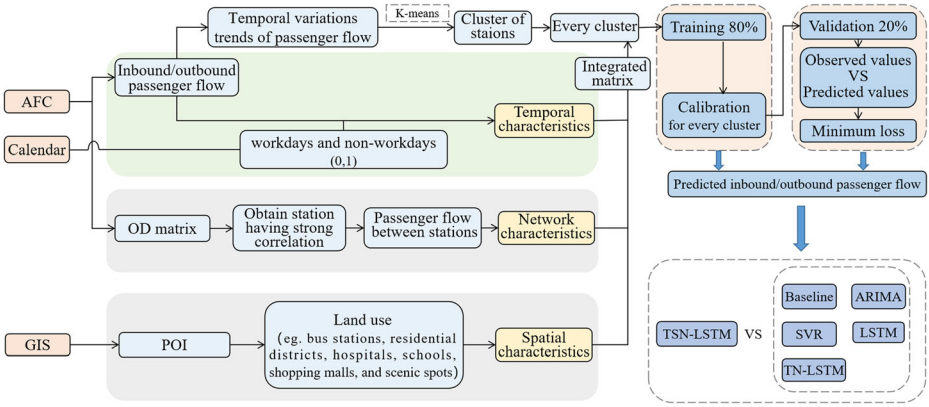



**Fig. 1** The structure of the TSN-LSTM model

$$c_t = f_t \odot c_{t-1} + i_t \odot \widetilde{c}_t, \tag{5}$$

$$h_t = o_t \odot tanh(c_t), \tag{6}$$

$$\sigma(x) = \frac{1}{1+e^{-x}}, \tag{7}$$

$$tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}, \tag{8}$$

where $x_t$ is the original input data about passenger flow at time $t$, σ(x) and tanh(x) are the sigmoid activation function and the hyperbolic tangent activation function, respectively, $h_{t-1}$ is the hidden state at time $t-1$, $W_i$, $W_f$, $W_o$ and $W_c$ are the respective weight matrices for the different gates, $b_i$, $b_f$, $b_o$, $b_c$ are their bias vectors for each function, $c_t$ is the cell state, $\odot$ is the Hadamard product.

As illustrated in Fig. 2, the three gates protect and control the status of the cell by permitting information to pass selectively. In other words, the information can be deleted or added, which is reflected in the cell utilizing the gates. The input data about passenger flow at time $t$ and the hidden output $h_{t-1}$ at time $t-1$, and cell state $c_{t-1}$ are given. Then the cell state $c_t$ and hidden output $h_t$ are obtained under the regulation of the three doors.

## 3.2 Deep learning-based metro passenger flow architecture

### 3.2.1 Clustering method of stations

K-means is undoubtedly the most popularly employed clustering algorithm [6]. K-means clustering not only has the advantages of simplicity and lower computational complexity [25], but also performs well in exploring and analyzing data where the primary objective is to

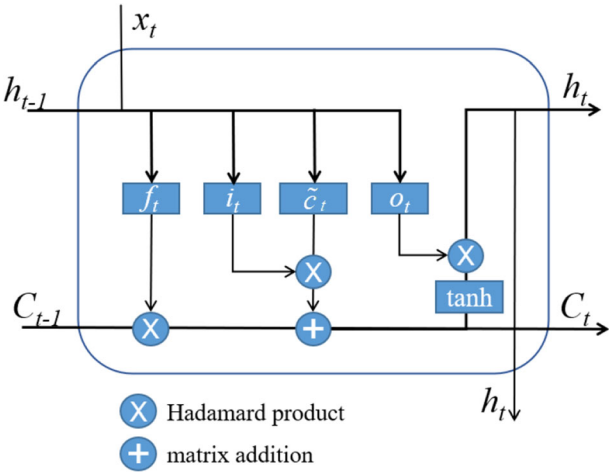



**Fig. 2** Illustration of the hidden layer unit of LSTM model

categorize similar observations into a cluster [12]. Furthermore, K-means is an unsupervised classification method of patterns into groups [16], namely, applied to exploratory data, although the classification standard is not undefined in advance. Therefore, we use the K-means algorithm to explore the similar variations in trends of passenger flow and land use, which provides a prerequisite step for networked traffic forecasting.

The passenger flow variation trend is consistent regardless of the time interval. It is worth noting that the greater the time interval, the more obvious the changes in passenger flow. As observed in Fig. 3, the peak values in intervals of an hour, half-hour, and 15-min are about 9000, 5000, and 2800, respectively, and the most common low levels in the three intervals are about 1000, 500, and 200. The difference values in the intervals of an hour, half-hour, and 15-min between the peak and stable values are 8000, 4500, and 2300, respectively. It means that the difference increases as the time interval enlarges and the changes in passenger flow are more obvious. Meanwhile, as the interval grows larger, the samples become smaller, making it impossible to better express the passenger flow variations feature. Therefore, we choose half an hour as the time interval to extract the trend of passenger flow.

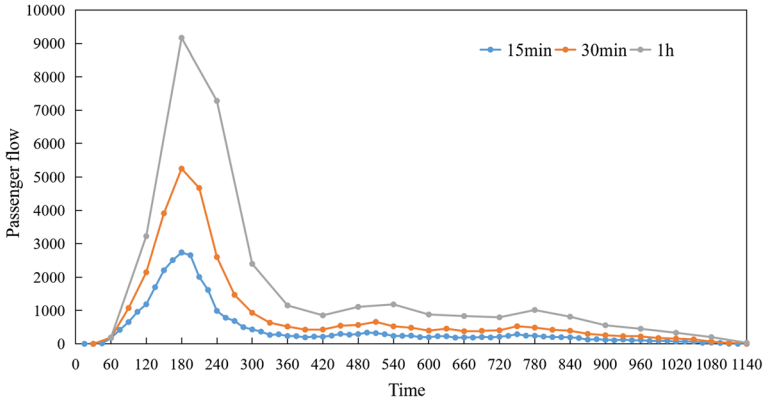



**Fig. 3** The passenger flow at different time interval

Every half-hour passenger flow of subway stations is calculated based on processed AFC data. The variable trends of passenger flow in some stations and land use around stations are similar. It's worth noting that the significant differences in passenger flow volume are presented at different stations. Therefore, the ratios of half-hour passenger flow over the maximum half-hour passenger flow [38] are introduced to eliminate the effects of passenger flow volume value for the extraction of stations that have similar characteristics in variable trends of passenger flow and land use. The ratios are indicated as:

$$P_{st} = \begin{pmatrix} \frac{r_{11}}{r_{\max 1}} & \frac{r_{12}}{r_{\max 1}} & \frac{r_{13}}{r_{\max 1}} & \cdots & \frac{r_{1n}}{r_{\max 1}} \\ \frac{r_{21}}{r_{\max 2}} & \frac{r_{22}}{r_{\max 2}} & \frac{r_{23}}{r_{\max 2}} & \cdots & \frac{r_{2n}}{r_{\max 2}} \\ \frac{r_{31}}{r_{\max 3}} & \frac{r_{32}}{r_{\max 3}} & \frac{r_{33}}{r_{\max 3}} & \cdots & \frac{r_{3n}}{r_{\max 3}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{r_{s1}}{r_{\max s}} & \frac{r_{s2}}{r_{\max s}} & \frac{r_{s3}}{r_{\max s}} & \cdots & \frac{r_{sn}}{r_{\max s}} \end{pmatrix}, \tag{9}$$

where $P_{st}$ represents the $n$th ratios of station $s$, $n$ is the number of time intervals at metro operating time in a day, $r_{maxs}$ is the maximum half-hour passenger flow at station $s$, $r_{st}$ is the $n$th half-hour passenger flow of station $s$.

The stations are clustered considering ratios of half-hour passenger flow, where the K-means algorithm is performed employing the matrix $P_{st}$. The cluster is formed by judging the distance between the centroid and the sample. Subsequently, iterative updating is carried out until the position of the centroid basically does not change based on Euclidean distance. The distortion function is used to search for the optimal number of clusters as presented:

$$di = \sum_{i=1}^{a} \sum_{s=1}^{e} \sum_{t=1}^{n} min \| p_{st} - c_i \|^2, \tag{10}$$

where $di$ is the distortion, $p_{st}$ is the value of passenger flow at $t$ in the station $s$, $c_i$ is the clustering centroid, $a$ and $e$ are the numbers of clusters and stations in each cluster, respectively, $e$ is the number of stations in each cluster. All the passenger flows at $t$ time are used as feature parameters to cluster the stations of the subway network. The main objective of K-means is to optimize the function in Eq. (10) [9].

A decrement of distortion between two adjacent clusters is given by:

$$de = |di_a - di_{a+1}|, \tag{11}$$

where $de$ represents the decrement. The proper number of clusters is determined based on the distortion function and the decrement using the "elbow method" [38].

### 3.2.2 Temporal characteristics

The flow at the same time interval on one day shares a higher similarity with the adjacent one. The trends of passenger flow on different days are essentially consistent; for example, Fig. 4 depicts the passenger flow in the Liuzhoudonglu station for five days. During the five-day morning peak, the passenger flow is relatively large, and the passenger flow remains low after ten o'clock. The concurrent trends of passenger flow for five days demonstrate a strong correlation between the adjacent days. Besides, the passenger flow in sequential dates well

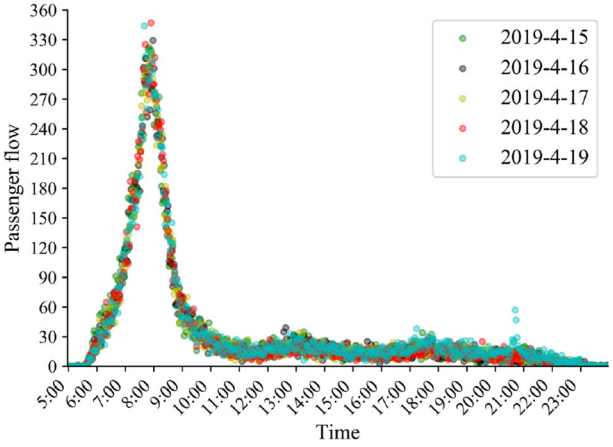



**Fig. 4** The variation of flow in Liuzhoudonglu station

reflects the regularity of the cycle, significantly between adjacent weeks. For instance, the passenger flow on weekdays is maintained at a certain level but is different from holidays or weekends. A cyclical trend is revealed among the weeks, as shown in Fig. 5, which presents the variation of flow in Liuzhoudonglu station from May 2 to June 2. It is noticeable that the government will adjust the holiday period in China. This means that there are holidays during the week and work during the weekend. Therefore, we propose to combine the variation in passenger flow with the data of workday and non-workday on the calendar, rather than separating holidays and weekends from all other dates, and avoid considering whether holidays and weekends overlap.

Suppose that $k$ is the station to be predicted. The inbound passenger flow is extracted from the processed data, and it is defined as $FI_k^t = \{fi_k^{t-m}, fi_k^{t-2m}, fi_k^{t_0-3m}, \cdots, fi_k^{t-nm}\}$, where $k$ is a predicted station, $m$ is a fixed time interval, $fi_k^{t-nm}$ is the inbound passenger flow of station $s$ at the time $t - nm$. The outbound passenger flow is defined as $FO_k^t = \{fo_k^{t-m}, fo_k^{t-2m}, fo_k^{t_0-3m}, \cdots, fo_k^{t-nm}\}$, $fo_s^{t-nm}$ is the outbound passenger flow of station $s$ at the time $t - nm$. The data of workday and non-workday on the calendar are not continuous numerical variables. A hot encoding is used to encode this categorical feature. For example, suppose '01' indicates a non-workday and '10' indicates a workday.

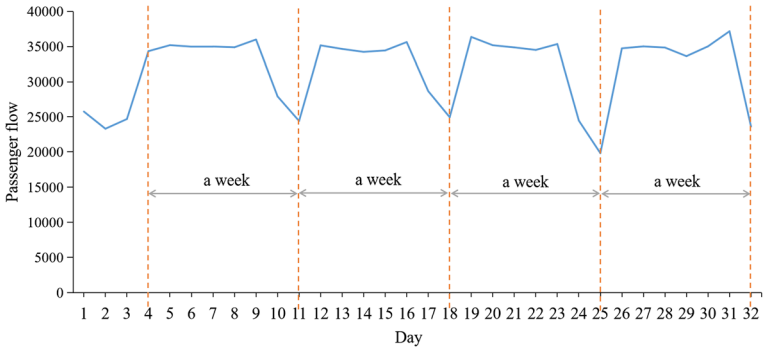



**Fig. 5** The variation of flow in Liuzhoudonglu station

### 3.2.3 Network characteristics

Normally, the input of the LSTM is a sequence that expresses the variation of historical passenger volume. According to this sequence, the predicted value of passenger flow is acquired. In fact, urban rail transit is a complex network system, and the passenger flows of some stations are closely related. So, passenger flow prediction is conducted considering not only the data of a single station but also the interactions between stations. Therefore, the OD matrix to predict passenger volume has been introduced for urban rail [37]. We exploit inbound/outbound time to search the passenger route for evaluation of the stations that have higher spatial correlation with a predicted station among the whole network. For example, the outbound time of the predicted station is used to search the passenger route, and sequentially, the impact of the passenger flow from other stations is explored. In addition, the determination of the stations that have high spatial correlation involves two parts: the passenger flow contribution for outbound passengers and passenger flow allocation from inbound passenger flow, and it is conducted by exploiting the OD matrix.

The expression of the OD matrix for passenger flow contribution is extracted. Based on the outbound time of passengers, the inbound stations are traced, and the passenger flow contributions from the stations of the whole urban rail transit network are extracted as:

$$f_{sk}^{t} = \begin{pmatrix} f_{1k}^{t-m} & f_{1k}^{t-2m} & f_{1k}^{t-3m} & \cdots & f_{1k}^{t-nm} \\ f_{2k}^{t-m} & f_{2k}^{t-2m} & f_{2k}^{t-3m} & \cdots & f_{2k}^{t-nm} \\ f_{3k}^{t-m} & f_{3k}^{t-2m} & f_{3k}^{t-3m} & \cdots & f_{3k}^{t-nm} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{sk}^{t-m} & f_{sk}^{t-2m} & f_{sk}^{t-3m} & \cdots & f_{sk}^{t-nm} \end{pmatrix}, \tag{12}$$

where $f_{sk}^{t-nm}$ is the passenger flow from station $s$ to station $k$ at the time $t - nm$. Here, the passenger flow of a day is calculated from one station to the station $k$:

$$d_{sk} = \sum_{p=0}^{n} f_{sk}^{t-pm}, \tag{13}$$

So, the sequence of passenger flow in a day from different stations to $k$ is denoted as $D_{sk}^{\boldsymbol{out}} = \{d_{1k}, d_{2k}, d_{3k}, \cdots, d_{sk}\}$.

$D_{sk}^{\boldsymbol{out}}$ is sorted from the largest to the smallest. The top $x$ stations are determined as the stations that have a higher spatial correlation. Then, passenger flow from station $j$ to station $k$ obtained at $m$ time interval as expressed as:

$$F_{jk}^{t} = \left\{ f_{jk}^{t-m}, f_{jk}^{t-2m}, f_{jk}^{t-3m}, \cdots, f_{jk}^{t-nm} \right\}, \tag{14}$$

where $j \in \{1, 2, 3, \cdots x\}$.

Based on the inbound time of passengers, the outbound stations are traced, and the passenger flows from station $k$ to stations of the whole urban rail transit network are extracted in the following:

$$f_{ks}^{t} = \begin{pmatrix} f_{k1}^{t-m} & f_{k1}^{t-2m} & f_{k1}^{t-3m} & \cdots & f_{k1}^{t-nm} \\ f_{k2}^{t-m} & f_{k2}^{t-2m} & f_{k2}^{t-3m} & \cdots & f_{k2}^{t-nm} \\ f_{k3}^{t-m} & f_{k3}^{t-2m} & f_{k3}^{t-3m} & \cdots & f_{k3}^{t-nm} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{ks}^{t-m} & f_{ks}^{t-2m} & f_{ks}^{t-3m} & \cdots & f_{ks}^{t-nm} \end{pmatrix}, \tag{15}$$

where $f_{ks}^{t-nm}$ is the passenger flow from station $k$ to station $s$ at the time of $t - nm$. Here, the passenger flow of a day is calculated from station $k$ to the one of the stations given by:

$$d_{ks} = \sum_{p=0}^{n} f_{ks}^{t-pm}. \tag{16}$$

So, the sequence of passenger flow in a day from station $k$ to different stations is denoted as $D_{ks}^{in} = \{d_{k1}, d_{k2}, d_{k3}, \cdots, d_{ks}\}$.

$D_{ks}^{in}$ is sorted from largest to smallest. The top $x$ stations are determined as the stations that have a higher spatial correlation. Then, passenger flow from station $k$ to station $j$ obtained at $m$ time interval as expressed as:

$$F_{kj}^{t} = \left\{ f_{kj}^{t-m}, f_{kj}^{t-2m}, f_{kj}^{t-3m}, \cdots, f_{kj}^{t-nm} \right\}, \tag{17}$$

where $j \in \{1, 2, 3, \cdots x\}$.

### 3.2.4 Spacial characteristics

In the planning of public transportation for subway stations, a radius of about 500 m is regarded as an appropriate service range, and it is also an acceptable and endurable distance for passengers [30, 32]. The land use around the predicted stations consists of bus stations, residential districts, hospitals, schools, shopping malls, and scenic spots. The land use around stations is investigated. For instance, suppose Liuzhoudonglu station is the predicted station. Liuzhoudonglu station on Nanjing Metro Line 3 is one of the stations with a huge passenger flow. The land use around Liuzhoudonglu station is presented in Fig. 6. The residential


**Fig. 6** The land use around Liuzhoudonglu station

districts occupy a large area. Moreover, many bus stations transport numerous travelers to the subway station. Therefore, these intensive distributions of residential districts and bus stations help to form the Liuzhoudonglu station's characteristics of massive inbound passenger flow in the morning peak hours and outbound passenger flow in the evening peak hours.

The POI data is extracted from the Baidu map and the number of these factors in the five hundred meters around the stations is investigated. The land use is integrated into:

$$N_s^t == \begin{pmatrix} st_1 & rq_1 & ho_1 & sch_1 & sho_1 & sce_1 \\ st_2 & rq_2 & ho_2 & sch_2 & sho_2 & sce_2 \\ st_3 & rq_3 & ho_3 & sch_3 & sho_3 & sce_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ st_s & rq_s & ho_s & sch_s & sho_s & sce_s \end{pmatrix}, \tag{18}$$

where $st_s$, $rq_s$, $ho_s$, $sch_s$, $sho_s$, and $sce_s$ are the number of bus stations, residential districts, hospitals, schools, shopping malls, and scenic spots around station $s$ respectively.

### 3.2.5 Training algorithm

The spatial-temporal network matrix is constructed as:

$$F_{in} == \begin{pmatrix} fi_1^{t-m} & f_{1j}^{t-m} & st_1 & rq_1 & ho_1 & sco_1 & shp_1 & sce_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ fi_1^{t-nm} & f_{1j}^{t-nm} & st_1 & rq_1 & ho_1 & sco_1 & shp_1 & sce_1 \\ fi_2^{t-m} & f_{2j}^{t-m} & st_2 & rq_2 & ho_2 & sco_2 & shp_1 & sce_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ fi_s^{t-m} & f_{sj}^{t-nm} & st_s & rq_s & ho_s & sco_s & shp_s & sce_s \end{pmatrix}. \tag{19}$$

It is integrated for the prediction of inbound passenger flow for one cluster, and the methods are the same for other clusters and outbound passenger flow. The training data is constructed. 80% of the data is the training set, while the remaining samples are the testing set. The Min-Max normalization method is utilized to scale investigated data in the range [0,1] for both the training and testing set. The supervised learning backpropagation and Adam algorithm are utilized to train the model. The parameters to be determined are epochs, the number of hidden units, and batch size. The error objective of performance metrics is minimized until the stopping criteria is reached.

The performance metrics used to evaluate the model are Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared ($R^2$) [15]:

$$\boldsymbol{RMSE} = \sqrt{\frac{\boldsymbol{1}}{n}\sum_{i=\boldsymbol{1}}^{n}\left(y_i - \widehat{y}_i\right)^{\boldsymbol{2}}}, \tag{20}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \widehat{y}_i\right|. \tag{21}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \widehat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2}, \tag{22}$$

where $y_i$ is the passenger flow during $i$th time interval. $\widehat{y}_i$ is the predicted passenger flow during $i$th time interval, $\overline{y}$ is the average value of all true passenger flow, $n$ is the number of time interval.

## 4 Results

### 4.1 Results of cluster and land use

The extracted data is the passenger flow of Nanjing subway stations every half-hour from April 15 to April 19 (Monday to Friday), 2019. Then the average data of these days is calculated and the subway stations are clustered according to passenger flow variations as mentioned in section 3.2. The variation of the distortion function with the number of clusters is indicated in Fig. 7. When the clustering number exceeds five, the decrement of distortion becomes small and remains steady, which means that the value of five can be considered as the "elbow method". Thus, the final cluster number is set as 5.

The passenger flow variation of clustering results is revealed in Fig. 8. The spatial distributions of stations in each cluster are shown in Fig. 9, where different colors express subway lines and stations in five clusters. The features of stations in the five clusters are analyzed as follows based on the land use composed of six factors: bus stations, residential districts, hospitals, schools, shopping malls, and scenic spots.

The passenger flow in cluster 1 presents a bimodal distribution. Compared to the other periods in a day, the passenger flow has a higher increment during the morning and evening peak hours. These stations in cluster 1 are mainly distributed far from the downtown area. Even two hours are required for travelers from some stations to the city center. Furthermore, the factors of land use are lower than any other cluster. There are some bus stations and residential districts around the subway station, and few hospitals, universities, shopping malls, and scenic spots around 500 m. Meanwhile, the values of passenger flow are extremely lower than the other clusters.

The passenger flow of the stations in cluster 2 remains relatively steady throughout the day, either at a high level or at a low level, without obvious morning or evening peak characteristics. The stations with higher passenger flow are hubs such as Nanjingnan Station and

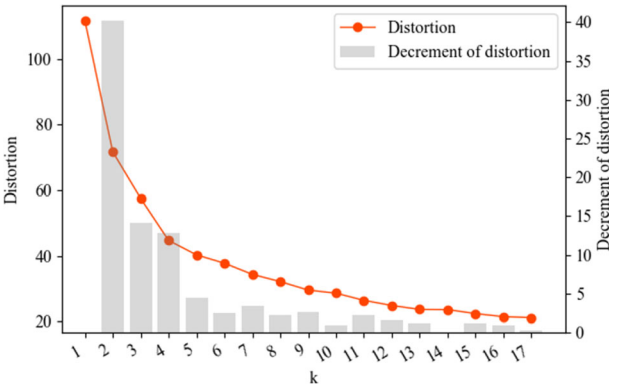
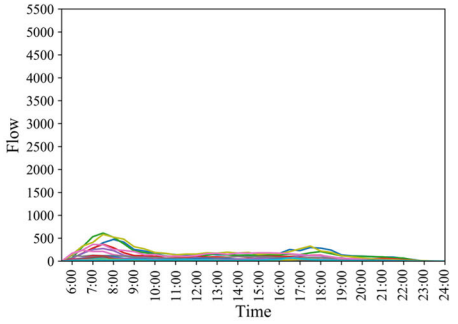
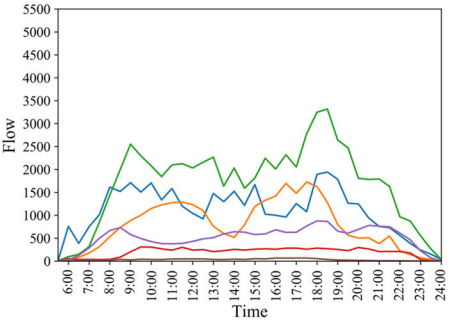


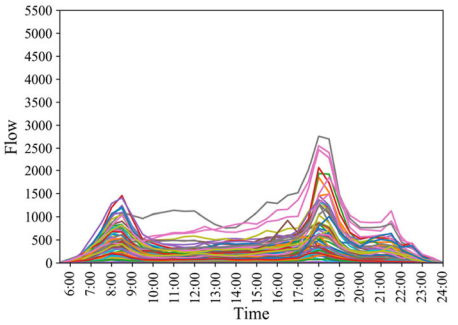

**Fig. 7** Distortion and decrement



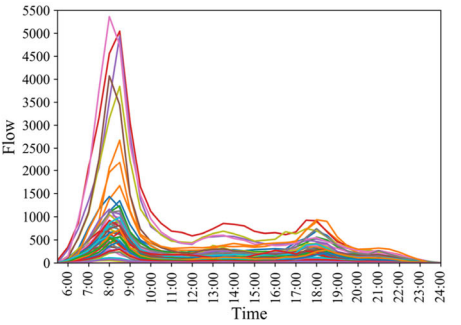

(a) Passenger flow distribution of stations in cluster 1



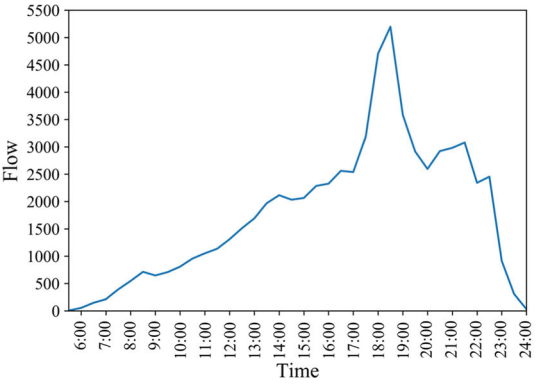

(b) Passenger flow distribution of stations in cluster 2




(c) Passenger flow distribution of stations in cluster 3




(d) Passenger flow distribution of stations in cluster 4




(e) Passenger flow distribution of stations in cluster 5

**Fig. 8** Passenger flow distribution of stations in different clusters

Lukoujichang Station. The most obvious feature of land use is that the number of bus stations is high, even reaching thirty around the stations. However, other indicators, such as hospitals and residential areas, are particularly low.

The characteristics of a bimodal distribution are indicated for the passenger flow in cluster 3. These stations are located in the downtown area, and many shopping malls, bus stations, residential districts, hospitals, and schools are distributed around the subway station. The land use of these clusters attracts people for employment and entertainment during the day.

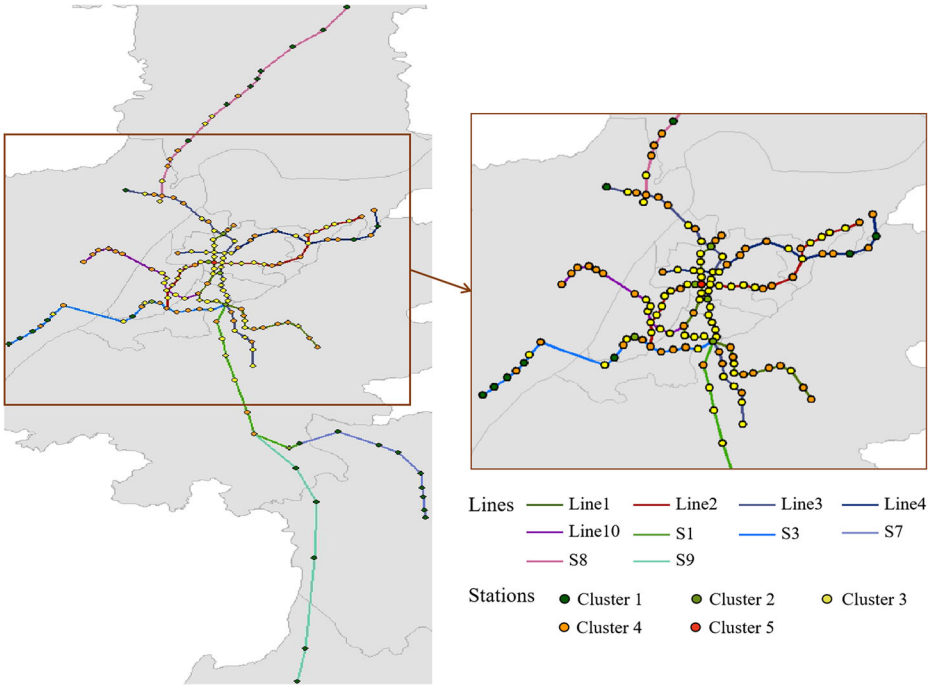



**Fig. 9** Spatial distribution of stations in different clusters in Nanjing

Furthermore, the inbound passenger flow during the evening peak hours climbs to a higher level than that in the morning peak hours.

The unimodal distribution of passenger flow is indicated in cluster 4. The stations in this cluster are mainly located around the downtown area. Residential districts occupy a lot of areas around these stations. In addition, the land use in cluster 4 has some bus stations, hospitals, schools, and shopping malls, which is less than that in cluster 3. The variation in passenger flow reveals obvious tidal characteristics affected by significant numbers of residential districts. Many passengers arrive at the station during the morning peak hours.

Only one station in cluster 5, Xinjiekou station, is located in the downtown area. There are more shopping malls in the surrounding area than in any other station. The region is a famous Chinese commercial center, attracting a large number of people for shopping or work. Therefore, a huge number of passengers leave this station in the afternoon and evening.

### 4.2 TSN-LSTM modeling results

The data collected from the Nanjing Metro System consists of 270,365,082 records from March 4, 2019 to June 2. The passenger flow in a given time period from 5:00 to 24:00 in a day is extracted from processed AFC data for all the stations. The time interval is 15 minutes in the case study. The data for workdays and non-workdays on the calendar from March 4 to June 2 is collected. The encoding is used to encode this data of discrete categories, where '01'

and '10' indicate the non-workday and workday, respectively. The passenger flow between the predicted stations and the stations in the top three strongest correlations is captured.

The data for temporal characteristics, network characteristics, and spatial characteristics are integrated according to the results of K-means clustering analysis. 80% of the data (930,468) is the training set, while the remaining samples are used as the testing set. The parameter adjusting is performed and the normalized forecast values are re-scaled. Finally, we obtain the 15-minute passenger flow prediction results for different clusters as shown in Table 2.

Table 2 denotes cluster 1 with the lowest RMSE and MAE and cluster 5 with the topmost RMSE and MAE. But it does not mean that cluster 5 has poor prediction performance because the station in cluster 5 has a large passenger volume. The $R^2$ is used to measure the linear relationship between predictive values and true values. The larger the value of $R^2$ is, the better the fitting effect is to some extent. The $R^2$ of cluster 5 reaches 0.955, which reveals the favorable prediction performance of TSN-LSTM. The TSN-LSTM model avoids the complexity of developing numerous models, so it spends less time and effort completing the prediction for the entire network of stations. For example, parameters are adjusted for training models using the proposed for 5 times, far below the 159 times in other models.

Two representative stations are chosen in every cluster, namely: 1. stations with low passenger flows and 2. stations with high passenger flows. The predicted results of two representative stations are shown in Fig. 10. The predicted model well captures the variation trend, and the results are favorably anastomotic with the observed data. The aggregation of the passenger flow in cluster 4 and cluster 5 is higher than that in the other clusters, and the observed values and predicted values are closer, which expresses the preferable predicted performance. It should be noted that the slope of passenger flow at station 2 in cluster 2 reaches 0.87, which means a better prediction, and the RMSE and MAE in cluster 1 are low. In summary, the proposed model can satisfy accuracy requirements on a network scale.

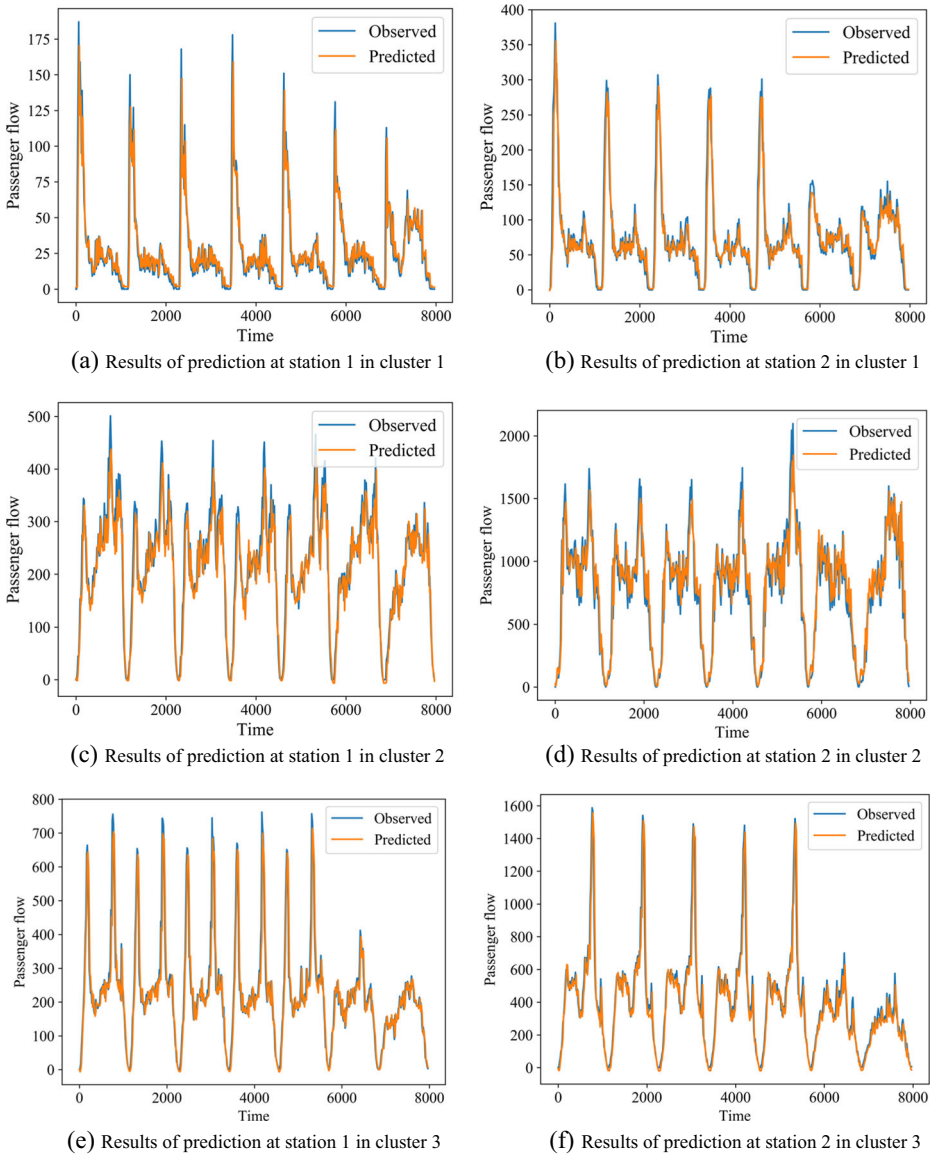
Four prediction models are constructed and compared with the proposed model, provided as follows:

**Baseline** Liu [23] mentioned that they utilized the nearest historical observations as a baseline. Based on this, the passenger flow at the nearest 15 minutes is regarded as the predicted value. For example, if the passenger flow at 7:15 is required to be predicted, the passenger flow at 7:00 is set as the prediction result to be compared with real values at 7:15. The performance of this model is used as a baseline.

**ARIMA** Three critical parameters to be determined are p (the order of autoregression), q (the order of moving average), and d (the order of difference). The Akaike information criterion

**Table 2** Performance of TSN-LSTM in different clusters

| | Clusters | Number of stations | RMSE | MAE | $R^2$ |
|---|---|---|---|---|---|
| passenger flow | 1 | 30 | 13.512 | 7.211 | 0.871 |
| prediction | 2 | 6 | 115.799 | 67.631 | 0.897 |
| Per | 3 | 65 | 46.315 | 24.407 | 0.903 |
| 15 min | 4 | 57 | 45.792 | 21.947 | 0.950 |
| | 5 | 1 | 165.044 | 96.124 | 0.955 |




(a) Results of prediction at station 1 in cluster 1

(b) Results of prediction at station 2 in cluster 1

(c) Results of prediction at station 1 in cluster 2

(d) Results of prediction at station 2 in cluster 2

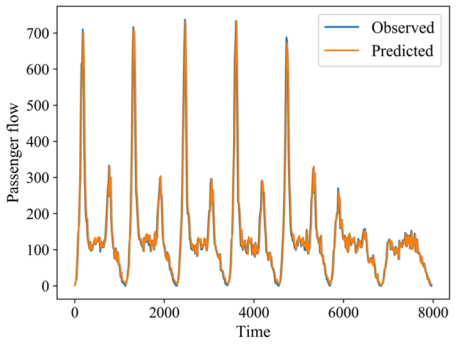
(e) Results of prediction at station 1 in cluster 3

(f) Results of prediction at station 2 in cluster 3

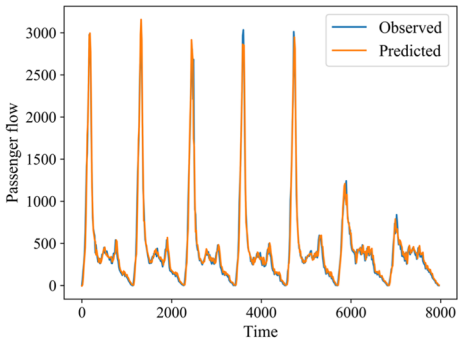
**Fig. 10** The results of passenger flow prediction in different clusters

(AIC) and Bayesian information criterion (BIC) contribute to searching for the optimal values of p and q [33]. We use the minimum value of the BIC to estimate the optimal values of p and q, considering the introduction of the number of observed values in the BIC formula and the large experimental sample size.

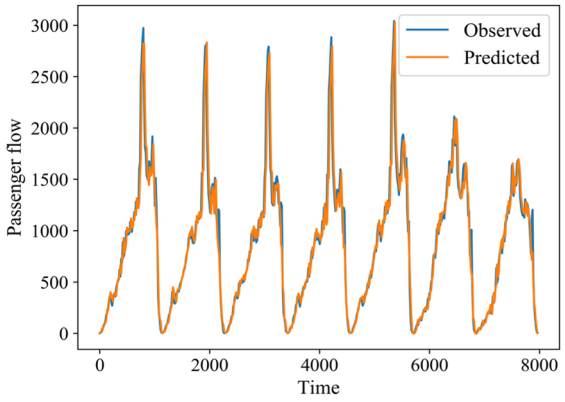
**SVR** 80% of the data is the training set, while the remaining samples are used as the testing set. The optimal SVR model with the most appropriate adjusting parameters of the kernel, C, gamma, cache size, degree, epsilon, etc. is obtained to predict the passage flow.




(g) Results of prediction at station 1 in cluster 4




(h) Results of prediction at station 2 in cluster 4




(i) Results of prediction at in cluster 5

**Fig. 10** (continued)

**LSTM** Similarly, 80% of the data is the training set, while the remaining samples are the testing set. The historical data is input, and the output is the predicted passenger volume at the station.

**TN-LSTM** Both the temporal dependencies and network characteristics are introduced in the TN-LSTM model.

Different prediction models are tested and compared based on the data of representative stations, the results of which are expressed in Table 3. From.

Table 3, we can observe that the TSN-LSTM has the best performance regardless of every cluster. On the contrary, the Baseline model has the lowest performance as the RMSE and MAE for station 2 in cluster 2 are 612.303 and 485.761, respectively. That is, the performance of other models is all above the baseline. Furthermore, in terms of MAE, a similar pattern to that of RMSE is shown. The ARIMA model performs relatively poorly on this dataset. In particular, the prediction accuracy is in a state of being out of control in the prediction for the high passenger flow. For example, the RMSE and MAE for station 2 in cluster 2 are 578.200 and 463.115, respectively. The RMSE and MAE for the station in cluster 4 reach 907.563 and 682.649, respectively. Prediction accuracy improves significantly in the SVR model relative to

**Table 3** The comparison of different prediction models

| Clusters | Cluster 1 | | | | Cluster 2 | | | | Cluster 3 | | | | Cluster 4 | | | | Cluster 5 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Station | Station 1 | | Station 2 | | Station 1 | | Station 2 | | Station 1 | | Station 2 | | Station 1 | | Station 2 | | | |
| metrics | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE | RMSE | MAE |
| Baseline | 43.112 | 28.437 | 89.395 | 61.508 | 150.404 | 119.257 | 612.303 | 485.761 | 217.798 | 155.127 | 414.817 | 301.484 | 181.932 | 115.510 | 798.620 | 470.644 | 924.101 | 727.626 |
| ARIMA | 31.602 | 21.541 | 63.891 | 42.543 | 149.352 | 102.995 | 578.200 | 463.115 | 162.182 | 110.935 | 399.892 | 255.340 | 128.968 | 75.712 | 565.257 | 324.678 | 907.563 | 682.649 |
| SVR | 16.794 | 9.101 | 24.538 | 16.240 | 41.800 | 31.115 | 191.285 | 144.665 | 61.655 | 38.343 | 92.961 | 57.983 | 44.151 | 26.375 | 163.194 | 84.249 | 190.442 | 107.710 |
| LSTM | 16.308 | 8.745 | 23.179 | 15.505 | 42.230 | 31.365 | 182.018 | 138.290 | 58.828 | 36.208 | 91.774 | 58.485 | 44.352 | 25.942 | 163.130 | 83.863 | 183.029 | 104.565 |
| TN-LSTM | 15.845 | 8.571 | 22.712 | 15.103 | 40.444 | 30.031 | 178.427 | 135.182 | 57.572 | 35.083 | 90.882 | 55.896 | 42.562 | 25.306 | 96.081 | 52.893 | 164.350 | 97.005 |
| TSN-LSTM | 15.834 | 8.536 | 22.479 | 14.906 | 38.192 | 29.046 | 176.748 | 133.909 | 57.498 | 34.672 | 84.995 | 51.697 | 34.098 | 21.458 | 87.126 | 52.108 | 161.247 | 93.073 |

ARIMA. For instance, the RMSE of station 1 in cluster 3 is reduced from 162.182 to 61.655. Meanwhile, the performances of LSTM and SVR show little difference.

The data of workday and non-workday and historical passenger flow directly affects the passenger flow of the predicted station. The RMSE is remarkably improved when temporal characteristics with the calendar elements are incorporated into the model, i.e., from 798.620 to 163.130 in cluster 4. The MAE is also found to improve.

The accuracy is significantly enhanced when network characteristics are considered based on temporal dependencies. The passenger flows of some important stations exert a significant influence on the prediction. For instance, RMSE is from 163.130 to 96.081 for station 2 in cluster 4.

The accuracy is further improved when the spatial characteristics are incorporated into the model, as the RMSE and MAE for station 2 in cluster 1 are 22.479 and 14.906 respectively, which is better as compared to the TN-LSTM model. Moreover, TSN-LSTM represents better forecast performance, even for the huge passenger volume, than the other models. For instance, it can be seen from Fig. 10 that the maximum value of passage flow per 15 min at station 1 in cluster 4 is about 700, and it is about 3000 at station 2 in cluster 4. The RMSE and MAE for station 1 in cluster 4 in TSN-LSTM are 34.098 and 21.458, respectively, and in LSTM they are 44.352 and 25.942, respectively. But the RMSE and MAE for station 2 in cluster 4 in TSN-LSTM are 87.126 and 52.108, and in LSTM they are 163.130 and 83.863, respectively. Also, the MAE in ARIMA and SVR for station 2 in cluster 4 is 324.678 and 84.249, respectively. Regardless of the clusters, the TSN-LSTM has better performance than any other model.

Relevant studies on STPFP in urban rail transit are summarized and compared with the predictive performance of our models. However, the values of RMSE and MAE are significantly influenced by the values of passenger flow in the station. The values of passenger flow used in the different cases in the previous studies vary considerably. To ensure comparability, we set a base model in relevant studies, and the increased ratio of the proposed model compared to the base model is calculated as follows:

$$RA_{RM} = \frac{\boldsymbol{RMSE}_{\boldsymbol{study}} - \boldsymbol{RMSE}_{\boldsymbol{base}}}{\boldsymbol{RMSE}_{\boldsymbol{base}}}, \tag{23}$$

$$RA_{MA} = \frac{MAE_{\boldsymbol{study}} - MAE_{\boldsymbol{base}}}{MAE_{\boldsymbol{base}}}, \tag{24}$$

where, $RA_{RM}$ is increased ratio of RMSE in the proposed model compared to the base model, $\boldsymbol{RMSE}_{\boldsymbol{study}}$ is the RMSE of the proposed model, $\boldsymbol{RMSE}_{\boldsymbol{base}}$ is the RMSE of base model, $RA_{MA}$ is increased ratio of MAE in proposed model compared to the base model. $MAE_{\boldsymbol{study}}$ is the MAE of the proposed model, $MAE_{\boldsymbol{base}}$ is the MAE of the base model. $RA_{RM}$ and $RA_{MA}$ are used to represent increased ratios for predictive accuracy compared to the base model.

The ARIMA, a typical model, is considered the base model. The comparison results are shown in Table 4. The HSTDL [39], CB-LSTM [38], DeepPF [23], and AS2S_1E3D [11] are the deep learning-based methods. $RA_{RM}$ in these models achieved 62.0%, 18.7%, 69.2%, and 24.9%, respectively. The increased ratios of RMSE and MAE in TSN-LSTM are 71.3% and 73.2%, respectively, compared to ARIMA, which is the highest among compared models. The TSN-LSTM adds the influence of land use factors to the prediction compared to other models. The inputs for temporal and network characteristics are also improved. In addition, putting

**Table 4** Comparison with STPFP models in the literature

| | Method type | $RA_{RM}$ | $RA_{MA}$ |
|---|---|---|---|
| This study | TSN-LSTM | 71.3% | 73.2% |
| Zhang et al. [39] | HSTDL | 62.0% | 61.5% |
| Zhang et al. [38] | CB-LSTM | 18.7% | 17.7% |
| Liu et al. [23] | DeepPF | 69.2% | 57.1% |
| Hao et al. [11] | AS2S_1E3D | 24.9% | – |

similar characteristics of station traffic together for batch prediction helps to improve the accuracy.

## 5 Conclusions

This study developed an improved TSN-LSTM for urban rail transit. Based on LSTM, the proposed model takes into account spatial characteristics in addition to temporal characteristics and network characteristics between different stations, and this more comprehensive consideration improves the accuracy of the model. Passenger flow variation, OD matrix of stations, and land use are extracted and introduced to represent temporal, network, and spatial characteristics. In addition, by incorporating the K-means algorithm, the simultaneous prediction of passenger flow for multiple stations is achieved. Based on the actual data of Nanjing Metro, the proposed TSN-LSTM was compared with other typical prediction models, i.e., the ARIMA model, the SVR model, and the general LSTM. The results show that TSN-LSTM has lower root mean square error, mean absolute error, and higher R-squared compared with the other models. Furthermore, the model maintains a precise prediction at the station despite having a large number of passengers, which illustrates the effectiveness of the proposed model. Meanwhile, the model can not only realize the prediction of the passenger flow for the stations in the whole urban network in a short period of time, but also save a substantial amount of time and effort.

However, because of the limitations of the data acquisition, this consideration is not elaborated in the land use of the predicted station. For future work, it is worthwhile to investigate the number of residents in the community, the number of students, etc., which may probably further increase the prediction accuracy. Moreover, in addition to the consideration of multivariate data, the interpretability of different influences on the prediction results should be enhanced rather than viewing deep learning as a blind box. It is necessary to explore the relationship between different influencing factors and the results in future work.

**Funding** This research has been supported by the Jiangsu Rail Transit Industry Development Collaborative Innovation Base Open Fund (No. GCXC2104 and No. GCXC2103).

**Data availability** The datasets generated during the current study are not publicly available as the data also forms part of an ongoing study.

## Declarations

**Competing interests** The authors have no relevant financial or non-financial interests to disclose. Some data and models that support the findings of this study are available from the corresponding author upon reasonable request.

## References


1. Ambati LS, El-Gayar O (2021) Human activity recognition: A comparison of machine learning human activity recognition: A comparison of machine learning approaches. J Midwest Assoc Inf Syst 2021(1):4. https://doi.org/10.17705/3jmwa.000065
2. Ambati LS, El-Gayar O, Nawar N (2020) Influence of the digital divide and socio-economic factors on prevalence of diabetes. Issues Inf Syst 21(4):103–113
3. Ambati LS, El-Gayar O, Nawar N (2021) Design principles for multiple sclerosis mobile self-management applications: a patient-centric perspective. In: AMCIS 2021 Proceedings, (pp. 11)
4. Amiri SS, Mostafavi N, Lee ER, Hoque S (2020) Machine learning approaches for predicting household transportation energy use. City Environ Interact 7:100044
5. Bandara K, Bergmeir C, Hewamalage H (2020) LSTM-MSNet: Leveraging forecasts on sets of related time series with multiple seasonal patterns. IEEE Trans Neural Netw Learn Syst 32(4):1586–1599
6. Celebi ME, Kingravi HA, Vela PA (2013) A comparative study of efficient initialization methods for the k-means clustering algorithm. Expert Syst Appl 40:200–210
7. El-Gayar OF, Ambati LS, Nawar N (2020) Wearables, artificial intelligence, and the future of healthcare. In: AI and Big Data's Potential for Disruptive Innovation (pp. 104–129)
8. Feng XX, Ling XY, Zheng HF, Chen ZH, Xu YW (2019) Adaptive multi-kernel SVM with spatial-temporal correlation for short-term traffic flow prediction. IEEE Trans Intell Transp Syst 20(6):2001–2013. https://doi.org/10.1109/tits.2018.2854913
9. Govender P, Sivakumar V (2020) Application of k-means and hierarchical clustering techniques for analysis of air pollution: a review (1980–2019). Atmos Pollut Res 11(1):40–56
10. Habtemichael FG, Cetin M (2016) Short-term traffic flow rate forecasting based on identifying similar traffic patterns. Transp Res Part C Emerg Technol 66:6661–6678
11. Hao S, Lee DH, Zhao D (2019) Sequence to sequence learning with attention mechanism for short-term passenger flow prediction in large-scale metro system. Transp Res Part C Emerg Technol 107:107287–107300
12. Hassan BA, Rashid TA (2021) A multidisciplinary ensemble algorithm for clustering heterogeneous datasets. Neural Comput & Applic 3:1–24
13. He Y, Li L, Zhu X, Tsui KL (2022) Multi-graph convolutional-recurrent neural network (MGC-RNN) for short-term forecasting of transit passenger flow. IEEE Trans Intell Transp Syst 23:1–20. https://doi.org/10.1109/tits.2022.3150600
14. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
15. Hyndman RJ, Koehler AB (2006) Another look at measures of forecast accuracy. Int J Forecast 22(4):679–688
16. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. ACM Comput Surv (CSUR) 31(3):264–323
17. Jeong YS, Byon YJ, Castro-Neto MM, Easa SM (2013) Supervised weighting-online learning algorithm for short-term traffic flow prediction. IEEE Trans Intell Transp Syst 14(4):1700–1707. https://doi.org/10.1109/tits.2013.2267735
18. Jia HW et al (2020) ADST: forecasting metro flow using attention-based deep spatial-temporal networks with multi-task learning. Sensors 20(16):4574. https://doi.org/10.3390/s20164574
19. Jiao PP, Li RM, Sun T, Hou ZH, Ibrahim A (2016) Three revised Kalman filtering models for short-term rail transit passenger flow prediction. Math Probl Eng 2016:1–10. https://doi.org/10.1155/2016/9717582
20. Li HY, Wang YT, Xu XY, Qin LQ, Zhang HY (2019) Short-term passenger flow prediction under passenger flow control using a dynamic radial basis function network. Appl Soft Comput 83:105620. https://doi.org/10.1016/j.asoc.2019.105620
21. Li Y, Zhu Z, Kong D, Han H, Zhao Y (2019) EA-LSTM: Evolutionary attention-based LSTM for time series prediction. Knowl-Based Syst 181:104785.1–104785.8

22. Li W, Guan H, Han Y, Zhu H, Wang A (2022) Short-term holiday travel demand prediction for urban tour transportation: a combined model based on STC-LSTM deep learning approach. KSCE J Civ Eng 26(9): 4086–4102. https://doi.org/10.1007/s12205-022-2051-8
23. Liu Y, Liu Z, Jia R (2019) DeepPF: A deep learning based architecture for metro passenger flow prediction. Transp Res Part C Emerg Technol 101:18–34
24. Ma X, Zhang J, Du B, Ding C, Sun L (2018) Parallel architecture of convolutional bi-directional LSTM neural networks for network-wide metro ridership prediction. IEEE Trans Intell Transp Syst 20(6):2278–2288
25. Nanda SJ, Panda G (2014) A survey on nature inspired metaheuristic algorithms for partitional clustering. Swarm Evol Comput 16:161–118. https://doi.org/10.1016/j.swevo.2013.11.003
26. Ni M, He Q, Gao J (2017) Forecasting the Subway passenger flow under event occurrences with social media. IEEE Trans Intell Transp Syst:1–10
27. Polson NG, Sokolov VO (2017) Deep learning for short-term traffic flow prediction. Transp Res Part C Emerg Technol 79:1–17
28. Roos J, Gavin G, Bonnevay S (2017) A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data. In: Coppola P (ed) Emerging technologies and models for transport and mobility, vol 26. Transportation Research Procedia, pp 53–61
29. Shahid F, Zameer A, Muneeb M (2021) A novel genetic LSTM model for wind power forecast. Energy 1: 120069
30. Sohn K, Shim H (2010) Factors generating boardings at metro stations in the Seoul metropolitan area. Cities 27(5):358–368. https://doi.org/10.1016/j.cities.2010.05.001
31. Sun YX, Leng B, Guan W (2015) A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. Neurocomputing 166:166109–166121. https://doi.org/10.1016/j.neucom.2015.03.085
32. Sung H, Oh JT (2011) Transit-oriented development in a high-density city: identifying its association with transit ridership in Seoul, Korea. Cities 28(1):70–82. https://doi.org/10.1016/j.cities.2010.09.004
33. Thiruchelvam L, Dass SC, Asirvadam VS, Daud H, Gill BS (2021) Determine neighboring region spatial effect on dengue cases using ensemble ARIMA models. Sci Rep 11(1):5873
34. Wei Y, Chen MC (2012) Forecasting the short-term metro passenger flow with empirical mode decomposition and neural networks. Transp Res Part C-Emerg Technol 21(1):148–162. https://doi.org/10.1016/j.trc.2011.06.009
35. Williams BM, Hoel LA (2003) Modeling and forecasting vehicular traffic flow as a seasonal ARIMA process: theoretical basis and empirical results. J Transp Eng 129(6):664–672. https://doi.org/10.1061/(asce)0733-947x(2003)129:6(664)
36. Wu J, Liu M, Sun H, Li T, Wang D (2014) Equity-based timetable synchronization optimization in urban subway network. Transp Res Part C Emerg Technol 51:1–18
37. Yang X, Xue QC, Ding ML, Wu JJ, Gao ZY (2021) Short-term prediction of passenger volume for urban rail systems: a deep learning approach based on smart-card data. Int J Prod Econ 231:107920. https://doi.org/10.1016/j.ijpe.2020.107920
38. Zhang J, Chen F, Shen Q (2019) Cluster-based LSTM network for short-term passenger flow forecasting in urban rail transit. IEEE Access 7:7147653–7147671
39. Zhang H, He J, Bao J, Hong Q, Shi X (2020) A hybrid spatiotemporal deep learning model for short-term metro passenger flow prediction. J Adv Transp 2020:Article ID 4656435

## Affiliations

**Ningning Dong[1] · Tiezhu Li[1] · Tianhao Liu[1] · Ran Tu[1] · Fei Lin[2] · Hui Liu[3] · Yiyong Bo[3]**

Ningning Dong
nningdong@163.com

Tianhao Liu
t.liu.tub@gmail.com

Ran Tu
turancoolgal@seu.edu.cn

Fei Lin
linfeicallme@163.com

Hui Liu
1357746565@qq.com

Yiyong Bo
41981070@qq.com

[1] School of Transportation, Southeast University, Nanjing 211189, China

[2] Nanjing Metro Group Co., Ltd., Nanjing 211806, China

[3] Nanjing Vocational Institute of Railway Technology, Nanjing 210031, China