Research on Short-time Inbound Passenger Flow Prediction of Urban Railways Considering Station Flow Heterogeneity

Jia He, Changfeng Zhu, Runtian He, Yunqi Fu, Jie Wang

Abstract—**Short-term passenger flow prediction for urban rail transit is imperative for effective real-time scheduling, efficient resource allocation, and prompt emergency response. However, conventional models face challenges in accurately capturing passenger flow features, which are influenced by numerous sources. Consequently, to address the limitations of traditional models, this paper investigates the improved dual-channel convolutional neural network (DC-CNN) and long-short-term memory network (LSTM) to construct a combined model to deal with complex spatiotemporal data, capturing non-linear features and sequence dependencies. Secondly, the impact of data quality, station heterogeneity, and model parameter training strategy on prediction accuracy was considered. A multi-dimensional index system was constructed based on the ontological attribute characteristics of subway stations and location environment features. The research sample was classified as a typical station using the k-means clustering algorithm. Furthermore, the model training phase incorporates a hybrid tuning strategy that integrates manual parameter optimization and automated hyperparameter search algorithms. An early-stop mechanism is also integrated to balance the model performance and training efficiency. To conclude the research, an example analysis is carried out by combining Hangzhou Metro AFC data. Based on the station POI data and historical passenger flow data, the stations are grouped into four categories (Cluster 0, 1, 2, 3). The MAPE values of the passenger flow prediction results are 14.02%, 18.12%, 13.57%, and 13.82%, respectively. Through experimental validation, it is determined that the prediction accuracy of the DC-CNN-BiLSTM prediction model optimized by considering site heterogeneity and using an improved training strategy is enhanced in all types of sites. Furthermore, the average absolute percentage error can be reduced by up to 1.62% compared with that of the original data without site segmentation. Additionally, the sensitivity of different hyperparameters to the prediction accuracy of the model demonstrates significant heterogeneity.**

The present study has the potential to contribute to the theoretical framework for predicting short-time passenger flow in urban rail transit.

Index Terms—**urban railway transport; k-means clustering; deep learning; predicting passenger flows**

I. Introduction

In the context of accelerating urbanization, rail transit has emerged as a pivotal component of the urban transportation infrastructure, assuming a predominant role in facilitating passenger mobility. The short-time passenger flow of urban rail transit is influenced by a multitude of factors, including time, space, holidays, and others, resulting in significant volatility and spatial and temporal complexity. Given the frequent fluctuations in passenger flow within urban rail transit systems, it is imperative to accurately capture the dynamic changes in passenger flow over brief periods, which is essential for supporting traffic scheduling and emergency response. Traditional time series forecasting methods are inadequate for accurately capturing the spatial and temporal dynamics of passenger flow. Consequently, the central objective of this study is to explore effective methodologies for integrating temporal and spatial characteristics within the rail transit system through the application of deep learning techniques. This endeavor aims to formulate a passenger flow prediction model that can comprehensively address the multifaceted influences of various sources, thereby enhancing the accuracy and generalizability of predictions.

An accurate analysis of passenger flow features of urban rail transit is fundamental to the construction of accurate passenger flow prediction models. Numerous results have emerged from related research. In a study by L. X. Si et al. [1], the authors examined the spatial distribution of the occupational and residential population in Hangzhou, as well as the spatial and temporal characteristics of subway commuting data. This analysis was based on the degree of occupational and residential spatial matching, as determined by cell phone signaling data from China Mobile. The findings led to preliminary conclusions regarding the reasonable degree of spatial and temporal features of occupational and residential spatial matching, as well as the spatial and temporal features of subway commuting in Hangzhou. In a related study, Y. S. Jiang et al. [2] proposed an acceptable site classification method based on the k-means++ algorithm

Manuscript received April 29, 2025; revised July 10, 2025.

This work was supported in part by the National Natural Science Foundation of China (No.72161024) and "Double-First Class" Major Research Programs, Educational Department of Gansu Province (No.GSSYLXM-04).

Jia He is a postgraduate student at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China(e-mail: 1983620746@qq.com).

Changfeng Zhu is a professor at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China(Corresponding author, phone: +86 189 1989 1566, e-mail: cfzhu003@163.com).

Runtian He is a doctoral candidate at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China(e-mail: heruntian8@gmail.com).

Yunqi Fu is a doctoral candidate at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China(e-mail: 13240002@stu.lzjtu.edu.cn).

Jie Wang is a doctoral candidate at School of Traffic and Transportation, Lanzhou Jiaotong University, Lanzhou 730070, China(e-mail: 1009696615@qq.com).

clustering site public features to cluster the passenger flow of each site. They established a fitting equation between the results of the passenger flow clustering and the multi-dimensional parameters of land use features and calculated to obtain the public features of the traffic flow of the sites in five broad categories. Z. J. Zhu et al. [3] analyzed the passenger flow distribution characteristics of rail transit origins and destinations (OD) by using the drift power law distribution model. They then used the community discovery method based on the Louvain algorithm to divide the rail transit network into different community structures. This analysis revealed the spatial correlation between different stations from the perspective of OD ridership.

The theoretical methods currently employed in rail passenger flow prediction research can be broadly categorized into three distinct classes.

Concerning prediction methods based on mathematical and statistical models, this approach is widely employed as a conventional classical passenger flow prediction method due to its robust theoretical foundation. J. Xiong et al. [4] proposed a method based on the Kalman filtering principle and gray correlation analysis to realize the short-term prediction of passenger flow and carried out the validity test of the prediction method. C. Q. Ma et al. [5] have proposed a differential integration moving average autoregressive model to predict short-term passenger flow for the entire network under various time granularities. G. Y. Zhang et al. [6] developed a research approach based on an enhanced ARIMA model for the prediction of short-duration passenger flow in urban rail transit, which can predict the amount of delay generated by the dependent variable, the lagged value, and the current value of the random error generated.

The advent of artificial intelligence has precipitated the advancement of intelligent algorithms for conventional machine learning predictive models. These models have been further developed to capture non-linear relationships and interaction effects between features in data with greater flexibility than mathematical-statistical models. With respect to predictive methods based on traditional machine learning, L. H. Li et al. [7] developed a short-term passenger flow prediction model based on the random forest regression algorithm, incorporating factors such as departure date and operation time. They evaluated the importance of passenger flow influencing factors by combining the OOB residual mean square. S. S. Liu et al. [8] proposed a least squares support vector machine (LSSVM) based on the optimization parameter of the Improved Particle Swarm Optimization Algorithm (IPSO) to address the passenger flow fluctuations during complex holidays and achieve sparsification of model solutions. J. Jin et al. [9] analyzed the spatio-temporal complexity characteristics and employed a backpropagation (BP) neural network to predict the zone passenger flow. This prediction was based on the inbound and outbound flows of the first three consecutive periods of the urban rail transit and the zone passenger flow data of the next period.

In the context of the proliferation of traffic data and the advent of advanced deep learning algorithms, time series prediction models have emerged as a prevalent approach for extracting temporal correlations in data to address traffic prediction challenges. In the domain of passenger flow forecasting, the utilization of deep learning methodologies for the extraction of features from passenger flow time series has been a subject of considerable research interest. Q. Ouyang et al. [10] have developed a historical LSTM model for passenger flow prediction at bus stops. This model is founded on the feature extraction of an Xgboost model, the encoding of historical data, the encoding of real-time data, and the decoding of a multilayer neural network. The LSTM model for real-time data has been employed for passenger flow prediction at bus stops. J. Li et al. [11] developed a high-speed railroad passenger flow prediction model based on LSTM, and the results indicated that the number of hidden units, neurons, and input step size have a significant impact on the accuracy of passenger flow prediction. T.D. Sajanraj et al. [12] proposed a methodology that employs a long short-term memory (LSTM) algorithm to circumvent the limitations of a station-specific model in conducting a global model search for all stations. In their study, C. Y. Lin et al. [13] examined the spatial correlation of subway stations within a subway line and the temporal correlation of time series in passenger flow prediction. They developed a long-short-term memory (LSTM) model applicable to the prediction of a single subway station. This model predicted subway passenger flow based on the land use around the subway station.

The majority of extant studies exclusively consider time series data, a methodological approach that falls short of meeting the demand for accurate prediction of passenger flow. These studies also tend to overlook the existence of rich spatial features in traffic data. Consequently, some scholars have introduced convolutional networks to assist the model in extracting spatial information. In their seminal work, J. L. Zhang et al. [14] proposed Conv-GCN, a pioneering deep learning architecture that fuses graph convolutional networks (GCN) and three-dimensional convolutional neural networks (3D CNN). This innovative framework not only handles three inflow and outflow modes (near-term, daily, and weekly) but also captures the spatio-temporal correlation and topological information of the entire network, which is then deeply fused.

In light of the growing body of research examining the factors influencing passenger flow characteristics, scholars have increasingly combined the time series model and a convolutional network to integrate the temporal and spatial dimensions of passenger flow in passenger flow prediction. This integration has been shown to enhance prediction accuracy. In the context of prediction based on combinatorial models, B. X. Cao et al. [15] have proposed a CNN-LSTM model that integrates built environment indicators with historical passenger flow and key built environment indicators as input variables to facilitate precise short-term passenger flow prediction. X. L. Ma et al. [16] have proposed a convolutional neural network (CNN)-based abstract traffic feature extraction and a network-wide traffic speed prediction method that learns the traffic as an image and can predict the traffic speed of large-scale and network-wide traffic with high accuracy. J. Liu et al. [17] have constructed a subway passenger flow prediction model based on ISSA-CNN-LSTM by using ISSA to optimize the hyperparameters of the CNN-LSTM model based on the analysis of the correlation between different types of subway stations passenger flow and weather factors.

In summary, some extant studies have given insufficient

consideration to the variability of passenger flow features and spatial distribution among different subway stations, making it difficult for the models to capture local flow patterns at the station level in a fine-grained manner. Furthermore, conventional time-series prediction methodologies predominantly emphasize the statistical characteristics of the time series itself, which poses significant challenges in fully integrating the intricate spatial dependencies. These methodologies frequently necessitate extensive data preprocessing and model training, consequently leading to a substantial increase in the time complexity of the model when dealing with large-scale site data. In the context of model construction, a single convolutional neural network can capture spatially localized features; however, it is deficient in its ability to portray in-depth temporal dynamics. Conversely, traditional LSTM models and their variants encounter limitations in their capacity to capture spatial information, which hinders their efficacy in dealing with complex spatio-temporal dependencies. Furthermore, the efficacy of several prediction models is constrained in the context of non-linear variation, extended time dependence, and fine-grained time prediction. This limitation impedes the ability to accurately predict short-term passenger flow, hindering adequate support for metro scheduling and emergency management protocols.

In order to address these issues, a deep learning prediction model has been constructed based on the combination of CNN and Bi-LSTM. This model mines the passenger flow characteristics of different types of stations and captures the changing law of short-term passenger flow of rail transportation more accurately through the fusion of spatial and temporal features. The purpose of this is to improve the accuracy and applicability of the prediction model. Grouping sites based on similarity through clustering effectively reduces the heterogeneity of the data and improves the ability to recognize local patterns, which in turn compensates for the lack of spatial heterogeneity mining ability of traditional models. Combining the advantages of CNN for spatial feature extraction and BiLSTM for in-depth modeling of bidirectional temporal dependencies significantly improves the spatial and temporal feature fusion capability for short-term passenger flow prediction, overcoming the limitations of traditional prediction models in modeling spatial and temporal dependency structures. The integrated prediction model has been demonstrated to be more effective in capturing the fluctuating characteristics of short-duration passenger flows.

The model integrates the structural framework of the rail transit station network with the temporal progression of passenger flow, thereby enabling the provision of relatively precise passenger flow prediction outcomes within a concise timeframe under complex spatial and temporal contexts. This integration signifies the model's substantial applicability and practical relevance.

II. Description of the Issue

The research for short-term inbound and outbound passenger flow prediction is essentially a passenger flow time-series prediction problem. That is to say, based on historical passenger flow data $\{X(t)|\ t \in T_{\text{past}}\}$, the passenger flow at any station s_i at a future point in time $t+\Delta t$ (Δt/min) is predicted by analyzing the multi-source factors affecting the passenger flow. These factors may include, for example, the spatial distribution characteristics of the station and weather conditions at the regional scale.

The urban rail transit network is modeled as a graph $G=(S,E)$, where the set of stations of the rail transit network $S=\{s_1,s_2,\ldots,s_N\}$, the number of stations is N; the set of edges between stations $E \subseteq S \otimes S$ is described as the spatial connectivity between stations.

Furthermore, a correlation matrix A is constructed, taking into account physical distance, historical passenger flow correlation, and spatial correlation. The application of artificial intelligence (A_{ij}) facilitates the determination of connection strength and similarity between s_i and s_j.

In light of the aforementioned prediction of future passenger flow, the construction of a deep learning prediction model f has been undertaken. The subsequent analysis will seek to ascertain the model's optimal prediction performance.

$$\left\{\hat{x}_{t+1},\hat{x}_{t+2},...,\hat{x}_{t+T}\right\}=f^{(\theta)}(X,G(A),W_{EA}) \tag{1}$$

Where, $\hat{x}_{t+k}$ is the predicted value of the model at time $t+\Delta t$ ($\Delta t \in [1,T]$); W_{EA} denotes data related to external factors; and θ is a model parameter.

Due to the temporal and spatial distribution characteristics of the passenger flow into the station, the model outputs the predicted value $\hat{\boldsymbol{X}}_{t+1:t+T}$, which is a two-dimensional matrix that responds to the spatial and temporal characteristics of the passenger flow as follows

$$\hat{\boldsymbol{X}}_{t+1:t+T}=\begin{bmatrix}\hat{x}_{t+1}^1 & \hat{x}_{t+1}^2 & \cdots & \hat{x}_{t+1}^N\\ \hat{x}_{t+2}^1 & \hat{x}_{t+2}^2 & \cdots & \hat{x}_{t+2}^N\\ \cdots & \vdots & \ddots & \vdots\\ \hat{x}_{t+T}^1 & \hat{x}_{t+T}^2 & \cdots & \hat{x}_{t+T}^N\end{bmatrix}\in\mathbb{R}^{T\times N} \tag{2}$$

To ensure that the results of the output of the model f are as accurate as possible, the deviation between $\left\{\hat{x}_{t+1},\hat{x}_{t+2},...,\hat{x}_{t+T}\right\}$ and the actual value of the future passenger flow $\left\{x'_{t+1},x'_{t+2},...,x'_{t+T}\right\}$ should be minimized under some error metric. Construct an appropriate loss function $\ell(\cdot,\cdot)$ to minimize the loss $\mathcal{L}$ as follows

$$\mathcal{L}=\sum_{t\in T_{\text{future}}}\ell(\hat{Y},Y) \tag{3}$$

To further represent the model prediction accuracy, based on the literature [25], the model loss function is constructed using the mean square error (MSE) as follows

$$\mathcal{L}(Y,\hat{Y})_{\text{MSE}}==\frac{1}{N\times T'}\sum_{i=1}^{N}\sum_{t=1}^{T'}(Y_{i,t}-\hat{Y}_{i,t})^2 \tag{4}$$

To summarize, the problem of predicting passenger flow is essentially solved by training a model f with historical data $\{X(t)|\ t \in T_{\text{past}}\}$ and appropriate exogenous variables $\{W_i(t)|\ t \in T_{\text{past}}\}$ and applying it to make predictions at future points in time $t \in T_{\text{future}}$.

III. RESEARCH METHODS

A. Site Segmentation Based On K-means

A comparative analysis of passenger flow features across distinct metro stations is conducted, with particular attention to the impact of POIs on these features. The utilization of clustering and synthesis techniques to identify commonalities in passenger flow patterns enables the grouping of stations with similar features into distinct categories. This approach facilitates the classification and coding of stations with significant variations in passenger flow features, thereby enhancing the accuracy of prediction models.

Given a collection of light rail stations, where $s_i \in \mathbb{R}^{N\times N}$ denotes the i^{th} station data point. To ensure the cohesion of clusters, in each iteration s_i is assigned to the cluster center $Q_j = \left\{ s_i : \left\| s_i - \mu_j \right\| \le \left\| s_i - \mu_l \right\|, \forall l = 1,\ldots,K \right\}$ that is closest to it. The cluster center is updated after each allocation $\mu_j = 1/\left|Q_j\right| \sum_{s_i \in Q_j} s_i$, which is the average of all data points in the current cluster. The instant algorithm converges when $\left\| \mu_j^{(t+1)} - \mu_j^{(t)} \right\| < \varepsilon \quad \forall j = 1,2,\ldots,K$.

To circumvent an excessive degree of randomization in the selection of cluster centers, the k-means++ algorithm is employed to enhance the initialization method. Specifically, an arbitrary point is initially selected as the first cluster center, and the subsequent cluster centers are then selected based on the probability distribution, which is determined by the distance $D(s)$ between the data points and the selected cluster centers. The initialized probability distribution operation $P(s)$ is as follows

$$P(s) = \frac{D(s)^2}{\sum_{s' \in N} D(s')^2} \tag{5}$$

The principle underlying this methodology is that the maximum contour coefficient indicates the optimal clustering effect. The contour coefficient operation $S(i)$ for any point i is as follows

$$S(i) = \frac{b(i) - a(i)}{max\{a(i), b(i)\}} \tag{6}$$

Where, $a(i)$ is the distance of the vector i to other points in the same cluster; $b(i)$ is the mean value of the distance of the vector i to the points contained in the nearest neighboring cluster. The mean value of the contour coefficient of all data points is taken as an assessment of the overall clustering effect.

To achieve the minimum value of the optimized objective function, the data is divided into K clusters $Q_1, Q_2, \ldots, Q_K$ as follows

$$\min_{\mu_1,\mu_2,\ldots,\mu_K} \sum_{j=1}^{K} \sum_{x \in C_j} \left\| x - \mu_j \right\|^2 \tag{7}$$

Where, μ_j is the center of the cluster Q_j ; $\| x - \mu_j \|$ represents the Euclidean distance between the data points and the cluster center. By iteratively updating the cluster assignments and cluster centers, the algorithm continuously approaches this optimal solution, which is the optimal number of clusters.

B. Combinatorial Predictive Modeling and Feature Engineering Construction

The integration of convolutional neural networks (CNNs) and bidirectional long short-term memory (BiLSTMs) has been demonstrated to be a highly effective approach for addressing the challenges posed by spatio-temporal features and long-term temporal dependencies, respectively.

This integration has enabled the construction of a combined prediction model that exhibits both efficiency and accuracy in predicting passenger flow. The structure of the predictive portfolio model is shown in Fig. 1.

To more systematically extract the value of the data and construct a high-quality feature set to significantly improve the accuracy and reliability of the passenger flow prediction model, the following steps should be taken to construct a feature engineering that can be better adapted to the short-time passenger flow prediction model:

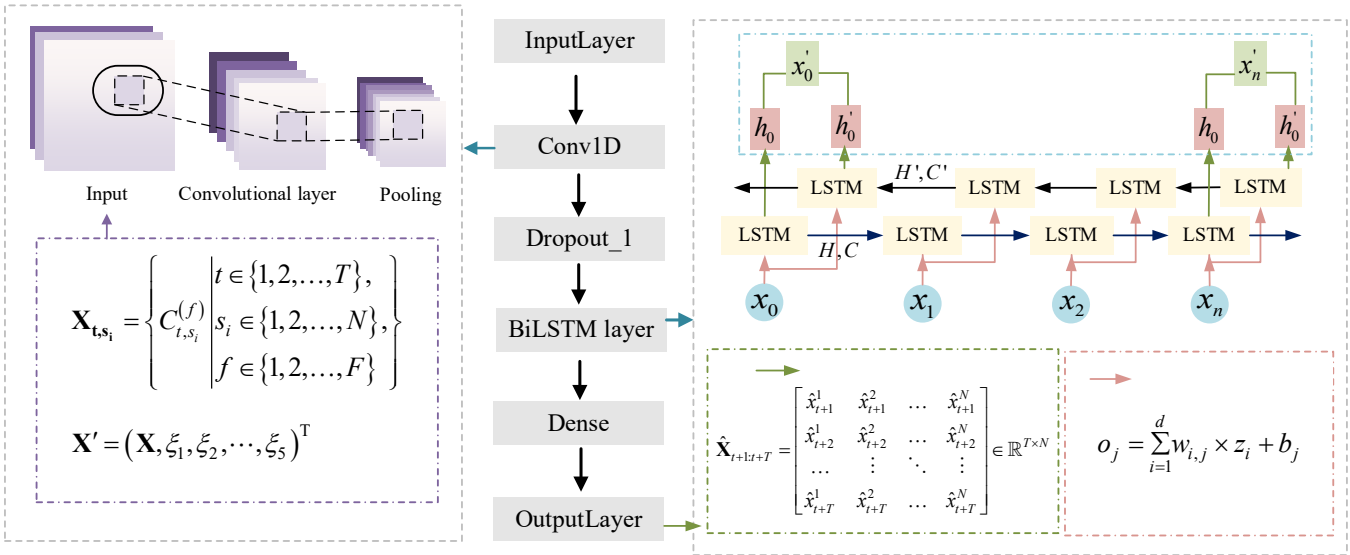


Fig. 1. Structure of the predictive portfolio model

- The creation of a feature engineering process to identify the key information dimensions in passenger flow forecasting. The identification and extraction of the base period hourly features. The hourly features of the base period are identified and extracted, and the "hourly" features are derived from the time index.

- The feature construction is engineered by applying the derivation technique to derive six time step rolling mean features based on normalized historical passenger flow to capture the short-term smoothing trend. Derived differential features with six time steps characterize the short-term rate of change and volatility of passenger flow.

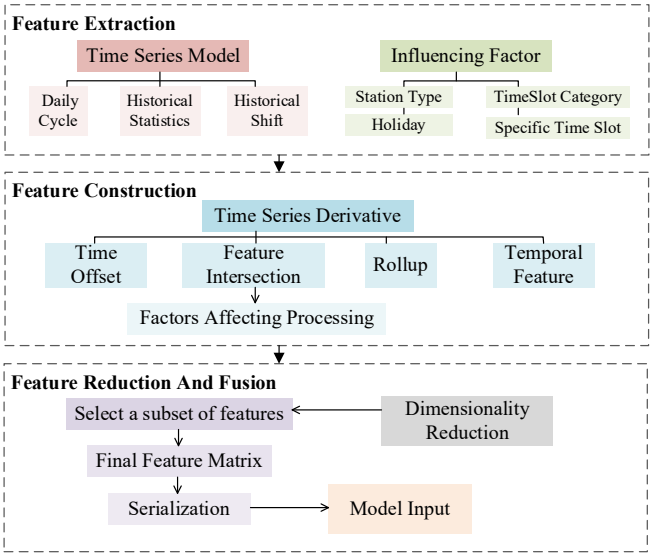


Fig. 2. Theoretical system for constructing passenger flow features engineering

Encoding multi-dimensional features, including different time slices of the time axis, multiple features, and their correlations. Suppose the input feature matrix $\boldsymbol{X}$ has three dimensions $N \times T \times F$, the number of features at each time step is F. The model input is considered a three-dimensional tensor $\boldsymbol{X} \in \mathbb{R}^{T \times N \times F}$. Coding the input matrix as a two-dimensional matrix $\boldsymbol{X}^{'} \in \mathbb{R}^{N \times F^{'}}$, where feature fusion $F^{'}$ obtains the final feature dimension.

Regarding the time series coding, a time granularity of t (min) is selected on the time axis, and the forecasting target is designated as the future ($\Delta t = 10$) short-term passenger flow. The features of the time dimension can be expressed as $\boldsymbol{T} = \{t_1, t_2, ..., t_T\}$, where t_i is a time slice indicating the corresponding time feature $t_i \in \mathbb{R}$ (e.g., hour, date, holiday identifier, etc.).

Calculate average passenger flow $\overline{x}_{r,s}^{m}$ based on time-series data, taking into account departure station types ξ_r and arrival station types ξ_s. The category of the fusion features of passenger flow space and time obtained by cluster analysis is denoted as $C_{r,s}^{m}$. The set of passenger flow data corresponding to each site in the spatial dimension is represented by $\boldsymbol{N} = \{x_{s_1}, x_{s_2}, \ldots, x_{s_{N-1}}, x_{s_N}\}$. The features of each site include, but are not limited to, the actual number of passengers at each site at different time steps, the date, whether it is a holiday, special events, etc. Let the eigenvector of the i^{th} site be, for the moment t.

The input space-time feature matrix $\boldsymbol{X}_{t,s_i}$ can be expressed as follows

$$\boldsymbol{X}_{t,s_i} = \left\{ C_{t,s_i}^{(f)} \middle| \begin{array}{l} t \in \{1,2,\ldots,T\}, \\ s_i \in \{1,2,\ldots,N\}, \\ f \in \{1,2,\ldots,F\} \end{array} \right\} \tag{8}$$

Where, $C_{t,s_i}^{(f)}$ denotes the f^{th} feature at t, s_i.

The term $\boldsymbol{X}'$ is defined as the input matrix of the combined forecasting model, which is the model used to predict future passenger flows as follows

$$\boldsymbol{X}' = \left(\boldsymbol{X}, \xi_1, \xi_2, \cdots, \xi_5\right)^{\boldsymbol{T}} \tag{9}$$

Where, $\boldsymbol{X}$ is the passenger flow matrix for each time segment; $\xi_1, \xi_2, \cdots, \xi_5$ is each influencing factor. The urban rail passenger flow data are aggregated at regular time intervals, resulting in the formation of two-dimensional data comprising time steps and the number of features. Relevant features are shown in Table I.

TABLE I
Factors Affecting Rail Ridership

variant	factor	data range
ξ_1	entry/exit times	06:00—23:00
ξ_2	days of the week	1,2,3,4,5,6,7
ξ_3	holiday	yes-1; no-0
ξ_4	high/peak conditions	morning peak-0; morning flat peak-1; evening peak-2; evening flat peak-3
ξ_5	type of site	depending on the results of site cluster 0,1,…

C. *Spatial feature extraction based on DC-CNN*

The spatial features of subway passenger flow include not only the geographical relationship between each station but also the spatial distribution pattern of passenger flow in different periods and different scenes. To accurately extract the spatial features of multi-scene passenger flow, a parallel processing Dual-Channel Convolutional Neural Network (DC-CNN) has been proposed. This network is designed to extract and integrate different types of spatial information, thereby obtaining a richer and multilevel representation of spatial features.

As can be seen in Fig. 3, the urban rail transit system is divided into grid cells based on the relative positions of the stations on the plan, and the spatial topological relationships of the stations are preserved. The pixel value of each cell is defined as the passenger flow of the corresponding station. This data is then constructed into a two-dimensional image, which is referred to as a spatio-temporal map of the passenger flow of urban rail transit stations.

The two-dimensional image under consideration is to be represented as a vector $X \in \mathbb{R}^{H \times W \times C}$, where H is the height of the image, W is the width of the image, and C is the number of channels of the image. The image is to be used in the construction of a convolutional neural network with multiple channel inputs $H \times W \times C^0$ and outputs $H \times W \times C^1$.

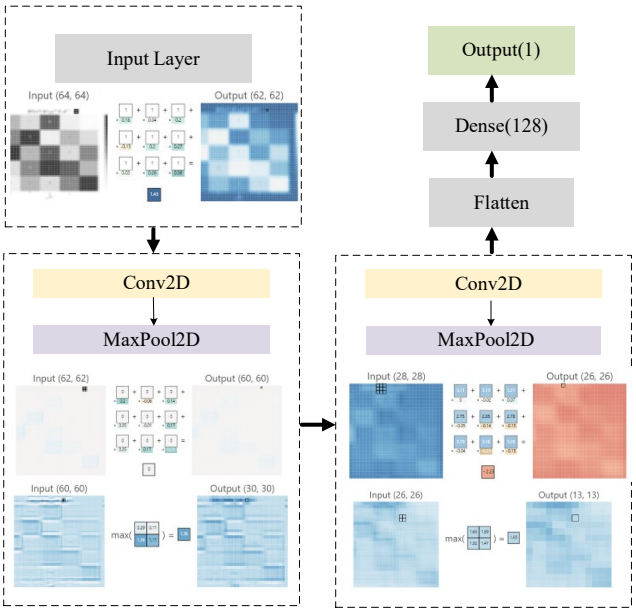


Fig. 3. Diagram of CNN operation

A convolutional layer is constructed to extract local features. The output value of the convolutional result $Y_{(i,j,k)}$ in the position (i,j) and channel c is operated as indicated in the following equation.

$$Y_{(i,j,c)} = \sum_{m=0}^{M-1}\sum_{n=0}^{N-1}\sum_{c=0}^{C-1} W_{m,n,c,k} \cdot X_{(i+m,j+n,c)} + b_k \tag{10}$$

Where, $X_{(i,j,c)}$ is the pixel value of the channel c at the position (i,j) in the input image; $W_{m,n,c,k}$ is the weight of the convolution kernel of size $M \times N$, where m and n is the index used to traverse the height and width of the convolution kernel; k is the output channel index; b_k is the bias term of the convolution kernel.

In light of the necessity to incorporate non-linear features, while concurrently seeking to circumvent the pitfalls of overfitting, the implementation of the ReLU activation function $f(x)_{\text{ReLU}}$ proves instrumental. This function is deployed to process the linear pre-activation feature maps derived from the convolution kernel, ensuring a systematic and methodical approach to model development.

$$f(x)_{\text{ReLU}} = \begin{cases} x, & x > 0 \\ 0, & otherwise \end{cases} \tag{11}$$

Subsequently, the value $Z_{(i,j,k)}$ of the position in the position of (i,j,k) in feature map can be obtained following activation function processing.

$$Z_{(i,j,k)} = f_{\text{Relu}}\left(Y_{(i,j,k)}\right) = \max\left\{0, Y_{(i,j,k)}\right\} \tag{12}$$

The maximum pooling method is employed to construct the pooling layer, which serves to reduce the dimensionality of the feature map while retaining the most salient features, thus reducing the computational burden. Pooling outputs the downscaled values $Z'_{\text{pool}(i,j,k)}$ as

$$Z'_{\text{pool}(i,j,k)} = \max_{m=0,A-1}\left\{\max_{n=0,B-1}\left\{i+m, j+n, k\right\}\right\} \tag{13}$$

Where: $A \times B$ is the size of the pooling window; i, j is the index of the output feature map in height and width directions.

Subsequently, the magnitude $H_{\text{pool}} \times W_{\text{pool}}$ of the feature map post-pooling is determined as follows

$$H_{\text{pool}} = \frac{H + 2p - M + s(s-a+1)}{s^2} \tag{14}$$

$$W_{\text{pool}} = \frac{W + 2p - N + s(s-b+1)}{s^2} \tag{15}$$

Where: s denotes the step size in the convolution operation; p denotes padding. The padding technique is introduced in response to the loss of data edge information and the limitation of CNN applications due to the reduction of output data size that accompanies the increase in the number of convolutional layers.

The feature map $\boldsymbol{Z}_{\text{pool}}$, which is obtained after multiple convolutions and pooling, is spread into a vector $z \in \mathbb{R}^d$. This vector is then linearly transformed to output a vector o_j by a fully connected layer as follows

$$o_j = \sum_{i=1}^{d} w_{i,j} \times z_i + b_j \tag{16}$$

Where, $w_{i,j}$ is the weight of the fully connected layer; b_j is the bias term; $d = H_{\text{pool}} \times W_{\text{pool}} \times C_{\text{pool}}$.

D. Temporal feature extraction based on Bi-LSTM

In essence, the majority of passenger trips are typically oriented around commuting between residential and occupational locations, exhibiting a high degree of regularity. Consequently, the temporal distribution of subway passenger flow characteristically exhibits unevenness. In consideration of the periodicity and cyclicity of time, the prediction results are contingent not only on historical data but also on future trends. To address this, the Bi-LSTM model is employed to capture the time-dependent characteristics in both directions. Bi-LSTM is an extended version of LSTM, consisting of forward and inverse LSTM units. Forward and reverse LSTMs are computed from the starting and ending points of the time series, respectively, and the outputs of both are combined to measure changes in urban rail passenger flow. The procedure is as follows

$$\overrightarrow{h_t} = \text{LSTM}_{\text{forward}}(x_t, \overrightarrow{h_{t-1}}) \tag{17}$$

$$\overleftarrow{h_t} = \text{LSTM}_{\text{backward}}(x_t, \overleftarrow{h_{t+1}}) \tag{18}$$

$$h_t = [\overrightarrow{h_t}, \overleftarrow{h_t}] \tag{19}$$

The final hidden state h_T of the Bi-LSTM is connected to a connected layer. The output of this layer is the predicted value of passenger traffic at a future moment t. The operation of this layer is as follows

$$\hat{y} = W_h \cdot h_T + b_h \tag{20}$$

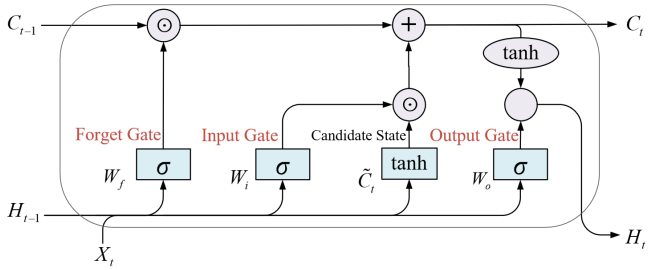


Fig. 4. Internal structure of the LSTM "neuron" at time t

The interior of each LSTM cell is illustrated in Fig. 4. The operations of the LSTM forget gate, input gate, and output gate are as follows

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{21}$$

Where, f_t is the activation value of the forgetting gate; W_f is the weight matrix; b_f is the bias vector; σ is the sigmoid activation function, $\sigma(t) = \frac{1}{1+e^{-t}}$.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{22}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{23}$$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{24}$$

Where, i_t is the activation value of the input gate; W_i, W_C is the weight matrix; b_i, b_C is the bias vector; $\tilde{C}_t$ is the candidate memory cell state; x_t represents the vector of inputs; C_t is the updated cell state; the activation function $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{25}$$

$$h_t = o_t \cdot \tanh(C_t) \tag{26}$$

Where, o_t is the activation value of the output gate; W_o is the weight matrix; b_o is the bias vector; h_t represents the hidden state vector.

The spatial features extracted in the CNN module are then fed into the LSTM part, which processes the dependencies in the time dimension and captures the time series features of the passenger flow. At each temporal interval t, the hidden state h_t of the LSTM model is updated by the following inputs: (1) the spatial features ; (2) the hidden state h_{t-1} from the previous moment. The update process is defined as follows

$$h_t = \mathrm{LSTM}(X_{t,:}, h_{t-1}) \tag{27}$$

Where, $X_{t,:}$ denotes the convolutional output features for all sites at the time step moment t.

E. Preprocessing of data

E.1. Data cleaning

To ensure the quality of data, missing values, outliers, and duplicates are treated in this paper. The interpolation method is employed to address the presence of missing values. If the missing values cannot be reasonably filled, the obviously unreasonable data points are deleted.

To identify outliers, the box-and-line plot method is introduced for statistical analysis. The upper and lower bounds are calculated based on the interquartile range (IQR), which is defined as follows

$$IQR_{\text{upper}} = Q_3 + 1.5 \times IQR \tag{28}$$

$$IQR_{\text{lower}} = Q_1 - 1.5 \times IQR \tag{29}$$

Where, Q_1 and Q_3 are the 25th and 75th percentiles of the data, $IQR = Q_3 - Q_1$. Each data point IQR_x is identified. If it satisfies either condition $IQR_x < IQR_{\text{lower}}$ or $IQR_x > IQR_{\text{upper}}$, then an exception exists for that data point.

In this paper, we employ the average value method to address the outliers. The anomalous data is to be presented on the day i of the cycle m. For each station s_i, the inbound passenger flow during the period $[t, t+\varepsilon]$ is denoted as $status = 0$, and the outbound passenger flow is denoted as $status = 1$. Subsequently, the mean value of the concurrent passenger flow on the same day of the cycle before and after the cycle in which the outlier is situated is designated as the outcome of the repair as follows

$$\overline{x}^{m,status}_{([t,t+\varepsilon],s_i)} = \frac{x^{m-1,status}_{[t,t+\varepsilon],s_i} + x^{m,status}_{[t,t+\varepsilon],s_i} + x^{m+1,status}_{[t,t+\varepsilon],s_i}}{3} \tag{30}$$

E.2. Standardization of data

To enhance the training of the model, the data undergoes normalization, resulting in processed data that possesses a mean of 0 and a standard deviation of 1. In this paper, the Z-score standardization is introduced to process each of the original sample data $x_{i,j}$, and the operation $\hat{x}_{i,j}$ of the standardized data is as follows

$$\mu_j = \frac{1}{N}\sum_{i=1}^{N} x_{i,j} \tag{31}$$

$$\sigma_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(x_{i,j} - \mu_j\right)^2} \tag{32}$$

$$\hat{x}_{i,j} = \frac{x_{i,j} - \mu}{\sigma} \tag{33}$$

Where, N is the sample size; μ_j is the mean; σ_j is the standard deviation.

F. Indicators for evaluation

Following extant literature concerning the evaluation of performance metrics for passenger flow prediction models [18], [19], [20], this paper selects the metrics most frequently employed in traffic flow prediction, including Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The following formulas calculate the values of these metrics.

$$f_{\text{RMSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{34}$$

$$f_{\text{MAE}} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{35}$$

$$f_{\text{MAPE}} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{36}$$

Where, y_i is the actual value; $\hat{y}_i$ is the predicted value.

TABLE II
Partial Historical Traffic Data Set

time	lineID	stationID	deviceID	status	userID	payType
2019-01-02 00:00:00	C	39	1824	0	B958313f7b5b847d32120b6fa97587b3a	1
2019-01-02 00:00:01	B	8	384	0	Bdd932cd325d8e6021e71d0f27a149073	1
2019-01-02 00:00:03	B	2	74	0	B32a6c9e89459cf3afac6956d0ba2349f	1
…	…	…	…	…	…	…

IV. Example Analysis

A. Data set

To train and validate the performance of the model and evaluate the predictive effect, using the real passenger flow dataset of urban rail transit, for example validation in this paper. The dataset is from the data of the automatic fare collection (AFC, Automatic Fare Collection) system of the Hangzhou Metro subway, and the period is from January 1 to January 28, 2019. The data includes the time of passengers entering and leaving the station, station information, etc. The original data format is shown in Table II.

The dataset is divided in order to prevent the dissemination of information during the training process and to guarantee the impartiality of the test results. The initial two weeks are allocated for the training set, the third week is designated for the validation set, and the remaining days are apportioned for the test set.

B. Analysis of passenger flow feature

A thorough examination of historical passenger flow data is imperative to elucidate the characteristics of passenger flow fluctuations and their potential influencing factors. Furthermore, a comprehensive analysis of the factors that may affect the characteristics of passenger flow fluctuations is necessary. The original dataset is first subjected to a series of cleansing and preprocessing steps, after which outliers remain in the processed data. Because reliance on the box-and-line plot method may result in the exclusion of inevitable fluctuations in passenger flow that are, in fact, realistic possibilities. However, these reasonable outliers must be retained to reflect the dynamic change characteristics of passenger flow more comprehensively when identifying passenger flow peaks, special events, extreme weather, etc. Therefore, the cleaning and retention of outliers must be balanced.

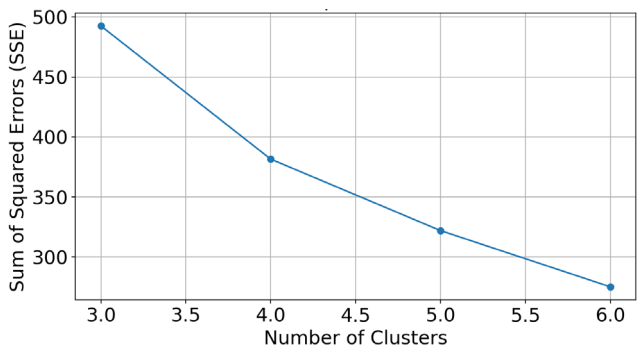


Fig. 5. Elbow method for cluster number determination

The categorization of sites is achieved through the implementation of the k-means clustering algorithm, a statistical method that involves the combination of data pertaining to historical passenger flow, geographic latitude and longitude of the site, and surrounding land use.

The final number of clusters is also influenced by the outcomes of the elbow method, a multifaceted approach designed to guarantee the rationality and stability of value selection k, as illustrated in Fig. 5.

By calculating the contour coefficients at different values k, the highest contour coefficient of approximately 0.375 is reached $k = 4$. At this point, the tightness of data within clusters and the separation of data between clusters reach an optimal balance. The site clustering results are shown in Fig. 6.

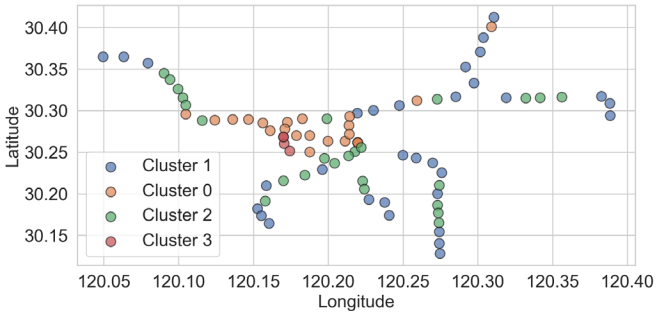


Fig. 6. Results of the k-means clustering of sites

As demonstrated in Fig. 6, the urban rail transit stations have been classified into four distinct types, designated as cluster 0, 1, 2, and 3, according to the results of the k-means clustering analysis. These categories represent the archetypal samples obtained through cluster analysis.

The Zhonghe North Road Station has been selected as a representative business office station within cluster 0. It is situated in the primary business district of Hangzhou City, near numerous office and commercial buildings, experiencing a high volume of passenger traffic with discernible morning and evening rush hours. The passenger flow in and out of the station fluctuates significantly, exhibiting clear peaks and valleys that reflect the observable commuting patterns. The passenger flow on weekdays exhibits a high level of variability, and the fluctuations in passenger flow demonstrate a strong time dependence and regularity. Given its status as a business district station, the passenger flow pattern necessitates precise short-term forecasting and possesses substantial representative value.

For cluster 1, the station of the University of Traditional Chinese Medicine is selected as a representative of typical university and research park stations. The university's location in the vicinity of prominent academic institutions

and scientific research facilities contributes to a stable yet less volatile flow of passengers, influenced by the rhythms of teaching and research activities. The inbound and outbound passenger flows exhibit stability, characterized by minor fluctuations, slow intra-day variations, and discernible intra-week rhythms. The impact of holidays during the semester is significant, and seasonal patterns are prominent. It is considered a representative of a low volatility, high regularity type.

The station is situated within a predominantly residential neighborhood, interspersed with limited commercial and living facilities. The regional pace of life influences its passenger flow. For these reasons, the Nanxing Bridge Station has been selected as the representative of Cluster 2, which is a typical residential area and mixed functional area station. The inbound peak is more pronounced, reflecting the travel patterns of residents as they commute to and from work during the morning and evening rush hours. The fluctuation of outbound passenger flow is relatively smooth, which makes it suitable for the study of passenger flow prediction of life service stations.

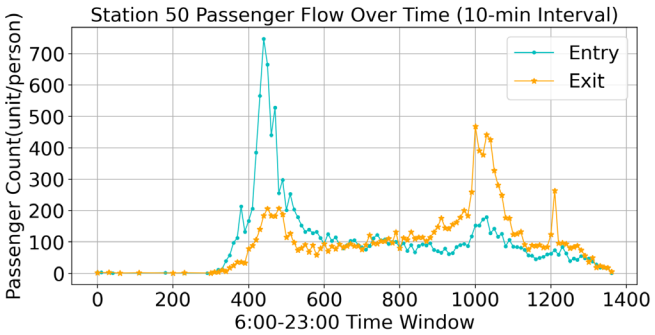
Fengqi Road Station sits in central Hangzhou. It connects subway lines and many bus routes, which makes it a busy and complex hub. It serves as a key downtown interchange station. Its inbound and outbound passenger flow fluctuates wildly and is significantly affected by commercial activities, holidays, urban activities, and other factors. The peak-hour passenger flow is highly concentrated and changes rapidly, and the difference in passenger flow between different days is noticeable. The model must consider non-linear and unexpected factors in passenger flow predictions. It works well as a representative of complex models.

Stations in each cluster have similarities in passenger flow characteristics (e.g., fluctuation amplitude, daily change pattern, the ratio of inbound and outbound stations, etc.), so these stations are selected as representatives of each cluster, which can better reflect the typical features of the stations in that category. These stations have relatively complete and high-quality data, with no obvious missing data or anomalies, which can ensure the reliability and consistency of the data in the analysis and modeling process.

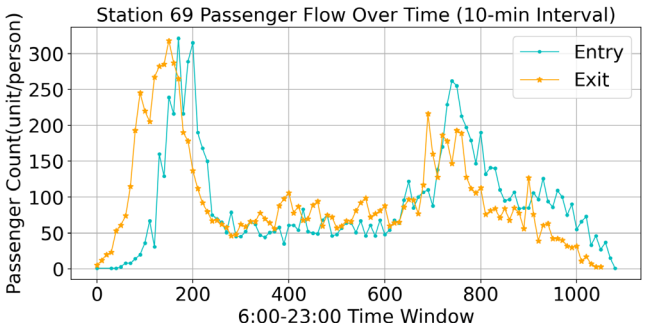
TABLE III
Results of the k-means Clustering of Sites

cluster 0	cluster 1	cluster 2	cluster 3
Zhonghe North Road Station	Chinese Medicine University Station	Nanxing Bridge Station	Fengqi Road Station
Xintang Station	Zhenning Road Station	Renmin Road Station	Longxiang Bridge Station
Xinfeng Station	Xixing Station	Wenze Road Station	Anding Road Station
…	…	…	…

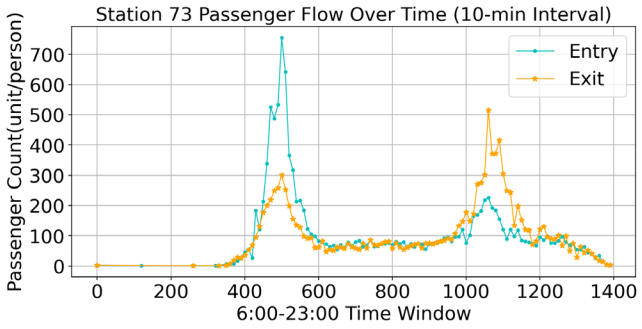
In order to evaluate the prediction performance of the model for different types of stations, according to the results of station clustering, take the period 6:00-23:00 to analyze the time sequence characteristics of the passenger flow of different categories of stations on an ordinary weekday with a time granularity of 10min, as shown in Figs. 9.



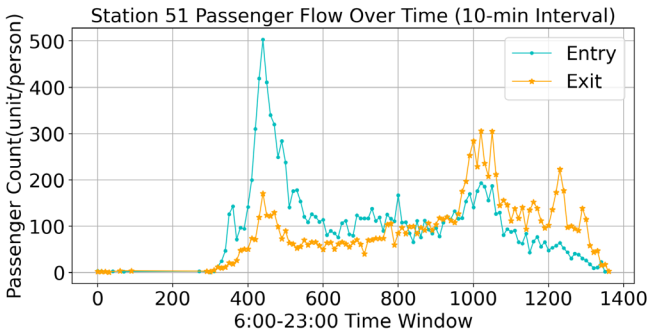
(a) cluster 0



(b) cluster 1



(c) cluster 2



(d) cluster 3

Fig. 7. Historical passenger flow timing features for different types of stations

As illustrated in Fig. 9, the historical passenger flow curves of different station types demonstrate that they exhibit distinct passenger flow dynamics. Passenger flows at all types of stations exhibit cyclical fluctuations, particularly during the morning and evening peak hours (approximately 6:00-9:00 a.m. and 5:00-8:00 p.m.). The trends of inbound and outbound passenger flows are generally similar; however, certain stations exhibit slightly staggered periods of high and low passenger traffic, which is indicative of the spatial and temporal features of passenger flows. Accordingly, to predict the short-term inbound passenger flow of urban rail transit, it

is necessary to train the model according to the clusters to which the stations belong separately. Because the clusters exhibit different fluctuation patterns, the prediction accuracy will be improved.

C. Experimental settings

In this paper, the solution of deep learning algorithms is based on the PyTorch framework, which is implemented on a Windows 11 64-bit operating system configured with the PyCharm 2021.3 (Community Edition) development environment. The GPU parallel computing architecture of the system is CUDA version 11.8, and the accompanying cuDNN version is 8.7.0. The models constructed in this paper are built using the Python 3.9 programming language and calling third-party libraries of Python, such as matplotlib, pandas, numpy, sklearn, and so on.

The initial model parameters are such that each channel of the DC-CNN part extracts data through a one-dimensional convolutional operation. This convolutional layer contains 32 convolutional kernels with a kernel size of 3. A padding method is used to maintain the output size. A ReLU activation function and a max pooling layer with a pooling window size of 2 back the convolutional layer. Additionally, a Dropout layer is employed, with the initial dropout rate set to 0.2. The BiLSTM part utilizes a two-layer LSTM network, with each layer comprising 128 hidden units. The batch_first=True setting is employed to accommodate the data dimensionality, and Dropout regularization is applied between layers, with an initial dropout rate set to 0.3.

To capture longer-term periodicity or trends, set the time window size to (6,24) based on the 10-minute data time frequency.

D. Analysis of predicted results

D.1. Hyperparametric analysis

According to the extant literature on the subject[21],[22], the judicious selection of step size is instrumental in ensuring that the model is capable of comprehensively capturing the time-varying patterns and periodic characteristics inherent in the passenger flow data. In turn, it enhances the model's capacity to discern alterations in short-term flow. Furthermore, there are discrepancies in the dynamic fluctuations of passenger flow across different subway stations and over various periods. The selection of an appropriate input step size facilitates the model's capacity to adapt to these distinct temporal patterns. During the hyperparameter tuning process, by systematically exploring the step-size parameter space, the model can be prevented from overfitting the data features of a specific historical window length, which helps to improve the applicability and stability of the model in the actual complex and changing environment.

Since the input step size directly affects the model's ability to learn time series patterns, it is the benchmark for all subsequent experiments. Therefore, experiments are designed to compare the prediction performance of different input steps of 30 minutes, 40 minutes, 50 minutes, 1 hour, and 2 hours, respectively, and to determine the optimal time window in combination with the location clustering results to avoid feature loss or noise interference due to an inappropriate period.

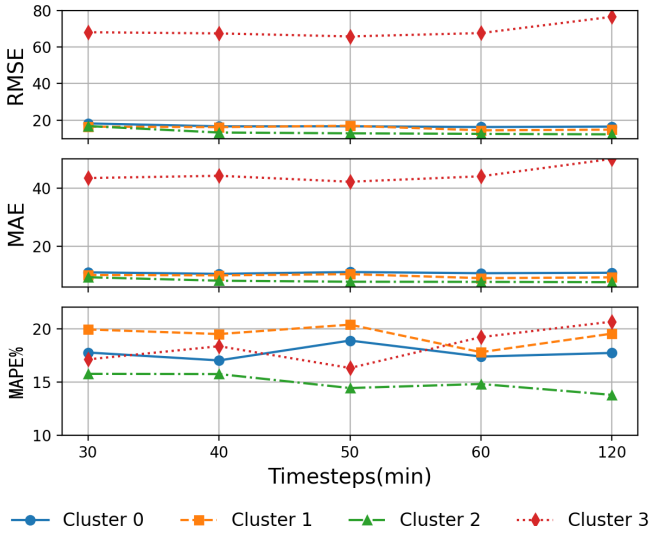


Fig. 8. Comparison of evaluation metrics for different input step prediction results for each type of site

As shown in Fig. 8, the utilization of RMSE, MAE, and MAPE as evaluation metrics has been demonstrated to reveal discrepancies in the performance of the prediction model across varying input steps. The sensitivity of the root mean square error (RMSE) and the mean absolute error (MAE) to variations in the input step size is found to be minimal, and a relatively smooth profile characterizes the resulting curve. The MAPE curve demonstrates significant fluctuations, suggesting that the input step size exerts a substantial influence on the model's prediction accuracy. It can be concluded that as the MAPE value decreases, the model's performance improves.

Based on this, the historical passenger flow characteristics of each type of station, the impact of high and low peak hours on the station's passenger flow, the commuting pattern, the nature of the surrounding land, as well as the changing pattern of the evaluation indices RMSE, MAE, MAPE and other factors are comprehensively considered, and the input step length of the cluster 0 and cluster 2 stations is determined to be 40 min; the input step length of the cluster 1 station is 1 hour (60 min); the input step length of the cluster 3 site is 50 min, and all subsequent experiments are based on this.

The learning rate constitutes the fundamental hyperparameter of model training, exerting a direct influence on the convergence speed and stability of gradient descent. A substantial body of research has demonstrated that the selection of an optimal learning rate during the model training phase is conducive to enhancing the model's predictive performance [23], [24]. To further study the impact of learning rate on the prediction performance of various types of sites, the changes in the prediction accuracy of the observed model when setting different learning rates (learning rate = 0.0001, 0.001, 0.005, 0.01) are used to determine the reasonable range quickly, which avoids the subsequent experiments wasting computational resources due to inappropriate learning rates and laying the foundation for the subsequent optimization of parameter combinations. Preliminary experimental findings indicate that RMSE and the MAE demonstrate minimal responsiveness to variations in the learning rate. Consequently, the emphasis is directed towards the examination of the change in MAPE, as illustrated in Fig. 9.

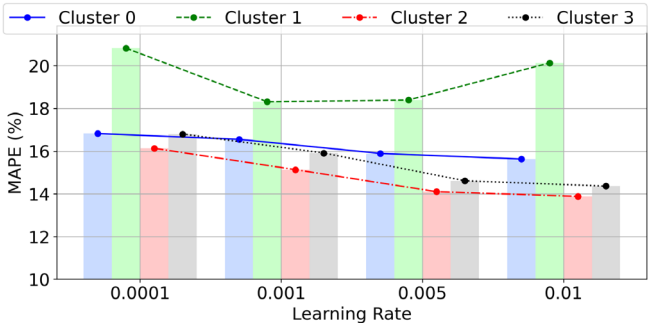
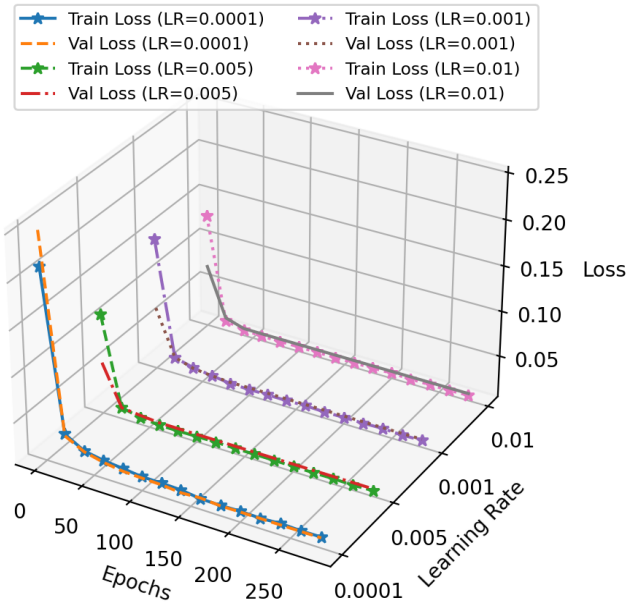


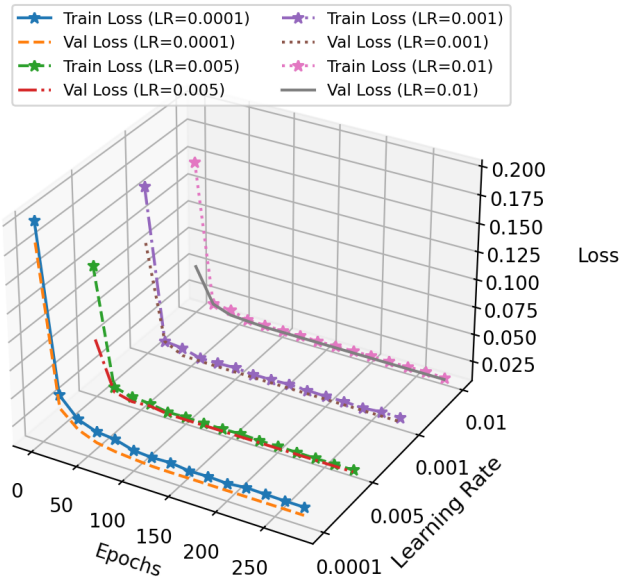
Fig. 9. Comparison of evaluation metrics for prediction results at different learning rate sizes for each type of site

As demonstrated in Fig. 9, the MAPE of Clusters 0, 2, and 3 exhibits a downward trend as the learning rate increases, which suggests that a higher learning rate facilitates the convergence of the model. Notably, Cluster 0 demonstrates a consistently low MAPE across all learning rates, indicating its superior stability and prediction accuracy. However, an elevated learning rate can adversely impact the prediction performance of Cluster 1, as evidenced by a substantial increase in MAPE at higher learning rates. This finding suggests that the model exhibits heightened sensitivity to the learning rate when predicting this particular type of site data. A critical evaluation of the relationship between convergence speed and prediction accuracy reveals that a learning rate of 0.01 is more appropriate for Cluster0, Cluster2, and Cluster3. Conversely, a lower learning rate of 0.001 is employed for Cluster 1 to avert performance degradation and ensure model stability.

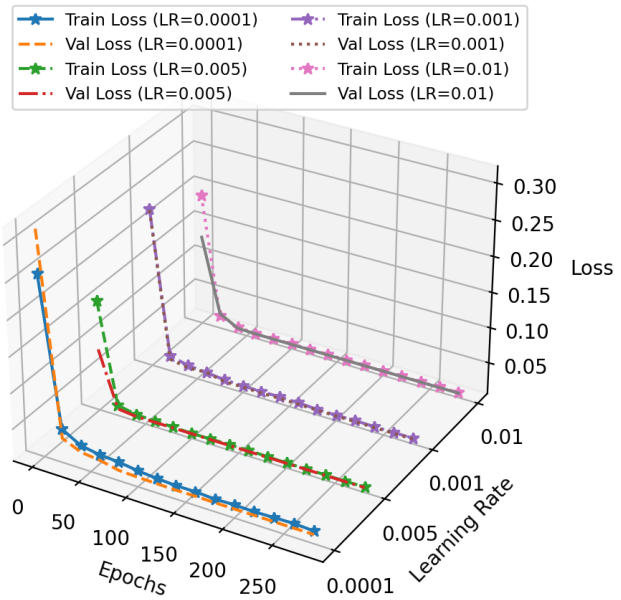
To reduce the risk of underfitting and overfitting the model, to effectively shorten the training time of the subsequent experiments, and to ensure that the model finds the number of rounds in which the validation set error is stable during a reasonable training phase, epochs tuning experiments are designed, and the training curves are shown in Fig. 10.



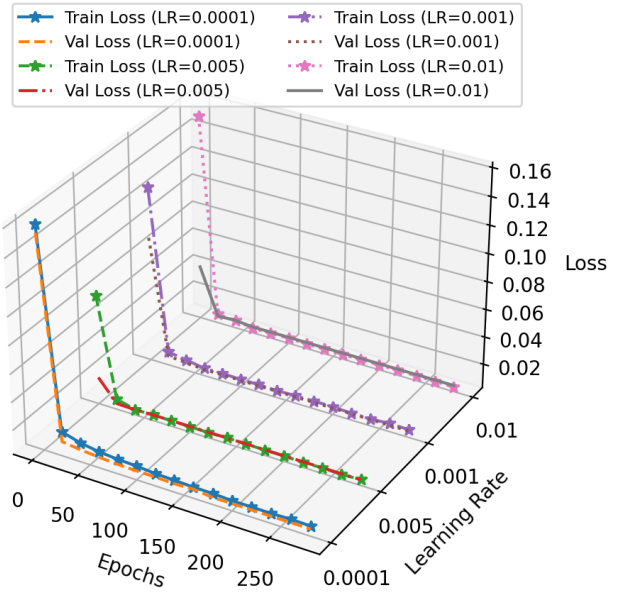
(a) cluster 0



(b) cluster 1



(c) cluster 2



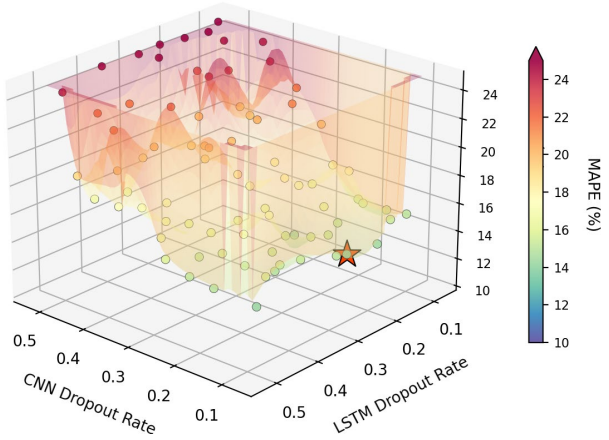
(d) cluster 3

Fig. 10. Model loss curves for each type of site at different learning rates and epoch settings

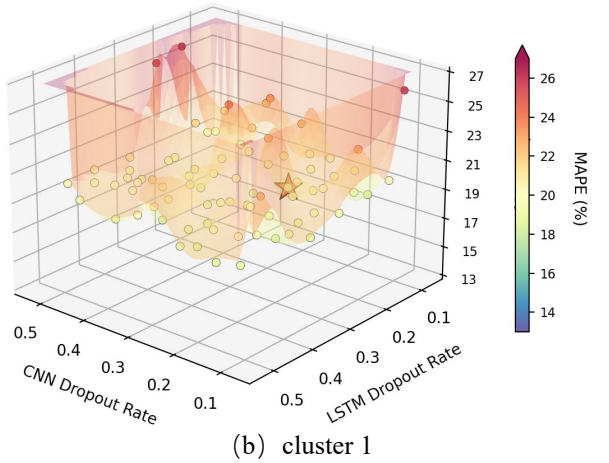
As can be seen from Fig. 10, preliminary observations indicate that the curve graph meticulously documents both the training set loss (Train loss) and the validation set loss (Val loss).

Furthermore, the loss function progressively converges to a steady state when epochs surpass 60, signifying that the model attains a stable performance level. This critical value can be regarded as the minimum epoch value required to adapt to the features of each type of site, maximizing the cost of training time under the premise of guaranteeing the performance of the model.

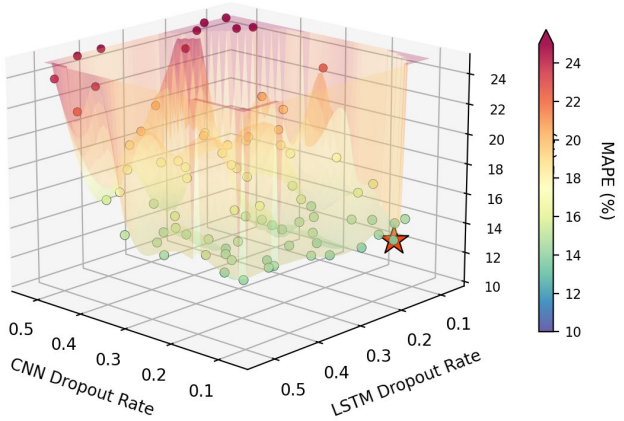
Under the condition that the previous study significantly improves the initial prediction accuracy of the model, considering the existence of limited training data, to further improve the generalization ability of the model and avoid the risk of overfitting, the impact of the regularization parameter of Dropout Rate (DRR) on the performance of the model is investigated. For each type of site model that has been preliminarily optimized, the dropout rate of the CNN and Bi-LSTM modules is systematically adjusted using the grid search method to assess its impact on the prediction accuracy of the models. The parameters are set within the range of 0.1-0.5, in which the two-stage method of rough searching every 0.1 and fine searching every 0.05 is introduced. Moreover, the early-stopping strategy is adopted to facilitate the search for an optimal solution that balances the searching efficiency and searching accuracy. As demonstrated in Fig. 16, the scatter plot illustrates the prediction evaluation indexes for various dropout combinations categorized by site type.



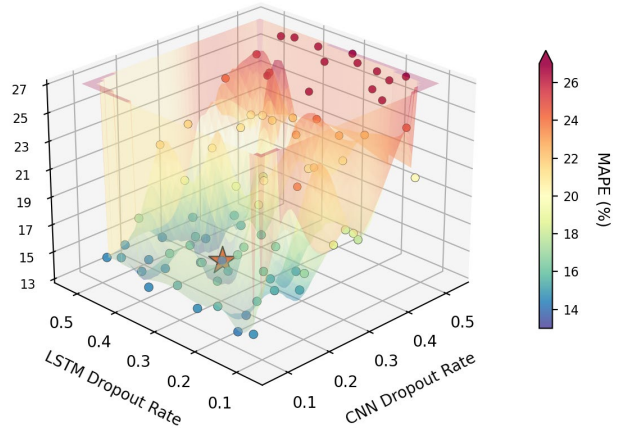
(a) cluster 0



(b) cluster 1



(c) cluster 2



(d) cluster 3

Fig. 11. Comparison of the predictive performance of different dropout combinations for various types of sites

As demonstrated by Fig. 11, the MAPE values of the prediction results exhibit substantial variations across different dropout combinations, exhibiting an overall positive correlation trend. That means the prediction accuracy of the model is comparatively higher under dropout combinations with lower values. The points in the figure that are labeled with stars correspond to the combinations that exhibited the smallest MAPE for all combinations of sample evaluation indexes in the experiment. These combinations can be considered the optimal ones. The four types of sites are 14.02%, 18.12%, 13.57%, and 13.82%, in that order. These values correspond to the optimal dropout (CNN, LSTM) combinations, respectively: cluster 0 (0.1, 0.25); cluster 1 (0.3, 0.15); cluster 2 (0.1, 0.15); cluster 3 (0.25, 0.35).

The results of the aforementioned hybrid parameter tuning experiments of manual parameter optimization and automated hyperparameter search algorithm are finally integrated. Based on the optimal parameter combinations with relatively balanced prediction performance and efficiency, the future 10-minute inbound passenger flows of Cluster 0, 1, 2, and 3 stations are predicted to ensure that the overall model performance is maximized.

The prediction results demonstrate a comparison between the actual and predicted passenger flow curves of the two cycles, as illustrated in Figs. 12, 13, 14, and 15, respectively.

As demonstrated in Figs. 12, 13, 14, and 15, the predicted and actual value curves of the four stations exhibit analogous undulation and change trends. The shaded segment of the

figure denotes the 10% error band. Except for individual exceptional extreme values, the majority of the prediction results and real values are contained within the error band. The prediction results align with the actual flow of passengers in the timing characteristics of the law, thereby reflecting the prediction performance of the model.

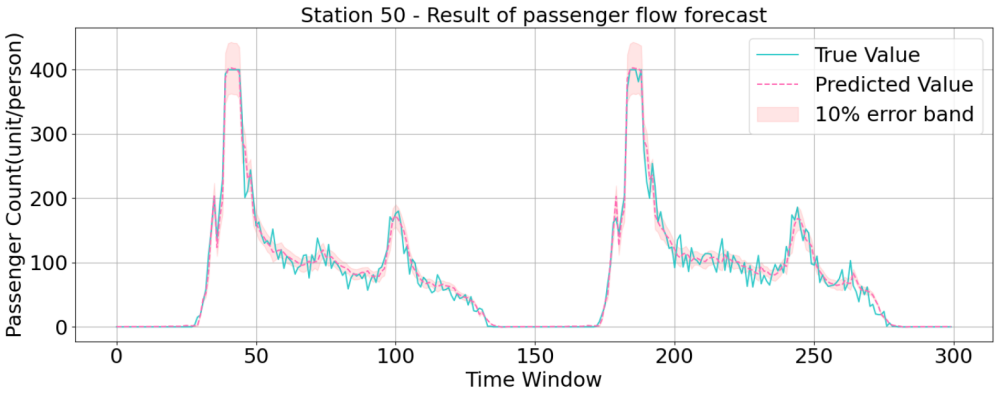


Fig. 12. Result of predicting 10-min of inbound passenger flow for cluster 0

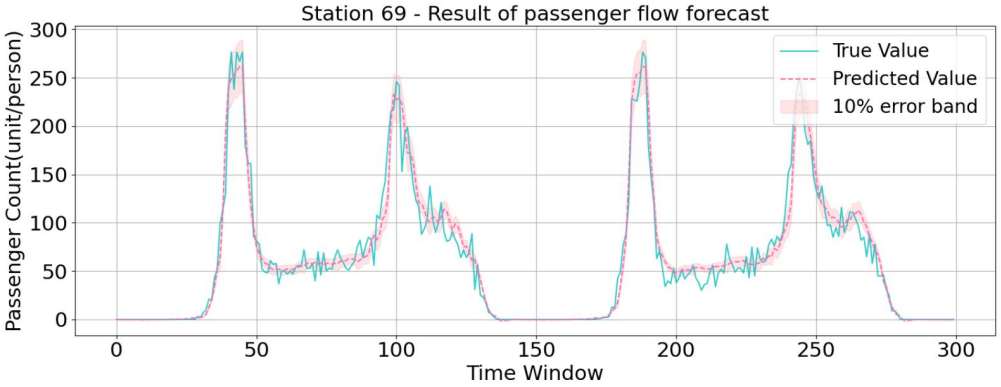


Fig. 13. Result of predicting 10-min of inbound passenger flow for cluster 1

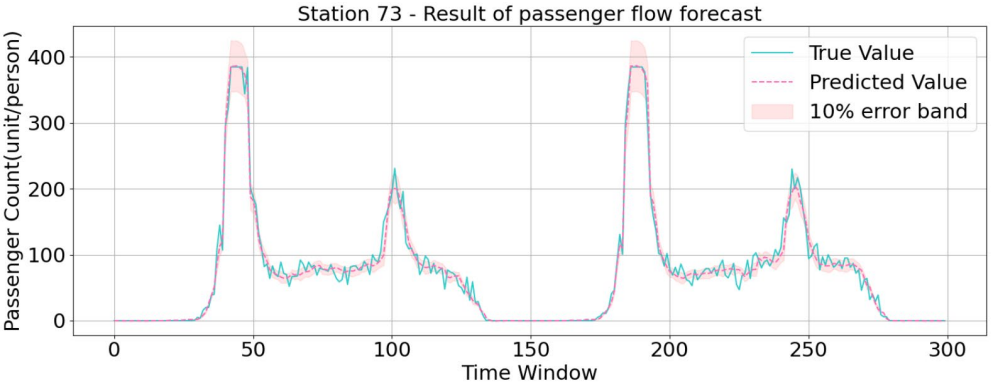


Fig. 14. Result of predicting 10-min of inbound passenger flow for Cluster 2

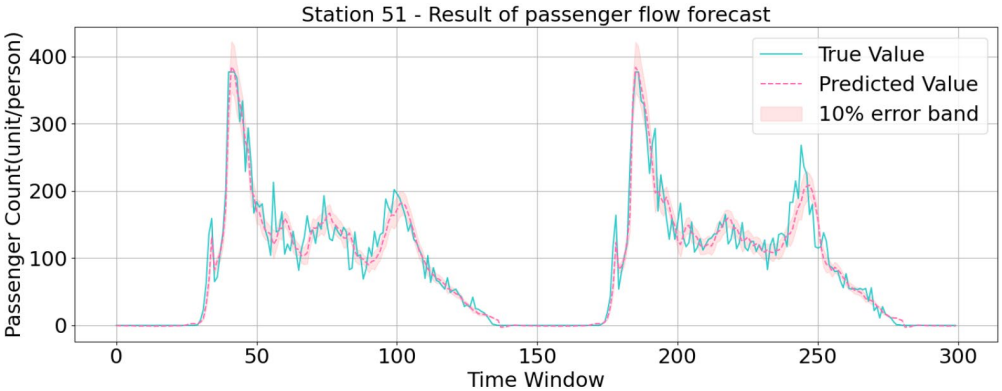


Fig. 15. Result of predicting 10-min of inbound passenger flow for Cluster 3

D.2. Comparative analysis of baseline model

In order to validate the effectiveness of the prediction model (hereafter abbreviated as DC-CNN-BiLSTM) proposed in this paper that combines site classification, the prediction accuracy of the following baseline model and the combined model are compared, and each evaluation metric is shown in Table 4.

TABLE IV
Comparison of the predictive effectiveness of different baseline models

model	Cluster	Evaluation indicators		
		RMSE	MAE	MAPE%
ARIMA	0	48.58	28.06	28.5
	1	55.71	40.54	41.2
	2	46.34	25.89	25.8
	3	50.66	29.95	30.1
RF	0	20.85	14.03	26.44
	1	22.61	13.46	31.66
	2	17.83	10.87	25.81
	3	28.48	18.76	29.67
LSTM	0	18.47	11.61	16.83
	1	17.37	10.69	20.83
	2	17.10	9.67	16.14
	3	73.26	44.70	16.81
CNN-BiLSTM	0	15.29	9.48	15.64
	1	16.12	9.71	18.32
	2	13.14	7.69	13.89
	3	69.95	42.56	14.37
DC-CNN-BiLSTM	**0**	**15.18**	**9.31**	**14.02**
	1	**15.49**	**9.4**	**17.85**
	2	**12.55**	**7.51**	**13.57**
	3	**32.68**	**20.36**	**13.82**

An analysis of Table 4 reveals that irrespective of the type of station for passenger flow prediction, the combination model exhibits superior prediction efficacy in comparison to a single model. The prediction accuracy of the deep learning model surpasses that of traditional mathematical statistics models, such as ARIMA, and machine learning models, such as RF. The evaluation indexes demonstrate enhancements, albeit to varying extents. The passenger flow patterns exhibited by clusters 0, 2, and 3 are characterized as being between regularity and volatility. The non-linear capability of the deep learning model is demonstrated to be advantageous, with the prediction results of MAPE reaching below 15% in all cases. As model complexity has increased and an early stop training strategy has been incorporated, the MAPE values of the model for passenger flow prediction have decreased, ranging from 14.02% to 13.57% and 13.82%, respectively. These results suggest that the model can achieve a more accurate prediction.

However, for Cluster 1, compared with other types of stations, the errors of all models are larger, with a minimum difference of 2.72% and a maximum of 15.4% in the MAPE values predicted for different stations. A subsequent analysis of the underlying reasons indicates that the presence of these stations in commercial areas or their status as interchange stations may be contributing factors. Additionally, the fluctuations in passenger flow observed at these stations could be a contributing element to the variations in performance metrics.

V. Procedure for Paper Submission

In this study, a combined DC-CNN-BiLSTM prediction model was constructed, and the effects of station heterogeneity and model training strategy on the prediction results were considered. The future 10-minute inbound passenger flow of urban rail transit was predicted, and the following summary was made based on the research work.

(1) A multi-dimensional index system is constructed, incorporating the ontological attribute features of subway stations and the location environment features. The historical passenger flow features of the stations are then fused with the newly constructed index system. The k-means clustering algorithm is subsequently employed to classify the rail transit stations. The objective is to accurately ascertain the various characteristics of passenger flow and to formulate a more rational urban rail transit station classification scheme. The research sample is divided into four categories of typical stations (Cluster 0, 1, 2, 3). The analysis shows that the different categories of stations present significant heterogeneity in the distribution of passenger flow in the spatial and temporal dimensions. This heterogeneity lays the foundation for differentiated passenger flow prediction.

(2) In each cluster, a representative station is selected as a typical case for in-depth data analysis, results presentation, and comparison. This selection is made to visually reflect the passenger flow characteristics of different types of stations.

Cluster 0: The station's passenger flow characteristics are typical, susceptible to the strong influence of the work cycle, and suitable for prediction modeling to deeply mine the changes of the peaks and valleys in the time sequence. The Zhonghe North Road station is selected as a typical station. Chinese Medicine University Station

Cluster 1: This station type reflects the daily travel patterns of university students and staff, which facilitates the establishment of a relatively smooth prediction model.

Cluster 2: This type of life service station is a prominent component of the passenger flow in the Hangzhou metro system, with a significant number of stations within the system. The Nanxing Bridge Station is selected as a representative station for this cluster.

Cluster 3: The complexity of this hub station type poses a challenge to short-term prediction. The passenger flow pattern is challenging to predict in the short term due to its high fluctuation and uncertainty. Fengqi Road Station is a representative station for this pattern. The selection of representative stations for each cluster facilitates the adoption of more targeted short-time inbound passenger flow prediction model training strategies for different types of stations in the example analysis.

(3) A combined prediction model of a convolutional neural network, combined with a two-way long and short-term memory network, was constructed to analyze and extract temporal and spatial features of passenger flow. This model was developed based on an analysis of historical passenger flow features of urban rail transit. The integration of spatial and temporal features is employed to forecast the inbound passenger flow of urban rail transit shortly. A comparison experiment was also designed with each baseline prediction model, which was verified by the example of subway AFC data, to show that the combined prediction model proposed in this paper has better prediction performance.

The traditional classical model's limited adaptability in complex scenarios poses a significant challenge in achieving accurate predictions for stations with more complex

passenger flow features. This discrepancy in prediction accuracy compared to other stations with relatively regular and stable passenger flow features is a notable concern.

(4) In particular, with regard to the training of models, the impact of varying model parameter values on the efficacy of training is taken into account. Parameter tuning of the model is conducted in consideration of the parameters of input step size, learning rate, number of training rounds, and combined discard rate, respectively. A hybrid tuning strategy has been developed that integrates manual parameter optimization and automated hyperparameter search algorithms. This strategy also incorporates an early-stop mechanism to balance model performance and training efficiency.

Following an examination of the training process with example validation, it was determined that the sensitivity of various hyperparameters to model prediction accuracy demonstrates significant heterogeneity. Furthermore, the degree of influence on the prediction effect of different types of sites varies. In addition to the advantages of the combined model structure on the improvement of prediction performance, for clusters 0-3, the DC-CNN-BiLSTM prediction results under the use of early-stopping and grid-search training strategies are significantly more accurate compared to the original data. Furthermore, the average absolute percentage error can be reduced by a maximum of 1.62%, which suggests that the high-quality model training does contribute to the improvement of prediction accuracy.

(5) Despite the enhancement in the precision of passenger flow prediction across all station types subsequent to the study outlined in this paper, substantial variations in the stability of prediction persist among distinct station types. Notably, within cluster 1 stations, the prediction accuracy remains suboptimal, underscoring the necessity for further optimization.

The analysis may be due to the fact that this study relies on historical card swipe data and does not integrate external variables such as unexpected events, which leads to increased prediction bias in extreme scenarios. It is expected that we will continue to improve prediction accuracy by exploring more advanced deep-learning architectures or hybrid models in future studies.

REFERENCES

[1] L. X. Si, M. R. Tang, H. Q. Tong, Y. C. Ju, and Z. J. Zeng, "Research on the characteristics of commuting passenger flow of Hangzhou metro," *Modern Urban Transit*, no.7, pp.100-108, 2024.

[2] Y. S. Jiang, G. S Yu, and L. Hu, "Refined Classification of Urban Rail Transit Stations Based on Clustered Station's Passenger Traffic Flow Features," *Journal of Transportation Systems Engineering and Information Technology*, vol.22, no.4, pp.106-112, 2022.

[3] Z. J. Zhu, Y. Zhang, J. R. Sun, S. Y. Zhang, B. R. Han, and Y. P. Zhao, "Community Structure Division and Ridership Characteristics Analysis of Rail Transit Stations Based on the Louvain Algorithm," *Journal of Transportation Engineering, Part A: Systems*, vol.150, no.8, 2024.

[4] J. Xiong, W. Guan, and Y. X. Sun, "Metro transfer passenger forecasting based on Kalman filter," *Journal of Beijing Jiaotong University*, vol.37, no.3, pp. 112-116+121, 2013.

[5] C. Q. Ma, P. K. Li, C. H. Zhu, and T. Tian, "Short-term passenger flow forecast of urban rail transit based on different time granularities," *Journal of Chang'an University: Natural Science Edition*, vol.40, no.3, pp.75-83, 2020.

[6] G. Y. Zhang, H. Jin, "Research on the prediction of short-term passenger flow of urban rail transit based on improved ARIMA model," *Computer Applications and Software*, vol.39, no.1, pp.339-344, 2022.

[7] L. H. Li, J. S. Zhu, L. X. Qiang, and Q. J. Qiao, "Study on Forecast of High-speed Railway Short-term Passenger Flow based on Random Forest Regression" *Railway Transport and Economy*, vol.39, no.9, pp.12-16, 2017.

[8] S. S. Liu, E. J. Yao, "Holiday Passenger Flow Forecasting Based on the Modified Least-Square Support Vector Machine for the Metro System," *Journal of Transportation Engineering, Part A: Systems*, vol.143, no.2, 2017.

[9] J. Jin, H. Y. Wang, M. Li, "Prediction of the Metro Section Passenger Flow Based on Time-Space Characteristic," *Applied Mechanics and Materials*, vol.2658, no. 97-400, pp.1038-1044, 2013.

[10] Q. Ouyang, Y. B. Lv, J. H. Ma, and J. Li, "An LSTM-Based Method Considering History and Real-Time Data for Passenger Flow Prediction," *Applied Sciences*, vol.10, no.11, 2020.

[11] J. Li, Q. Y. Peng, C. Wen, "Short term passenger flow prediction of high speed railway based on LSTM deep neural network," *Systems Engineering - Theory & Practice*, vol.41, no.10, pp.2669-2682, 2021.

[12] T.D. Sajanraj, J. Mulerikkal, S. Raghavendra, R. Vinith, and V. Fabera, "Passenger flow prediction from AFC data using station memorizing LSTM for metro rail systems," *Neural Network World*, vol.31, no.3, pp.173-189, 2021.

[13] C. Y. Lin, K. Wang, D. Y. Wu, and B. W. Gong, "Passenger Flow Prediction Based on Land Use around Metro Stations: A Case Study," *Sustainability*, vol.12, no.17, pp.6844-6844, 2020.

[14] J. L. Zhang, F. Chen, Y. N. Guo, and X. H. Li, "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, 2020.

[15] B. X. Cao, Y. X. Li, Y. Y. Chen, and A. A. Yang, "CNN-LSTM Model for Short-Term Passenger Flow Forecast Considering the Built Environment in Urban Rail Transit Stations," *Journal of Transportation Engineering, Part A: Systems*, vol.150, no.11, 2024.

[16] X. L. Ma, Z. Dai, Z. B. He, J. H. Ma, Y. Wang, and Y. P. Wang, "Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction," *Sensors*, vol.17, no.4, pp.818, 2017.

[17] J. Liu, Q. L. He, Z. K. Yue, and Y. L. Pei, "A Hybrid Strategy-Improved SSA-CNN-LSTM Model for Metro Passenger Flow Forecasting," *Mathematics*, vol.12, no.24, pp.3929-3929, 2024.

[18] L. Zeng, Z. N. Li, J. Yang, and X. Y. Xu, "CEEMDAN-IPSO-LSTM: A Novel Model for Short-Term Passenger Flow Prediction in Urban Rail Transit Systems," *International Journal of Environmental Research and Public Health*, vol.19, no.24, pp.16433-16433, 2022.

[19] Q. C. Chen, D. Wen, X. Q. Li, D. J. Chen, H. X. Lv, J. Zhang, and P. Gao, "Correction: Empirical mode decomposition based long short-term memory neural network forecasting model for the short-term metro passenger flow," *PloS one*, vol.15, no.3, pp. e0231199, 2020.

[20] L. X. Wei, D. J. Guo, Z. L. Chen, J. C. Yang, and T. L. Feng, "Forecasting Short-Term Passenger Flow of Subway Stations Based on the Temporal Pattern Attention Mechanism and the Long Short-Term Memory Network," *ISPRS International Journal of Geo-Information*, vol.12, no.1, pp.25-25, 2023.

[21] J. N. Sun, X. F. Ye, X. C. Yan, T. Wang, and J. Chen, "Multi-Step Peak Passenger Flow Prediction of Urban Rail Transit Based on Multi-Station Spatio-Temporal Feature Fusion Model," *Systems*, vol.13, no.2, pp.96-96, 2025.

[22] Nithin K. Shanthappa, Raviraj H. Mulangi, and Harsha M. Manjunath, "Deep learning-based public transit passenger flow prediction model: integration of weather and temporal attributes," *Public Transport*, pp.1-24, 2024.

[23] Z. W. Hou, J. Han, and G. Yang, "Analysis of Passenger Flow Characteristics and Origin–Destination Passenger Flow Prediction in Urban Rail Transit Based on Deep Learning," *Applied Sciences*, vol.15, no.5, pp.2853-2853, 2025.

[24] L. N. Liu, Y. G. Liu, X. F. Ye, "Multi-sequence spatio-temporal feature fusion network for peak-hour passenger flow prediction in urban rail transit," *Transportation Letters*, vol.17, no.1, pp.86-102, 2025.

[25] X. Zhao, C. X. Li, X. T. Zou, X. W. Du, and Ahmed Ismail, "Passenger Flow Prediction for Rail Transit Stations Based on an Improved SSA-LSTM Model," *Mathematics*, vol.12, no.22, pp.3556-3556, 2024.

Jia He was born in Ningxia, China, in 2001. She received her Bachelor's degree in Traffic and Transportation from Chang'an University, China, in 2023. Currently, she is pursuing a Master's degree in Traffic and Transportation Engineering at Lanzhou Jiaotong University. Her research interests include urban rail transit operations and passenger travel characteristics.