Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A TCN-Transformer Hybrid Deep Learning Model for Multivariate Time Series Forecasting with Multi-source Feature Fusion

Shijie CEN
Jing LU
7187541180qq.com

Zhijian LAI
Wei HUANG

Research Article

Keywords: multivariate time series forecasts, hybrid deep learning models, temporal convolutional networks, transformer models, multi-source features, passenger flow forecasts

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11158856/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

A TCN-Transformer Hybrid Deep Learning Model for Multivariate Time Series Forecasting with Multi-source Feature Fusion

Shijie CEN[1], Jing LU（Corresponding author）*[1], Zhijian LAI [1,2], Wei HUANG[1]*

([1]People's Public Security University of China, Xicheng District, Beijing 100038, China

[2]Guangxi Police College, Nanning, Guangxi 530028, China)

Keywords: multivariate time series forecasts, hybrid deep learning models, temporal convolutional networks, transformer models, multi-source features, passenger flow forecasts

Abstract: Accurate multivariate time series forecasting is a fundamental task for intelligent operation and decision-making across industrial scenarios. Existing hybrid time series models fail to simultaneously capture fine-grained local fluctuations and long-range periodic dependencies, and often neglect the coupled effects of heterogeneous external features, resulting in limited accuracy under frequent external shocks. To address these issues, this paper proposes a dual-branch parallel TCN-Transformer hybrid deep learning model for daily multivariate time series forecasting. First, a multi-source feature dataset is constructed by integrating historical values, calendar attributes, policy variables and meteorological features. A dual-branch encoding architecture is designed: the TCN branch extracts local short-term fluctuations via dilated causal convolutions, while the Transformer branch captures long-term periodic dependencies through multi-head self-attention. Adaptive average pooling and feature concatenation achieve deep fusion of local and global representations. Taking daily inbound passenger flow of the Hong Kong-Macau Ferry Terminal as an empirical case, experiments on a three-year dataset show that the proposed model achieves an R^2 of 0.953, an RMSE of 357.93 and an MAE of 274.27 on the test set, outperforming single TCN, Transformer, LSTM and CNN baselines. The model provides an effective solution for multivariate time series forecasting with multi-source heterogeneous features.

1. Introduction

Accurate passenger flow forecasting at border checkpoints provides an essential foundation for scientific decision-making in checkpoint management. As global cross-border mobility continues to increase, the number of people entering and leaving China through border checkpoints rises rapidly. This growth also increases the risks of clearance congestion, crowding, stampedes, and other forms of disorder, posing substantial challenges to national border security. Against this background, forecasting inbound and outbound passenger flows has become an important research topic in checkpoint safety management.

Existing passenger flow forecasting methods for border checkpoints can generally be divided into two categories. The first category comprises traditional forecasting approaches represented by exponential smoothing and regression analysis. For example, Wu (2011)[1] develops a forecasting model based on triple exponential smoothing to predict annual inbound and outbound passenger volumes in a specific region. Yu (2016)[2] applies exponential smoothing to historical passenger transport data to forecast passenger volumes at the Heihe checkpoint. Chi and Zhang (2013)[3] use regression analysis to predict annual cross-border passenger flows at the Zhuhai-Macau checkpoints. The second category includes shallow machine learning methods represented by BP neural networks and random forests. For instance, Gao et al. (2008)[4] propose a BP neural network-based method for forecasting the number of illegal border crossings.Most of these methods rely primarily on historical passenger flow data and do not adequately account for the effects of external factors such as calendar attributes, weather conditions, and immigration policies. They therefore have limited ability to capture nonlinear patterns in passenger flow,

which restricts forecasting accuracy. In addition, previous studies mainly focus on macro-level annual forecasts of inbound and outbound passenger flows, while relatively little attention is paid to fine-grained daily forecasting. As a result, existing approaches provide insufficient decision support for precise day-to-day checkpoint management. With the rapid transition toward smart border checkpoints, passenger flow forecasting should increasingly adopt artificial intelligence-based approaches. Deep learning models can be used to develop more accurate and operationally useful forecasting methods, thereby supporting checkpoint authorities in addressing the security challenges associated with growing cross-border passenger flows.

In recent years, gradient boosting algorithms (Qian et al., 2020)[5], long short-term memory (LSTM) models (Li et al., 2022)[6], and their improved variants (Cheng et al., 2022; Zhang et al., 2021; Zhang et al., 2022)[7][8][9] have been widely applied to passenger flow forecasting. In addition, Transformer models have achieved considerable success in forecasting tasks across tourism, transportation, hydrology, finance, and other fields because of their powerful attention mechanisms and sequence modeling capabilities (Kim et al., 2026; Meng & Shi, 2025; Wen et al., 2023; Yang, 2026)[10][11][12][13], providing useful methodological insights for forecasting inbound and outbound passenger flows at border checkpoints.

In tourism forecasting, Li et al. (2025)[14] propose a Transformer-based framework for predicting visitor flows at tourist attractions and employ the tree-structured Parzen estimator algorithm to perform single-step and multi-step forecasting for seven scenic areas. In the field of transportation, He et al. (2025)[15] develop a particle swarm optimization-based GRU-Transformer model to forecast short-term sectional passenger flows in urban rail transit by integrating multiple attribute-related factors. To address massive traffic flow datasets, Tao et al. (2026)[16] propose a stacked denoising and wavelet lightweight Transformer model, termed SDWLT, for traffic speed forecasting. Wang et al. (2020)[17] develop a traffic flow forecasting method based on a spatiotemporal graph neural network. The method first constructs a graph neural network with a positional attention mechanism to model spatial dependencies among road network nodes and then employs a gated recurrent unit and a Transformer to capture short-term and long-term temporal dependencies, respectively. Yan et al. (2022)[18] propose a traffic forecasting algorithm called T-former, which uses multi-head attention and stacked layers to learn dynamic and hierarchical patterns from sequential data. A global encoder and a global-local decoder are further employed to extract and fuse spatiotemporal features. Yu et al. (2025)[19] develop a combined CL-Transformer model to forecast arriving passenger flows at multiple types of external passenger transport hubs, demonstrating the broad applicability of the model. In hydrological runoff forecasting, Huang et al. (2026)[20] construct a hybrid Transformer-GRU model that integrates runoff-related factors to predict daily inflows at different cascade reservoirs, achieving satisfactory forecasting performance.

Although Transformer-based models and their hybrid variants have achieved strong performance in forecasting passenger flows for urban transportation systems such as metro and bus networks, inbound passenger flows at border checkpoints exhibit fundamentally different driving characteristics. First, they are strongly affected by external event shocks, including immigration policy adjustments and extreme weather events, rather than by the relatively continuous and stable fluctuations observed in urban transportation systems. Second, their periodic patterns are more complex because they reflect multiple calendar effects, such as public holidays and weekends across different cross-border regions, with varying levels of influence. Third, border checkpoint passenger flows exhibit greater sparsity and abrupt variation, including anomalously low observations close to zero, which imposes higher requirements on model robustness. Existing models are not specifically designed for these characteristics, and their direct transfer to border checkpoint scenarios may therefore result in substantial forecasting bias.

To address these limitations, this study proposes a hybrid model that integrates TCN and Transformer architectures and explicitly incorporates the coupled effects of calendar attributes, immigration policies, weather conditions, and other external factors for daily inbound passenger flow forecasting at border checkpoints. Compared with existing methods, the proposed model provides three main advantages. First, it constructs a multi-source daily feature dataset, overcoming the limitations of coarse temporal granularity and dependence on a single type of feature in previous border checkpoint passenger flow studies. Second, it develops a dual-branch parallel TCN-Transformer encoding architecture that jointly models and deeply integrates local passenger flow patterns and global temporal dependencies. Third, it is specifically adapted and optimized for the distinctive characteristics of border checkpoint passenger flows, thereby achieving higher forecasting accuracy and stronger generalization robustness in border checkpoint applications. Validation using data from the Hong Kong–Macau Ferry Terminal checkpoint demonstrates that the proposed hybrid model achieves strong predictive accuracy and generalization capability, providing technical support for scientific decision-making in practical applications such as checkpoint clearance management.

2. Development of the Inbound Passenger Flow Forecasting Model

Inbound passenger flow at border checkpoints is characterized by strong periodicity, pronounced nonlinearity, and complex interactions among multiple influencing factors. Traditional forecasting methods often have limited capacity to model such complex patterns. This study evaluates several deep learning models, including LSTM, CNN, and Transformer, and selects a hybrid TCN-Transformer model as the basic framework for systematically forecasting daily inbound passenger flow at border checkpoints.

2.1. TCN Model

The temporal convolutional network (TCN), introduced by Bai et al. (2018)[21] based on a one-dimensional convolutional architecture, is a temporal modeling network that performs reliably in extracting local temporal features and multiscale variations. Compared with conventional CNNs, TCN is better suited to time-series forecasting. By integrating causal convolution, dilated convolution, and residual connections, TCN preserves temporal causality while efficiently modeling long-range dependencies.

(1) Causal convolution strictly follows the chronological order of a time series. A unidirectional offset is applied to the convolution kernel to ensure that the output at the current time step depends only on historical observations and does not incorporate future information. It is expressed as follows:

$$y(t) = \sum_{i=0}^{k-1} W_i \times x(t-i) \quad (1)$$

Where $y(t)$d enotes the output at the current time step, $x(t\text{-}i)$denotes the input at a historical time step, W_i enotes the input at a historical time step.

(2) To capture long-term temporal dependencies, TCN employs dilated convolution, in which the convolution kernel processes the input sequence at specified intervals. This operation exponentially enlarges the receptive field while maintaining relatively low computational complexity. It is expressed as follows:

$$y(t) = \sum_{i=0}^{k-1} W_i \cdot x(t-d \times i) \quad (2)$$

where d is the dilation factor, which determines the temporal interval between the current time step and the historical observations involved in the convolution.

(3) Residual connections directly add the input of each convolutional block to its output, thereby forming a residual block. This structure helps alleviate the vanishing-gradient problem and facilitates the training of deeper networks. It is expressed as follows:

$$X' = ReLU\{BatchNorm[Conv1D(X)]\} + X \quad (3)$$

where X denotes the output sequence of the TCN $BatchNorm$ denotes the normalized input data

used to improve training stability, $ReLU$ denotes the activation function.

2.2. Transformer Model

The Transformer model is first proposed by Vaswani et al. (2017)[22] in 2017. Through self-attention and multi-head attention mechanisms, the model effectively captures long-range dependencies between arbitrary time steps in sequential data. Its parallel computational architecture also improves computational efficiency, giving the model substantial advantages in long-sequence forecasting tasks. Its main components are described below.

(1) The self-attention mechanism establishes relationships among different time steps within a sequence by dynamically assigning attention weights, thereby enabling global information modeling. Given an input sequence $X=(x_1,x_2,\cdots,x_n)$, the model transforms the original features into query *Q*, key *K*, and value *V* matrices through learnable linear projections:

$$Q=X'W_Q, K=X'W_K, V=X'W_V \tag{4}$$

Where W_Q, W_K, and W_V are learnable weight matrices. The self-attention output is then calculated as follows:

$$Attention(Q,K,V)=softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{5}$$

Where d_k denotes the dimensionality of the key vectors. The scaling factor $\sqrt{d_k}$ normalizes the dot products and prevents excessively large values from causing unstable gradients or adversely affecting model training.

(2) To improve the ability of the model to capture temporal dependencies at multiple scales, the Transformer employs a multi-head attention mechanism. This mechanism divides the feature space into multiple subspaces and independently performs attention calculations within each subspace, allowing the model to learn diverse temporal dependency patterns. The resulting representations are concatenated and linearly transformed to integrate information from multiple perspectives. The multi-head attention mechanism is expressed as follows:

$$MultiHead(Q,K,V)=Concat\left(head_1,head_2,\cdots,head_h\right)W_0 \tag{6}$$

$$head_i=Attention\left(QW_{Q_i},KW_{K_i},VW_{V_i}\right) \tag{7}$$

(3) Because the Transformer contains neither recurrent nor convolutional structures, it cannot inherently identify the sequential order of observations. Positional encoding is therefore introduced to explicitly incorporate temporal order information. A commonly used approach is sinusoidal positional encoding, which embeds positional information into the feature vectors through fixed sine and cosine functions. It is calculated as follows:

$$\begin{cases} PE_{(pos,2i)}=sin\left(\dfrac{pos}{10000^{2i/d}}\right) \\ PE_{(pos,2i+1)}=cos\left(\dfrac{pos}{10000^{2i/d}}\right) \end{cases} \tag{8}$$

where *pos* denotes the position index, *i* denotes the dimension index, and *d* denotes the dimensionality of the input representation.

（4）The standard Transformer adopts an encoder-decoder architecture. The encoder consists of multiple identical layers, each containing a self-attention module and a feed-forward neural network, which jointly extract global representations from the input sequence. The decoder extends the encoder structure by incorporating a masked self-attention module. This module prevents the model from accessing information from future time steps, thereby preserving temporal causality throughout the forecasting process.

2.3. Design of the TCN-Transformer Hybrid Model

Inbound passenger flow at border checkpoints typically exhibits two types of temporal patterns that are difficult to capture using a single model architecture. The first comprises short-term abrupt variations, including weekend peaks, weather

disruptions, temporary immigration policy adjustments, and sudden surges during public holidays. Capturing these patterns requires a model with a fine-grained local receptive field and a rapid response to changing features. The second comprises long-term temporal dependencies, including weekly, monthly, and quarterly cycles, persistent policy effects, and evolving passenger flow trends. Modeling these patterns requires strong global perception and the ability to capture long-range temporal relationships. Although TCN efficiently extracts local trends and detailed fluctuations through dilated causal convolutions and residual structures, its ability to model long-term dependencies across distant time periods remains limited. Conversely, the Transformer establishes global temporal relationships through self-attention, but it is less efficient in capturing local high-frequency fluctuations and is more susceptible to noise.

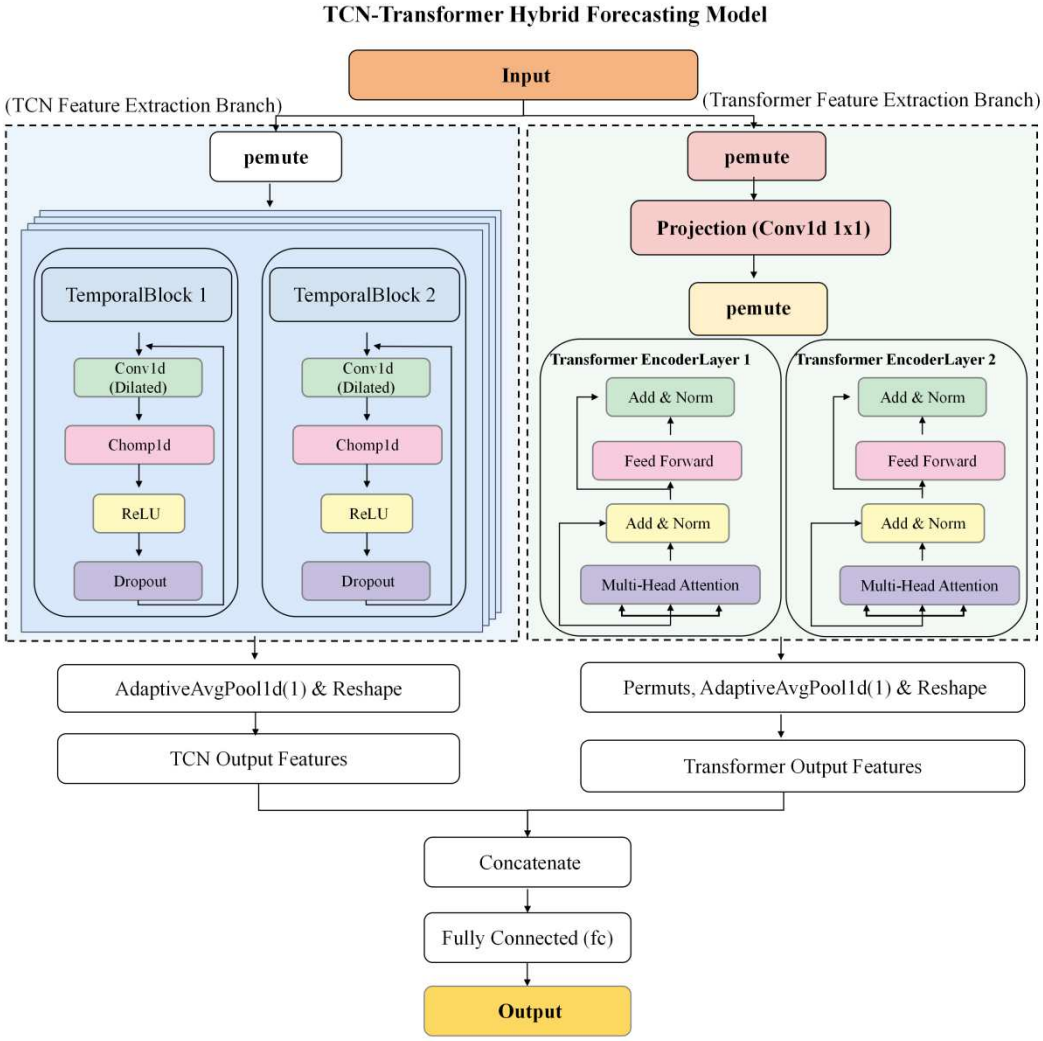


Figure 1. Architecture of the TCN-Transformer Hybrid Model

To characterize the complex temporal patterns of inbound passenger flow, including short-term abrupt variations, long-term periodicity, holiday-induced pulse effects, and external shocks caused by policy and weather conditions, this study develops a TCN-Transformer hybrid forecasting model that combines the local feature extraction capability of TCN with the global dependency modeling capability of the Transformer. The model adopts a dual-branch parallel encoding architecture. The TCN branch preserves its advantages in extracting local details, capturing short-term fluctuations, and responding

rapidly to abrupt changes, while the Transformer branch retains its strong capability to model global relationships, long-term dependencies, and multiscale trends. Through the deep fusion of fine-grained local features and global long-range dependencies, the proposed model is well suited to the distinctive characteristics of inbound passenger flow at border checkpoints, which include frequent short-term disruptions, complex long-term patterns, and strong interference from external factors. It therefore enables passenger flow forecasting with high accuracy and strong robustness. The overall architecture of the hybrid model is presented in **Figure 1.**

The hybrid model takes multi-source fused features as input, including historical inbound passenger flow, calendar attributes, immigration policies, weather conditions, and other multidimensional time-series information. These features are then processed synchronously by two parallel feature extraction branches. In the TCN branch, stacked causal convolutional residual blocks, referred to as Temporal Blocks, serve as the core units. Causal convolution ensures that the forecasting process strictly follows temporal order and prevents data leakage caused by the inclusion of future information. Dilated convolution progressively enlarges the receptive field across layers, enabling the model to capture multiscale short-term fluctuations without substantially increasing computational cost. Residual connections and Chomp1d cropping layers improve the stability of deep network training and alleviate the vanishing-gradient problem. This branch extracts fine-grained features, including short-term fluctuations, local trends, and high-frequency abrupt changes in passenger flow. It is particularly suitable for capturing sharp variations associated with weekend peaks, sudden weather changes, and temporary policy adjustments at border checkpoints.

In the Transformer branch, a 1×1 convolutional projection layer first aligns the feature dimensions, after which the resulting representations are passed through multiple Transformer encoder layers. Multi-head self-attention dynamically learns the association weights between different time steps and constructs global dependency representations. This mechanism enables the model to capture long-range temporal features, including weekly, monthly, and quarterly cycles, long-term trends, lagged policy effects, and cumulative holiday effects. By performing attention calculations in parallel across multiple representation subspaces, the branch identifies passenger flow evolution patterns from different perspectives and substantially improves the global representation of complex passenger flow sequences.

After the two branches generate global feature vectors with fixed dimensions, adaptive average pooling compresses the sequential representations and aligns their dimensions. Feature concatenation is then applied to deeply integrate the local features extracted by the TCN branch with the global features learned by the Transformer branch. Finally, the fused representation is mapped through a fully connected layer to produce the predicted inbound passenger flow.

3. Data Description

3.1. Data Sources

This study uses inbound passenger flow at the Hong Kong-Macau Ferry Terminal checkpoint as the empirical case. The dataset is obtained from the Hong Kong Open Data Portal and covers three consecutive years from March 1, 2023, to March 1, 2026. It contains four categories of daily data, including inbound passenger flow, calendar attributes, immigration policies, and weather conditions.

Different calendar types exert varying effects on inbound passenger flow at border checkpoints. In this study, calendar attributes are classified into public holidays, weekends, and weekdays. According to their relative influence on inbound passenger flow, these categories are ordinally encoded as 2, 1, and 0, respectively, allowing the model to identify and learn differences across calendar types.

Inbound passenger flow is also affected by immigration policies. Restrictive policies are encoded as -1, policies without a clear directional effect are encoded as 0, and facilitative policies are encoded as 1. The classification criteria and corresponding codes for immigration policies are presented in **Table 1**.

Table 1 Coding Scheme for Immigration Policies

Immigration policy type	Code	Description
Restrictive	-1	The checkpoint is temporarily closed or passenger clearance is restricted because of maintenance, major events, or other circumstances.
Neutral	0	Normal clearance operations
Facilitative	1	Facilitative measures, such as expanding the pilot scope of talent visas for Hong Kong and Macau, implementing multiple-entry endorsements, and introducing facial recognition clearance, are adopted to promote passenger movement through the checkpoint.

Weather conditions represent another important determinant of inbound passenger flow. Four weather indicators with relatively strong effects on checkpoint passenger flow are selected, namely rainfall, mean temperature, mean wind speed, and low-visibility duration. Low visibility refers to visibility below 8 km and is an important safety-related indicator in aviation, maritime transport, and other transportation sectors. To enable the model to more accurately capture the effects of weather conditions on inbound passenger flow, the original values of all four weather indicators are retained. A sample of the raw data for the Hong Kong-Macau Ferry Terminal checkpoint is presented in **Table 2.**

Table 2. Sample of the Raw Dataset

Date	Inbound passenger flow (passengers)	Calendar code	Policy code	Rainfall (mm)	Mean temperature (°C)	Mean wind speed (km/h)	Low-visibility duration (h)
2023-03-01	6893	0	0	0	19.7	8.3	0
2023-03-02	6248	0	0	0	19.8	17.3	0
2023-03-03	7233	0	0	0	18.6	18.7	0
2023-03-04	7502	1	0	0	19.6	12.8	0
2023-03-05	10102	1	0	0	19.7	16.5	0
…	…	…	…	…	…	…	…
2026-02-27	9352	0	1	0.3	21.4	13.1	0
2026-02-28	9660	1	1	39	20	28.3	0
2026-03-01	10842	1	1	0.1	20	17.4	0

3.2. Analysis of Inbound Passenger Flow Characteristics

The temporal trend and distributional characteristics of inbound passenger flow at the Hong Kong-Macau Ferry Terminal checkpoint are presented in **Figure 2** and **Figure 3**, respectively. As shown in **Figure 2**, the inbound passenger flow series exhibits clear periodic fluctuations and a relatively stable short-term trend, with pronounced peaks on weekends and public holidays. In addition, the series contains a small number of extreme low-value observations close to zero. These values are mainly attributable to temporary control measures implemented in response to extreme weather events, such as typhoons and heavy rainfall, as well as other special incidents. They therefore represent occasional disturbances rather than systematic trends, and their interference with model training remains manageable.

As shown in **Figure 3**, the mean inbound passenger flow at this checkpoint is 9,211 passengers, while the median is 8,805 passengers. The close proximity between these two values indicates that the overall passenger flow distribution is approximately symmetric and unimodal. On most days, inbound passenger flow is concentrated within the range of

4,272 to 14,244 passengers. The sample standard deviation is 2,216 passengers, which is substantially lower than the mean, suggesting that the magnitude of passenger flow variation is relatively limited. The boxplot also shows only a very small number of outliers beyond the upper and lower whiskers, indicating strong series stability and generally high data quality.

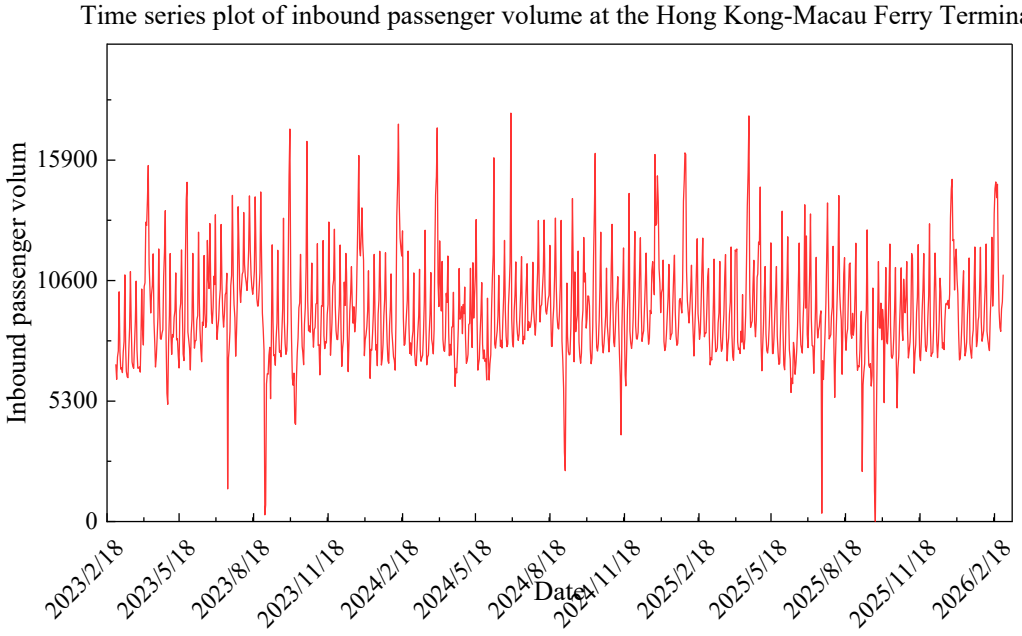


Figure 2. Time Series of Inbound Passenger Flow at the Hong Kong-Macau Ferry Terminal Checkpoint

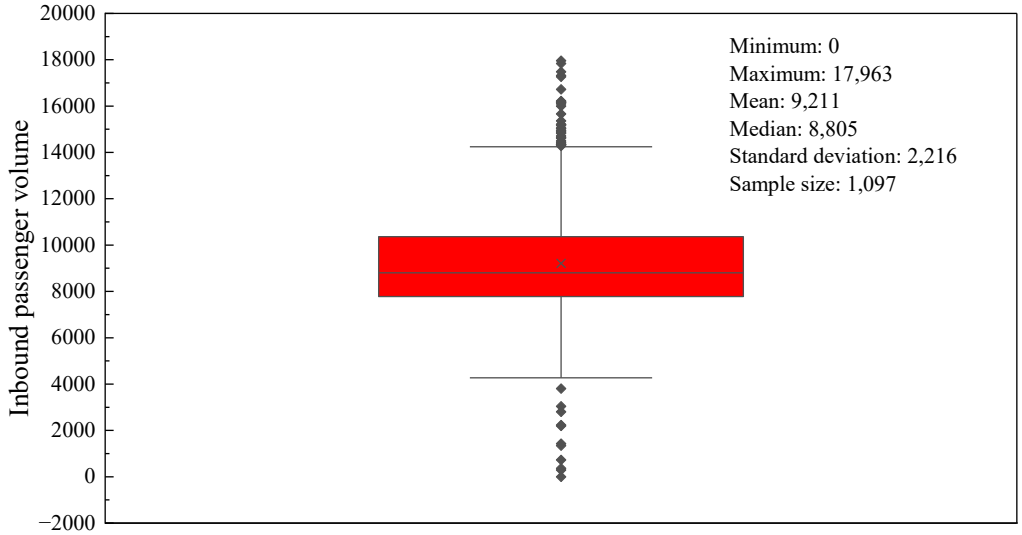


Figure 3. Boxplot of Inbound Passenger Flow at the Hong Kong–Macau Ferry Terminal Checkpoint

3.3. Data Preprocessing

3.3.1. Outlier Treatment

Extreme low-value observations close to zero are identified and corrected using the three-sigma rule. Assuming that the inbound passenger flow series approximately follows a normal distribution, the mean μ and standard deviation σ are calculated, and observations satisfying $x<\mu-3\sigma$ are classified as outliers. The resulting threshold is 2,563 passengers, and observations below this value are therefore treated as anomalous. These outliers are replaced using a three-day moving average based on adjacent observations. This procedure reduces the interference of extreme values and noise during model training while preserving the original temporal trend and fluctuation patterns of the passenger flow series to the

greatest extent possible, thereby improving data quality and modeling stability.

3.3.2. Data Standardization

To eliminate the effects of differences in scale among variables, Z-score standardization is separately applied to the input features and target variable as follows:

$$X' = \frac{x - \mu}{\sigma} \tag{9}$$

where μ and σ denote the mean and standard deviation of the corresponding feature or target variable, respectively.

Continuous variables, including historical passenger flow, rainfall, mean temperature, mean wind speed, and low-visibility duration, are standardized to have a mean of 0 and a standard deviation of 1, which facilitates model convergence. The standardization parameters are estimated exclusively from the training set. The test set is then transformed using the mean and standard deviation derived from the training set, thereby preventing information leakage and ensuring the scientific rigor and reproducibility of the experiments.

3.3.3. Dataset Partitioning

The dataset is divided into training and test sets in strict chronological order using an 8:2 ratio. The resulting training set contains 878 days, while the test set contains 219 days. This time-ordered partitioning prevents future information from leaking into the model training process.

3.4. Model Evaluation

To evaluate the forecasting accuracy of the proposed model for inbound passenger flow at border checkpoints, four performance metrics are employed: mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination (R^2). These metrics assess model accuracy and robustness from different perspectives. Their calculation formulas are as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \tag{10}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \tag{12}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2} \tag{13}$$

Where n where n denotes the number of samples, y_i denotes the observed passenger flow for the i-th sample, $\hat{y}_i$ denotes the predicted passenger flow for the i-th sample, and $\bar{y}$ denotes the mean of the observed passenger flow values. Lower MSE, RMSE, and MAE values indicate smaller forecasting errors, whereas an R^2 value closer to 1 indicates a stronger goodness of fit and better explanatory performance.

4. Experimental Design and Result Analysis

4.1 Experimental Setup and Parameter Configuration

To ensure the reproducibility and stability of experimental results, all model training, data processing and comparative experiments in this paper are conducted under identical hardware and software environments. The hardware environment adopts a portable computing device equipped with an Intel(R) Core(TM) i5-8265U processor running at 1.60 GHz, 8 GB of RAM and 512 GB of NVMe SSD local storage. The graphics configuration includes an NVIDIA GeForce MX250 discrete graphics card and an Intel(R) UHD Graphics 620 integrated graphics card. The software environment runs on a 64-bit Windows operating system, and a deep learning framework is built based on the Python programming language to accomplish data preprocessing, model construction, training and validation, and result visualization. On this basis, key parameters of the proposed TCN-Transformer hybrid prediction model are specified to achieve optimal feature extraction and

prediction performance. The detailed configuration of model hyperparameters is shown in **Table 3**.

Table 3 Model parameter configuration

Parameter	Value	Remarks
TCN network channel dimension	[32, 64]	Two residual blocks are used, with channel numbers set to 32 and 64 sequentially.
Causal convolution kernel size	3	—
Dimension of Transformer feed-forward hidden layer	64	—
Number of multi-head attention heads	2	—
Number of Transformer encoder layers	2	—
Temporal sliding window length	12	—
Learning rate of Adam optimizer	5×10^{-4}	—
Epochs	100	—
Batch size	16	—
Dropout rate	0.2	Random dropout probability applied after convolutional layers in the TCN module.
Dilation factor	2^n	The dilation rate of each TCN layer increases in powers of 2 to enlarge the receptive field.

4.2. Analysis and Evaluation of Experimental Results

To evaluate model performance, four widely used models, namely TCN (Hewage et al., 2020; Jiang et al., 2019; Wan et al., 2019)[23][24][25], Transformer (Gao et al., 2023)[26], LSTM (Wu et al., 2020)[27]and CNN (Yang et al., 2019)[28], are selected for comparison with the TCN-Transformer model proposed in this paper. All comparative models utilize the full set of features and maintain consistent structures and parameters. The prediction performance of each model on the test set is illustrated in **Figure 4**. In terms of overall fitting effect, the prediction curve of the proposed TCN-Transformer model is highly consistent with the actual passenger flow time series over the entire test period. The model can accurately reproduce periodic fluctuations of passenger flow, holiday peaks, and abrupt characteristics driven by exogenous shocks such as policy adjustments and weather conditions, without significant phase lag or systematic bias, demonstrating outstanding performance in temporal feature capture and dynamic tracking. By contrast, each baseline model presents obvious prediction limitations. The CNN model generates highly volatile prediction results and suffers from severe systematic overestimation of passenger flow peaks, failing to effectively characterize the inherent temporal patterns of passenger flow. The LSTM model can grasp the overall trend, but exhibits phase lag in capturing peak and valley values, accompanied by an excessively smoothed prediction curve and loss of local detailed features. The single TCN model achieves preferable fitting performance for short-term fluctuations, but is insufficient in capturing long-range dependencies and shows obvious underestimation for high-amplitude passenger flow peaks. The single Transformer model can reproduce the overall periodic pattern, but has weak tracking ability for local abrupt changes and stable deviations in peak and valley prediction, with overall robustness inferior to that of the TCN-Transformer model.

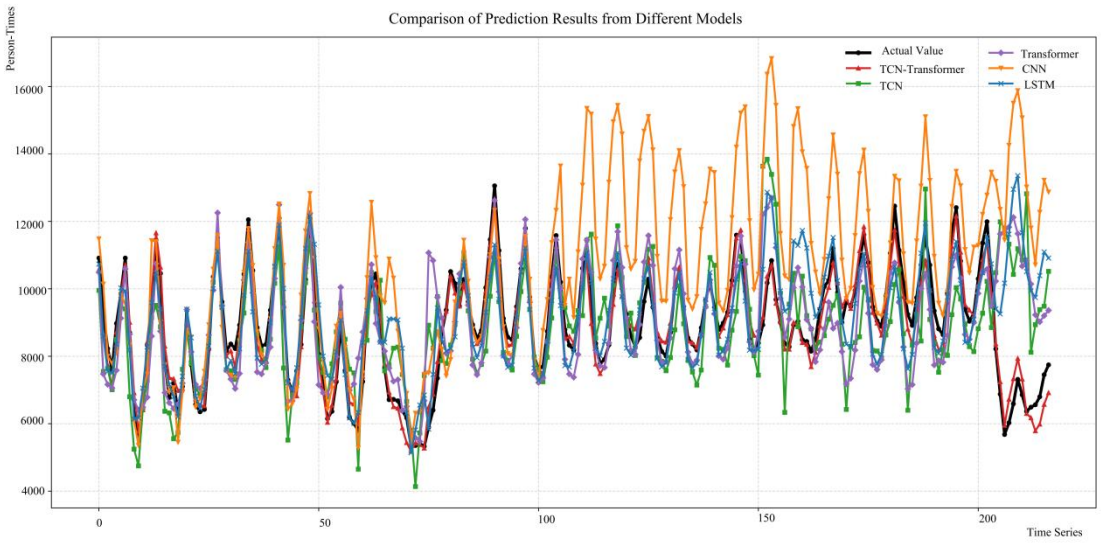


Figure 4. Comparison of Forecasting Results across Different Models

On the basis of the cross-model comparison, further attention is devoted to the inherent prediction performance of the TCN-Transformer model, and an experiment is carried out on the fitting curve of actual and predicted inbound passenger flow values in the test set, with the results presented in **Figure 5.** As shown in the figure, the constructed TCN-Transformer hybrid model achieves generally outstanding prediction performance for port inbound passenger flow. The prediction curve maintains a high degree of consistency with the actual passenger flow curve in terms of temporal trend, fluctuation morphology and peak-valley distribution, and can accurately reproduce the dynamic characteristics of passenger flow driven by weekly variations, holiday effects, as well as policy and weather disturbances. Meanwhile, the model possesses strong capacity to capture passenger flow peaks and valleys, with no evident phase lag, systematic overestimation or underestimation, and exhibits favorable fitting and tracking performance for both high-frequency fluctuations and local abrupt changes. The 95% confidence interval in the figure effectively covers the vast majority of actual observations, which indicates that the model prediction results have reliable statistical significance and uncertainty representation capability, and the prediction errors are dominated by random residuals and remain at a low level on the whole.

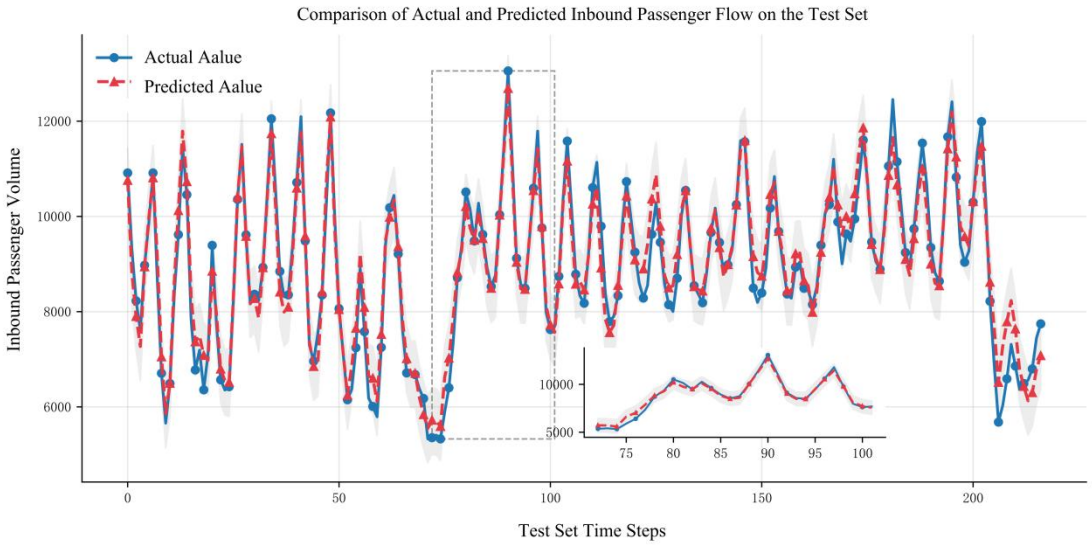


Figure 5. Fitted Curves on the Test Sets

To further quantitatively verify the overall fitting accuracy and generalization ability of the model, scatter plots of the fitting effect for the training and test sets are drawn, with the results shown in **Figure 6.** As can be seen from the figure, samples from both the training set and the test set are tightly distributed around the y=x diagonal line, with no obvious systematic deviation or heteroscedasticity, indicating that the model has favorable linear consistency in predicting passenger flow magnitude. The training set achieves a coefficient of determination R^2 of 0.969, a root mean squared error of 310.699, and a mean absolute error of 242.558. For the test set, the coefficient of determination R^2 is 0.953, the root mean squared error is 357.93, and the mean absolute error is 274.27. Although the test set errors are slightly higher than those of the training set, the overall difference is small. This result shows that the model fully learns the historical patterns of passenger flow without overfitting, and still possesses stable and reliable prediction capability on unseen new samples, with strong overall robustness.

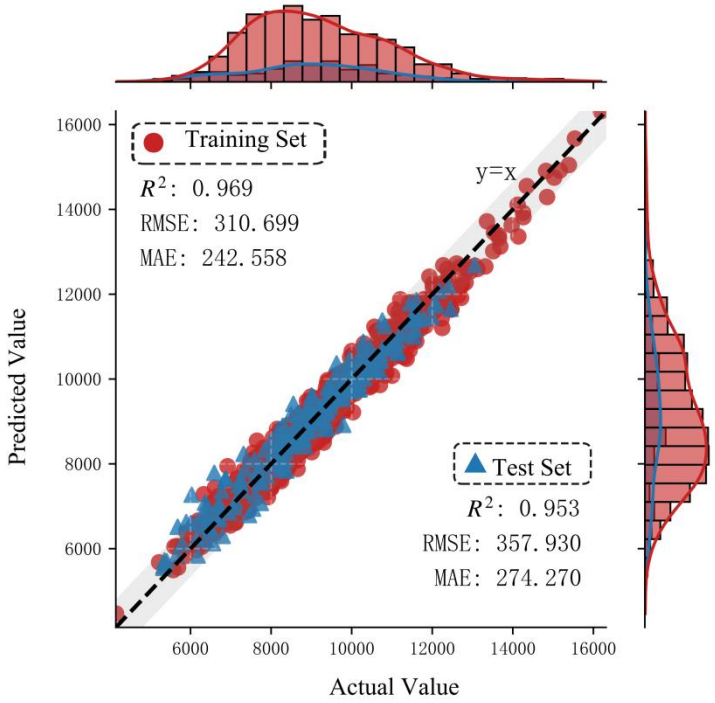


Figure 6. Scatter Plot of Observed and Predicted Values

Overall, the TCN-Transformer model combines the strength of TCN in local temporal feature extraction and the ability of Transformer to model long-range dependencies. It maintains high prediction accuracy across low, medium and peak passenger flow ranges, and achieves balanced fitting performance for both extreme peaks and normal fluctuation intervals. The model can effectively characterize the complex variation patterns of port inbound passenger flow driven by multiple factors such as holidays, policies and weather conditions, and its overall fitting performance meets the high-precision requirements of time series prediction.

4.3. Hyperparameter Analysis

To clarify the influence of various key parameters on the prediction performance of the TCN-Transformer model, this paper conducts a hyperparameter sensitivity analysis. The experiment adopts the univariate control method, which means only one hyperparameter is adjusted at a time while the rest remain unchanged. By observing the variation trends of two error metrics, MAE and MSE, on the test set, the influence patterns of different parameters on the prediction accuracy and stability of the model are identified. This study selects six core hyperparameters for analysis from two dimensions: model structure and training mechanism, namely

channel dimension, number of encoder layers, number of attention heads, feed-forward network dimension, learning rate and temporal window length. Under the premise of variable control, the value of each hyperparameter is adjusted one by one and the corresponding prediction results are recorded. Finally, the relationship between parameter changes and model performance is presented by means of visualization, with detailed results shown in **Figure 7.**

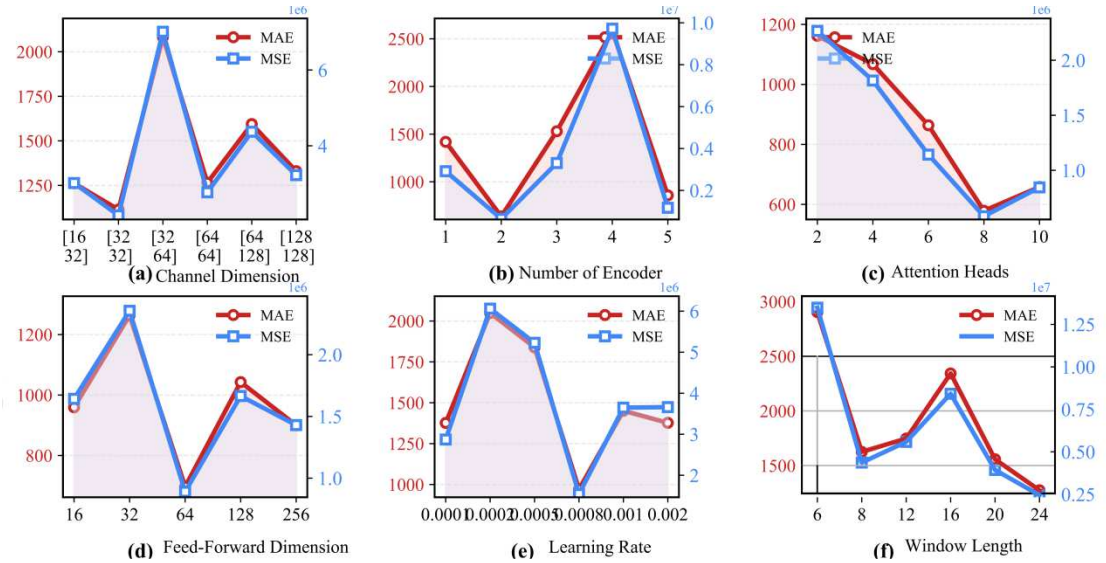


Figure 7. Hyperparameter Analysis

As observed from **Figure 7**, all hyperparameters exert significant effects on model performance in terms of overall variation trends, and the error curves follow a typical pattern of decreasing first and then increasing, which indicates the existence of an optimal interval rather than a monotonic relationship. In subfigure (a), the error is the lowest when the channel number is set to [32, 32], followed by the [64, 64] configuration. A moderate increase in channel number can enhance feature extraction capability, whereas excessively high dimensionality leads to overfitting and rising errors. In subfigure (b), the model achieves the best prediction performance with 2 encoder layers. Insufficient layers result in inadequate feature extraction, while an excessive number of layers causes model redundancy and gradient degradation. In subfigure (c), the prediction error reaches its minimum when the number of attention heads is 8. Too few heads fail to capture multi-scale temporal dependencies, whereas an excessive number may cause attention dispersion and dimensional fragmentation. In subfigure (d), the model converges most stably when the feed-forward dimension ranges from 64 to 128. An excessively small dimension leads to insufficient representation capacity, while an overly large dimension increases computational redundancy. In subfigure (e), the optimal convergence state is achieved at a learning rate around 8×10^{-4}. An excessively high learning rate causes training oscillation and hinders convergence, while an overly low learning rate results in slow convergence and a tendency to fall into local optima. In subfigure (f), the error is minimized when the temporal window length is approximately 24 days, followed by 8 days. This pattern is highly consistent with the periodic characteristics of port inbound passenger flow. An excessively short window loses historical information, while an overly long window introduces redundant noise.

The visualized results of hyperparameter sensitivity provide a basis for the reasonable determination of parameters in this study. Targeted adjustment of parameter settings during model training can mitigate overfitting or underfitting caused by improper parameter selection, thereby effectively improving the prediction accuracy and robustness of the model.

5. Conclusion

To address the issues that annual coarse-grained data can hardly satisfy the demand for refined management and control and the coupled influences of multi-source external factors are insufficiently characterized in the daily prediction of port inbound passenger flow, this paper proposes a TCN-Transformer hybrid prediction model and draws the following conclusions.

(1) a multi-source feature dataset integrating historical daily inbound passenger flow, date attributes, entry-exit policies and weather conditions is constructed, which makes up for the limitation that existing studies mainly rely on single historical passenger flow data.

(2) Second, the proposed model extracts local short-term fluctuation features through the TCN branch and captures long-term periodic dependencies via the Transformer branch. Through adaptive feature fusion, the model can effectively decouple the dual characteristics of port passenger flow, namely frequent short-term abrupt changes and complex long-term patterns. On the dataset of the Hong Kong-Macau Ferry Terminal, the model achieves a coefficient of determination R^2 of 0.953, a root mean squared error of 357.93 and a mean absolute error of 274.27 on the test set, with outstanding fitting accuracy and generalization performance.

(3) Third, compared with existing prediction methods based on LSTM, single TCN or single Transformer, the proposed model exhibits more stable prediction performance in typical scenarios including weekend peaks, holiday passenger surges and policy adjustment periods.

(4) Fourth, this study introduces the Transformer hybrid model into the field of port passenger flow forecasting for the first time, and verifies its applicability in multi-source heterogeneous feature fusion and complex time series modeling. Follow-up research can further explore the application of graph neural networks in multi-port collaborative prediction, as well as transfer learning methods targeting the data cold start problem of new ports.

References:

[1] Wu, C. F. (2011). Application of triple exponential smoothing in forecasting the number of inbound and outbound passengers. Journal of Langfang Teachers University (Natural Science Edition), 11(1), 13–14. (in Chinese)

[2] Yu, L. (2016). Forecasting of freight volume and traffic volume at Heihe highway port. Economic Research Guide, (2), 54–56. (in Chinese)

[3] Chi, D. Z., & Zhang, Y. (2013). A study on forecasting methods for cross-border passenger transport demand at Zhuhai–Macau ports. Journal of Highway and Transportation Research and Development (Applied Technology Edition), 9(5), 248–251. (in Chinese)

[4] Gao, H. K., Wang, C., Zhang, J., et al. (2008). Forecasting the development trend of current illegal entry and exit activities based on an improved BP neural network. Journal of Computer Applications, (S1), 330–332. (in Chinese)

[5] Qian, H., Yang, Z., Weng, J., et al. (2020). Short-term passenger flow prediction of metro stations around sports events based on AFC data. In Proceedings of the 20th COTA International Conference of Transportation Professionals (pp. 70–73).

[6] Li, Q., Wang, B., Jin, J., Zhang, T., & Wang, X. (2022). Automatic sleep staging model based on the bi-directional LSTM convolutional network and attention mechanism. CAAI Transactions on Intelligent Systems, 17(3), 523–530. https://doi.org/10.11992/tis.202103013 (in Chinese)

[7] Cheng, Z., Lu, J., Zhou, H., et al. (2022). Short-term traffic flow prediction: An integrated method of econometrics and hybrid deep learning. IEEE Transactions on Intelligent Transportation Systems, 23(6), 5231–5244.

[8] Zhang, J., Che, H., Chen, F., Ma, W., & He, Z. (2021). Short-term origin–destination demand prediction in urban rail transit systems: A channel-wise attentive split-convolutional neural network method. Transportation Research Part C: Emerging Technologies, 124, 102928. https://doi.org/10.1016/j.trc.2020.102928

[9] Zhang, X., Zhong, S., & Mahadevan, S. (2022). Airport surface movement prediction and safety assessment with spatial–temporal graph convolutional neural network. Transportation Research Part C: Emerging Technologies, 144, 103873.
[10] Kim, S., Lee, H. J., & Rey, C. S. (2026). A convolutional neural network–Transformer denoiser for low-signal-to-noise-ratio galaxy spectra: Stellar population recovery in synthetic tests. The Astronomical Journal, 171(6), Article 325.
[11] Meng, X. F., & Shi, H. Y. (2025). A survey of Transformer-based time-series forecasting methods. Journal of Frontiers of Computer Science and Technology, 19(1), 45–64. (in Chinese)
[12] Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., & Sun, L. (2023). Transformers in time series: A survey. In Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence (pp. 6778–6786).
[13] Yang, J. (2026). Expert Transformer: Fine-grained mixture-of-experts for time-series forecasting. Expert Systems with Applications, 324, 132634.
https://doi.org/10.1016/j.eswa.2026.132634
[14] Li, X., Zhang, X., Yu, L. A., et al. (2025). Short-term tourist flow forecasting for scenic areas based on an improved Transformer model. Chinese Journal of Management Science, 33(2), 105–117. (in Chinese)
[15] He, M., Zuo, J., & Yu, Z. (2025). Short-term sectional passenger flow forecasting for urban rail transit considering the fusion of multiple attribute factors. Journal of Shenzhen University (Science and Engineering), 42(6), 723–732. (in Chinese)
[16] Tao, Z., Feng, H., & Guo, R. (2026). Traffic speed prediction based on a stacked denoising and wavelet lightweight Transformer. Engineering Applications of Artificial Intelligence, 174, 114547.
[17] Wang, X., Ma, Y., Wang, Y., Jin, W., Wang, X., Tang, J., Jia, C., & Yu, J. (2020). Traffic flow prediction via spatial temporal graph neural network. In Proceedings of The Web Conference 2020 (pp. 1082–1092). Association for Computing Machinery.
https://doi.org/10.1145/3366423.3380186
[18] Yan, H., Ma, X., & Pu, Z. (2022). Learning dynamic and hierarchical traffic spatiotemporal features with Transformer. IEEE Transactions on Intelligent Transportation Systems, 23(11), 22386–22399.
[19] Yu, J. B., Weng, J. C., Lin, P. F., et al. (2025). Passenger arrival flow forecasting model for external passenger transport hubs based on a hybrid Transformer. Journal of Jilin University (Engineering and Technology Edition), 55(7), 2251–2259. (in Chinese)
[20] Huang, D., Liu, X., Ma, A. G., et al. (2026). Daily runoff forecasting for cascade reservoirs based on a Transformer–GRU model. China Rural Water and Hydropower, 1–17. Advance online publication. (in Chinese)
[21] Bai, S., Kolter, J. Z., & Koltun, V. (2018). Convolutional sequence modeling revisited. In Proceedings of the 6th International Conference on Learning Representations: Workshop Track.
[22] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In I. Guyon et al. (Eds.), Advances in Neural Information Processing Systems (Vol. 30, pp. 5998–6008). Curran Associates, Inc.
[23] Hewage, P., Behera, A., Trovati, M., Pereira, E., Ghahremani, M., Palmieri, F., & Liu, Y. (2020). Temporal convolutional neural network for effective weather forecasting using time-series data from the local weather station. Soft Computing, 24, 16453–16482.
[24] Jiang, W., Chen, J., & Chen, Y. (2019). Remaining useful life prediction for mechanical equipment based on temporal convolutional network. In Proceedings of the 14th IEEE International Conference on Electronic Measurement & Instruments (pp. 1192–1199). IEEE.
[25] Wan, R., Mei, S., Wang, J., Liu, M., & Yang, F. (2019). Multivariate temporal convolutional network: A deep neural networks approach for

multivariate time series forecasting. Electronics, 8(8), 876. https://doi.org/10.3390/electronics8080876
[26] Gao, R., Wan, Y. L., Shao, X. K., et al. (2023). Traffic flow forecasting model based on an improved spatiotemporal Transformer. Computer Engineering and Applications, 59(7), 250–260. (in Chinese)
[27] Wu, J., Hu, K., Cheng, Y., Zhu, H., Shao, X., & Wang, Y. (2020). Data-driven remaining useful life prediction via multiple sensor signals and deep long short-term memory neural network. ISA Transactions, 97, 241–250.
[28] Yang, B., Liu, R., & Zio, E. (2019). Remaining useful life prediction based on a double-convolutional neural network architecture. IEEE Transactions on Industrial Electronics, 66(12), 9521–9530.

Funding: This research is supported by the Fundamental Research Funds for the Central Universities ∗Grant Nos. 2025bsky007 and 2025bsky019) and the Research Program of the China-ASEAN ∗Guangxi) Institute of Police Studies ∗Grant No. DMJXZX202505).

Author Biographies: Cen Shijie ∗born in 1996), male, is from Guiping, Guangxi. He is currently a doctoral candidate, and his research interests include immigration administration and artificial intelligence.Email: 2925618373@qq.com.

Corresponding author: Lu Jing ∗born in 1972), female, is from Changchun, Jilin. She holds a doctoral degree and is a doctoral supervisor. Her research focuses on migration management. Email: 718754118@qq.com.

Lai Zhijian ∗born in 1995), male, is from Ganzhou, Jiangxi. He is currently a doctoral candidate and lecturer, and his research interests include statistical analysis and data modeling.

Huang Wei ∗born in 1998), male, is from Hezhou, Guangxi. He is currently a masters student, and his research focuses on criminal investigation.