sustainability

MDPI

Article

Multi-Model Attention Fusion Multilayer Perceptron Prediction Method for Subway OD Passenger Flow under COVID-19

Yi Cao and Xue Li *

School of Transportation Engineering, Dalian Jiaotong University, Dalian 116028, China
* Correspondence: li2394058728@djtu.edu.cn; Tel.: +86-159-4114-9425

Abstract: At present, machine learning has been successfully applied in many fields and has achieved amazing results. Meanwhile, over the past few years, the pandemic has transformed the transportation industry. The two hot issues prompt us to rethink the traditional problem of passenger flow forecasting. As a special structure embedded in the machine learning model, the attention mechanism is used to automatically learn and calculate the contribution degree of input data to output data. Therefore, this paper uses the attention mechanism to find the best model to predict OD passenger flow under COVID-19. Holiday characteristics, minimum temperature, COVID-19 factors, and past origin-destination (OD) passenger flow were used as input characteristics. In the first stage, the attention mechanism was used to capture the advantages of the trained random forest, extreme gradient boosting (XGBoost), gradient boosting decision tree (GBDT), and Adaboost models, and then the MLP was trained. Afterward, the weight distribution of the two models is carried out by using the historical passenger flow. The multi-model attention+ MLP model was used to evaluate the OD passenger flow prediction of Dalian Metro Line 1 under COVID-19. All the possible choices in this process were taken as a comparison experiment. The results show that only the fusion model combining the attention mechanism of random forest and XGBoost with MLP has the highest prediction accuracy.

Keywords: COVID-19; passenger flow forecast; MLP; attention mechanism; ensemble algorithm

check for updates

Citation: Cao, Y.; Li, X. Multi-Model Attention Fusion Multilayer Perceptron Prediction Method for Subway OD Passenger Flow under COVID-19. *Sustainability* **2022**, *14*, 14420. https://doi.org/10.3390/su142114420

Academic Editor: Elżbieta Macioszek

Received: 6 October 2022
Accepted: 2 November 2022
Published: 3 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

As an important part of the public transportation system, the subway system is favored by more and more residents for its convenience, speed, green, and environmental protection. Therefore, urban rail transit passenger flow prediction is an important part of urban management, especially in subway cities, which can reasonably arrange subway operation plan to avoid energy waste.

Since 2019, with the outbreak of COVID-19, the travel mode of residents has changed [1], and residents are no longer willing to travel by public transport. Due to the uncertainty of the epidemic, the randomness of passenger flow, and the imbalance of the impact of COVID-19, the prediction of subway OD passenger flow is more complicated than that in the normal period, which brings difficulties to the prediction of passenger flow.

The statistical method is a traditional forecasting method, which combines a certain mathematical distribution with the changes in passenger flow, and takes historical data as the reference to calibrate model parameters, including the time series prediction method and regression prediction method. For example, Meng et al. [2] segment and model historical passenger flow data at the same time. The moving average method was used to predict the passenger flow at the same time interval, and the real-time passenger flow was used for correction and testing. Kang et al. [3] used Gaussian process regression to predict short-term traffic flow.

However, statistical methods have only a great impact on passenger flow prediction with a certain mathematical distribution, so they require high regularity of data distribution,

while machine learning trains models with its own characteristics, so it is universal despite the generally low accuracy of prediction. Mohammed and Kianfar [4] applied a deep neural network, random forest, gradient elevator, and generalized linear model to the passenger flow prediction of Interstate 64 in St. Louis, MO, USA. The results showed that the results obtained by these four prediction methods were very similar. However, the random forest model performs slightly better than the models obtained by the other three methods. Sun et al. [5] used the data of Xi'an Metro Line 2 to prove that the prediction junction effect of XGBoost is better than that of the back propagation neural network model and ARMA model. Huang et al. [6] used three integration methods, including Adaboost, to integrate five models: long short-term memory (LSTM), linear regression, K-nearest neighbor (KNN), XGBoost and gate recursive unit (GRU), and then predicted the single bus trip time. Xu et al. [7] realized short-term passenger flow prediction by establishing the GBDT model and proved that the model was superior to the linear model and back propagation neural network. In 2019, an artificial neural network trained by Gallo et al. [8] using simulated data obtained from the dynamic loading process of the railway line achieved suitable prediction results in the subway system. Yang et al. [9] integrated the long-term dependence characteristics and short-term characteristics of passenger flow in passenger flow data, thus overcoming the limitations caused by the time lag in previous forecasting work. However, they did not take into account other factors, such as weather, location, and information about stations and passengers. Li et al. [10] took seven weather factors as external influencing factors, screened five internal influencing factors from historical data, and proposed the LSTM deep network prediction model. Li et al. [11] proposed a novel multi-scale radial basis function network to predict the passenger flow of abnormal before 30 min. However, passenger flow is unpredictable under COVID-19, and previous models of prediction do not perform well in current forecasts.

In recent years, with the rise of attention mechanisms, it has been applied to traffic prediction by many scholars. Inspired by human visual mechanisms, attention mechanisms are widely used in data prediction to focus selectively on important parts. For example, GUO S. et al. [12] combined spatio-temporal convolutional network and attention mechanism to predict traffic flow. Wang et al. [13] coded AFC information and proposed a real-time passenger flow prediction algorithm based on the attention mechanism.

Statistical methods and machine learning methods have their own strengths, so in order to predict passenger flow more accurately, many scholars combine two or three prediction models. Firstly, Bai [14] used autoregressive integrated moving average (ARIMA) to predict subway passenger flow under normal conditions, then analyzed the influencing factors of abnormal traffic and proposed a combined model of ARIMA and multiple regression to predict the flow under abnormal conditions. Sun et al. [15] used the wavelet method to decompose passenger flow into sequences of different frequencies and then used the support vector machine (SVM) model to predict the passenger flow of different sequences. Zhao et al. [16] proposed a prediction method combining empirical mode decomposition and LSTM for short-term passenger flow prediction.

Most of the existing studies used different forecasting methods to study similar factors to predict passenger flow without comprehensively considering the impact of COVID-19 on residents' travel. Practice shows that when there are confirmed cases in the surrounding environment, residents' dependence on public transport is reduced, and travel for the purpose of entertainment is significantly reduced, leading to a decrease in passenger flow [17,18]. When the surrounding environment is safe, passenger flow gradually resumes. Machine learning is flexible and can handle complex data, so it is very suitable for the complexity of the subway OD passenger flow under COVID-19. Random forest, logistic regression (LR), classification and regression trees, XGBoost, Adaboost, light gradient boosting machine (LightBGM), GBDT, and MLP were used to predict OD passenger flow under COVID-19. The results showed that MLP had the best prediction effect, followed by random forest, XGBoost, Adaboost, and GBDT. However, through the analysis of the results, it is found that different single models have different performance effects on

different types of OD prediction. Therefore, this paper proposes an integrated model of attention mechanism and MLP according to the influencing factors such as COVID-19, weather, and holidays, in which random forest, Adaboost, GBDT, and XGBoost are taken as four sub-models of attention mechanism. All possibilities in the prediction process were used as comparative tests. The results show that the multi-model attention+ MLP model still has better predictive performance.

The remainder of this article is as follows. Section 2 is data collation, station classification, and characteristics of passenger flow of different types of stations under COVID-19. Section 3 is to introduce the basic model. Section 4 is the construction of the multi-model attention+ MLP model. In Section 5, we find the most suitable multi-model attention+ MLP model for OD passenger flow prediction under COVID-19 through cases. Finally, Sections 6 and 7 present the discussions and conclusions of this paper. The article structure is shown in Figure 1.

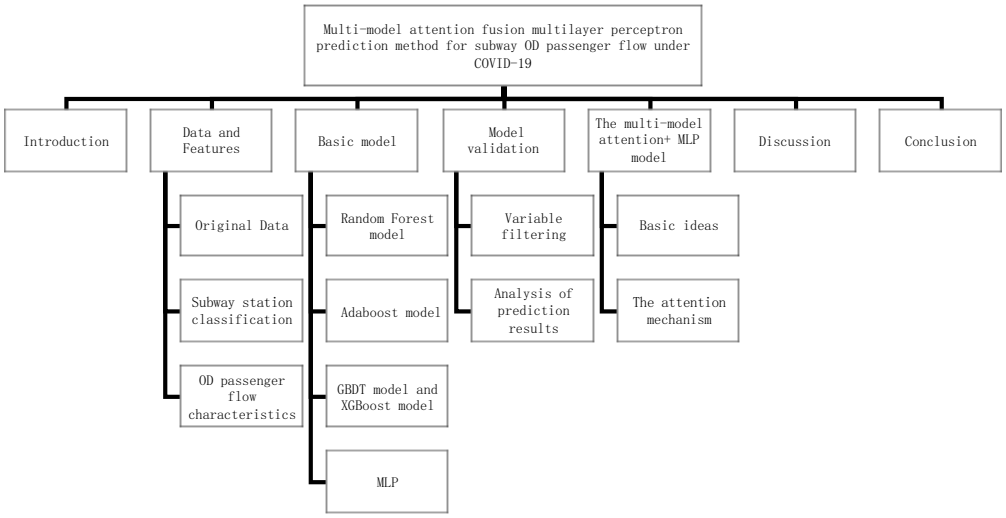


Figure 1. Article structure diagram.

2. Data and Features

2.1. Original Data

The historical subway OD data we use are from the card swiping data of Dalian Line 1 and Line 2 provided by the Metro Corporation in the project study compiled by the "14th Five-Year" Development Planning report of Dalian Urban Rail Transit, and the weather data are from the Meteorological Network, an open platform that provides data for various researchers. We summarized and sorted out the individual passenger flow data, eliminated invalid and abnormal data, and took the card replacement data from 1 July to 14 August as the training set and the card replacement data from 15 August to 30 August as the test set. Part of the sorted passenger flow data is shown in Table 1.

Table 1. Part of the sorted passenger flow data.

	1 July	2 July	3 July	4 July	5 July	…	26 August	27 August	28 August	29 August	30 August
Hua'nan Square–Qianshan Road	199	185	191	244	191	…	144	149	237	242	222
Hua'nan Square–Songjiang Road	244	236	253	188	176	…	157	162	356	276	282
Hua'nan Square–Dongwei Road	490	454	481	404	338	…	315	299	549	566	524
Hua'nan Square–Chunliu	474	470	467	341	235	…	267	244	420	374	355
…	…	…	…	…	…	…	…	…	…	…	…

2.2. Subway Station Classification

The same epidemic may have different impacts on the travel of residents with different land-use characteristics [19]. Therefore, according to the use attributes of different lands within 600 m around the station, we divided the station into five types: transfer, residence, employment, entertainment, and hub, as shown in Figure 2. As a result, 25 different types of OD station combinations were produced (Table 2).

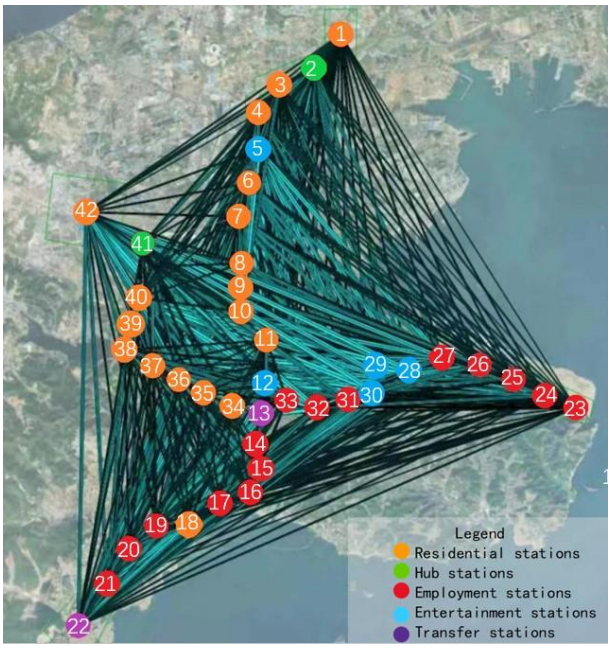


Figure 2. Classification and distribution of stations.

Table 2. The station corresponding to the number.

Number	Station Name	Number	Station Name	Number	Station Name	Number	Station Name	Number	Station Name	Number	Station Name
1	Yaojia	8	Dongwei Road	15	**Convention & Exhibition Center**	22	Hekou	29	Youhao Square	36	Malan Square
2	Dalian North Railway Station	9	Chunliu	16	Xinghai Square	23	Haizhiyun	30	Qingniwaqiao	37	Wanjia
3	Huabei Road	10	Xianggong Street	17	**2nd Hospital of Dalian Medical University**	24	Donghai	31	Yierjiu Street	38	Hongqi West Road
4	Hua'nan North	11	Zhongchang Street	18	Heishijao	25	Donggang	32	Renmin Square	39	Hongjin Road
5	Hua'nan Square	12	**Xinggong Street**	19	Xueyuan Square	26	Conference Center	33	Lianhe Road	40	Honggang Road
6	**Qianshan Road**	13	**Xi'an Road**	20	**Dalian Maritime University**	27	Dangwan Square	34	Dalian Jiaotong University	41	Airport
7	Songjiang Road	14	**Fuguo Street**	21	Qixianling	28	Zhongshan Square	35	Liaoning Normal University	42	Xinzhaizi

The stations mentioned in Section 5.2 are in bold.

2.3. OD Passenger Flow Characteristics

The outbreak of COVID-19 caused a change in the philosophy of travel of the population, which in turn led to a significant reduction in the share of public transport in the travel system. The share of public transportation in the whole travel system was significantly reduced [20]. Taking an outbreak in Dalian City in late July 2020 as an example, four sets of

OD passenger flow data were randomly selected from July to August 2020. The specific trends are shown in Figure 3.

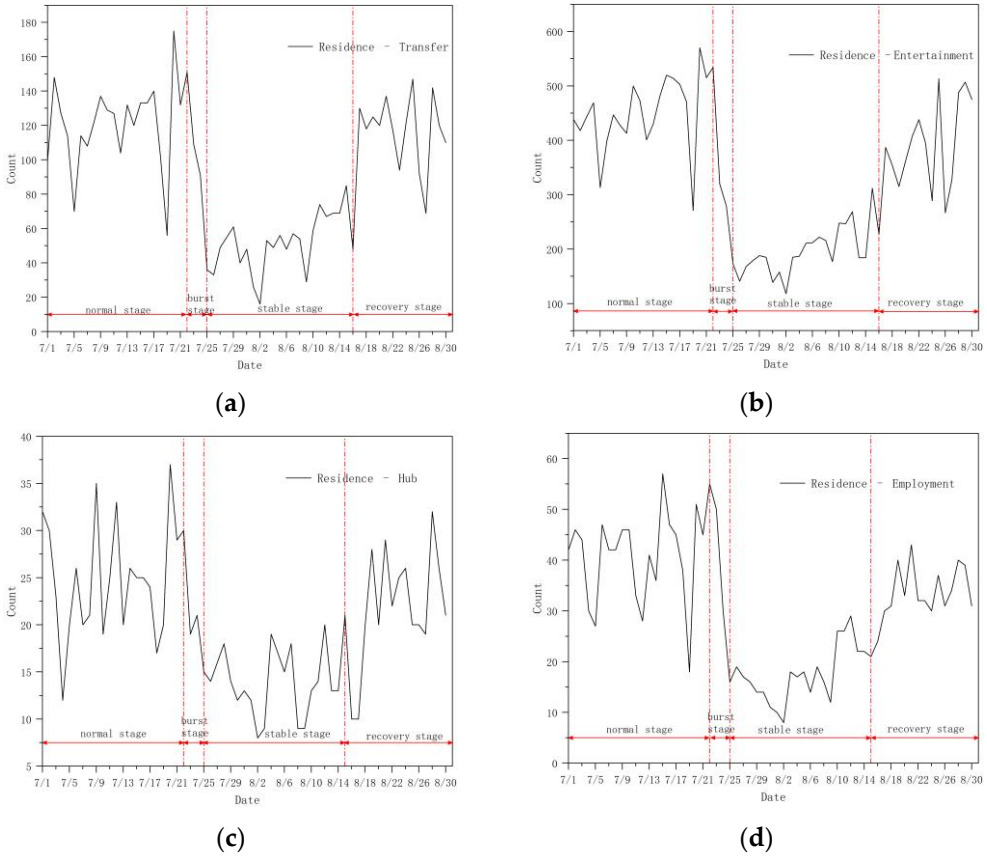


Figure 3. Historical OD passenger flow. (**a**) Historical passenger flow of Residence–Transfer. (**b**) Historical passenger flow of Residence–Entertainment. (**c**) Historical passenger flow of Residence–Hub. (**d**) Historical passenger flow of Residence–Employment.

An epidemic cycle can be divided into three stages, namely the burst stage (22 July to 25 July), the stable stage (26 July to 16 August), and the recovery stage (17 August to 30 August). During the burst stage, all four ODs experienced a "precipitous" decline in passenger traffic within 3–4 days, mainly due to the sudden outbreak of COVID-19, which led to a decline in the number of travelers. The main reason for this is the increased risk of going out due to the sudden outbreak and the reduced willingness of residents to travel by public transport [21]. At this stage, the "lag" in subway travel during COVID-19 is mainly due to the fact that the number of confirmed cases is usually announced in the afternoon or evening of the same day when most residents have already completed their travel activities for the day. It can be seen that the characteristics of COVID-19 on the previous day often have an impact on the passenger flow on the next day. In the burst stage, the OD of the Residence–Transfer has the largest change and compared with the passenger flow on 22 July, the passenger flow on 25 July decreased by 76.2%, the smallest change was a 50% decrease in the Residence–Hub, and the decline in the Residence–Entertainment and Residence–Employment was similar, 67.6% and 70.9%, respectively.

During the stable stage of COVID-19, the passenger flow of the four OD types maintains the characteristics of low-value cycle fluctuation, as shown by passenger flow on the weekend during the week, reaching the minimum value of the week, but the passenger flow is much lower than the regular traffic. During the recovery stage, the local policy of "opening entertainment venues" was announced, resulting in a continuous increase in passenger flow from the stable stage of COVID-19. Under normal conditions, the stations with less

traffic recover faster in this stage, while the stations with more normal traffic recover more slowly. On 21, 25, and 28 August, there were three extreme values for Residential-Transfer, Residential-Entertainment, and Residential-Employment, respectively, mainly due to the tendency of people to start "retaliatory" travel during breaks or special holidays after a period of closure.

3. Basic Model

3.1. Random Forest Model

The random forest contains multiple decision trees, which are trained randomly using the training set, and finally, the results obtained from all decision trees are averaged to obtain the final results [22]. Therefore, its core is to build multiple decision trees based on the feature attributes. One of the advantages of random forest is its suitable operability and interpretability [23]. In the OD passenger flow prediction under COVID-19, it is necessary to select a certain amount of historical data as the training set, analyze and select the influencing factors as the feature set, randomly select samples and influencing factors in the training set and feature set to build a decision tree, multiple decision trees to build a decision forest, fit the prediction model and get the prediction result. In order to improve the operation efficiency, it is necessary to stop splitting and pruning the decision tree.

3.2. Adaboost Model

The Adaboost model can be used for both classification and regression problems. The main idea is to adjust the sample weights by calculating the sample error rate and retrain the sub-models, and then integrate the sub-models that meet the requirements to get the final results. In the prediction process, the assigned weights appear in the training sub-model and the aggregated output sub-model prediction results, respectively. The former is to give more weight to the sub-models with unsatisfactory results to make them more visible in the next prediction, while the latter is to give more weight to the sub-models with satisfactory results [24], as shown in Figure 4.

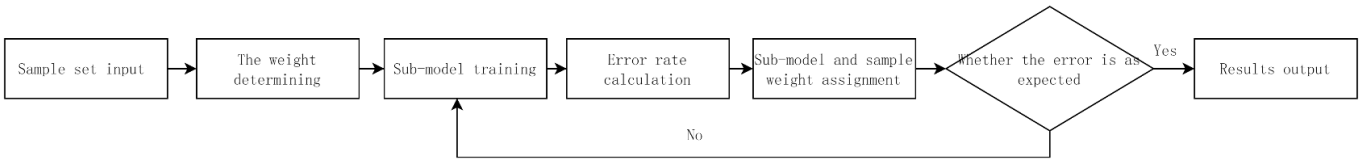


Figure 4. Adaboost model flow chart.

3.3. GBDT Model and XGBoost Model

The GBDT model is an integrated model that implements a learning optimization process using an additive model and a forward distribution algorithm. It contains multiple regression trees and uses the difference between the predicted and true values of the previous round of regression trees as a target of prediction to train new regression trees so that the final output is the sum of all regression trees [25].

Suppose the data set P (X, Y), where X is the sample data, and each sample datum contains multiple features, Y is the target value of the sample data, and there are a total of N samples. The loss function is $L(y, f(x))$, and the final regression tree is F_M.

(1) To initialize $F_0(x)$:

$$F_0(x) = arg_c min \sum_{i=1}^{N} L(yi, c) \quad (1)$$

(2) For M regression trees established: m = 1, 2, 3 . . . M:

a. Calculating the pseudo-residual of the m-th regression tree:

$$r_{m,i} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x)}\right]_{F(x)} = F_{m-1}(x) \quad (2)$$

b. By fitting $r_{m,i}$, we can obtain the leaf node $R_{m,j}$ of the m-th tree, where $j = 1, 2, 3 \ldots j_m$, j_m is the number of leaves.

c. Calculating the leaf nodes:

$$C_{m,j} = arg_c min \sum_{x_i \in R_{m,j}} L(y_i, F_{m-1}(x_i) + c) \tag{3}$$

where c is the constant.

d. The new $F_m(x)$ function is:

$$F_m(x) = F_0(x) + \sum_{m=1}^{M} \sum_{j=1}^{J_m} C_{m,j} I \quad x \in R_{m,j} \tag{4}$$

(3) The expression form of the final regression tree is:

$$F_M(x) = F_0(x) + \sum_{m=1}^{M} \sum_{j=1}^{J_m} C_{m,j} I \quad x \in R_{m,j} \tag{5}$$

The core idea of the XGBoost model is to use the prediction residuals of the previous sub-model to fit the constructed new sub-model. Unlike the GBDT, the complexity of the model is controlled by using a regular function, so it is an "upgraded" version of the GBDT model. The objective function of the model is shown in Equation (6).

$$minz = \sum_{i=1}^{n} L(y_i, \hat{y}_i) + \sum_{k=1}^{k} \Omega(f_k) \tag{6}$$

where $L(x, y)$ is the loss function, y_i is the true value of the i-th sample, $\hat{y}_i$ is the predicted value of the sample, $\Omega(x)$ is the canonical function, and f_k is the complexity of the k-th sub-model.

If the sub-model is a regression tree, the canonical function is related to the number of leaf nodes and leaf node values of the tree, as shown in Equation (7).

$$\Omega(f_k) = \alpha T + \frac{1}{2} \lambda w_i \tag{7}$$

where T is the number of leaf nodes, w_i is the value of the i-th node, and α and λ are parameters.

3.4. MLP

The increasingly strong use of neural networks has made it possible to combine the high ability to represent features with reinforcement learning [26]. MLP is a feedforward artificial neural network that exploits a supervised learning technique called backpropagation for training [27], which consists of an input layer, a hidden layer, and an output layer. MLP is formed when two neighboring layers are fully connected [28]. Its core is to construct a mapping relationship in the hidden layer that can fit the feature vector and the target relationship, as shown in Equation (8).

$$X' = \varsigma(Z) \quad Z = WX + b \tag{8}$$

where X is the input variable, X' is the output variable, $\sigma(X)$ is the activation function, W is the weight matrix, and b is the bias vector.

4. The Multi-Model Attention+ MLP Model

4.1. Basic Ideas

The fusion of models can get a better prediction effect [29]. Therefore, based on the attention mechanism, this paper tries to find a fusion algorithm to predict the OD passenger flow of the COVID-19 subway, namely the multi-model attention+ MLP model. Firstly, the correlation measure is used to screen the influencing factors. Second, the training set is used to train the sub-model of the attention mechanism. Finally, the output result of the trained MLP and the output result of the attention mechanism is weighted to obtain the final result, as shown in Equation (9). The flow of the prediction algorithm is shown in Figure 5.

$$\begin{gathered} y_t = \alpha_1 f(\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4) + \alpha_2 \hat{y}_5 \\ \alpha_1 + \alpha_2 = 1 \end{gathered} \tag{9}$$

where α_1, α_2 are the weights, $\hat{y}_1$, $\hat{y}_2$, $\hat{y}_3$, $\hat{y}_4$, and $\hat{y}_5$ are the prediction values of random forest, Adaboost, GBDT, XGBoost, and MLP, respectively, y_t is the final result of prediction, and $f(x)$ is the final output of the attention mechanism.

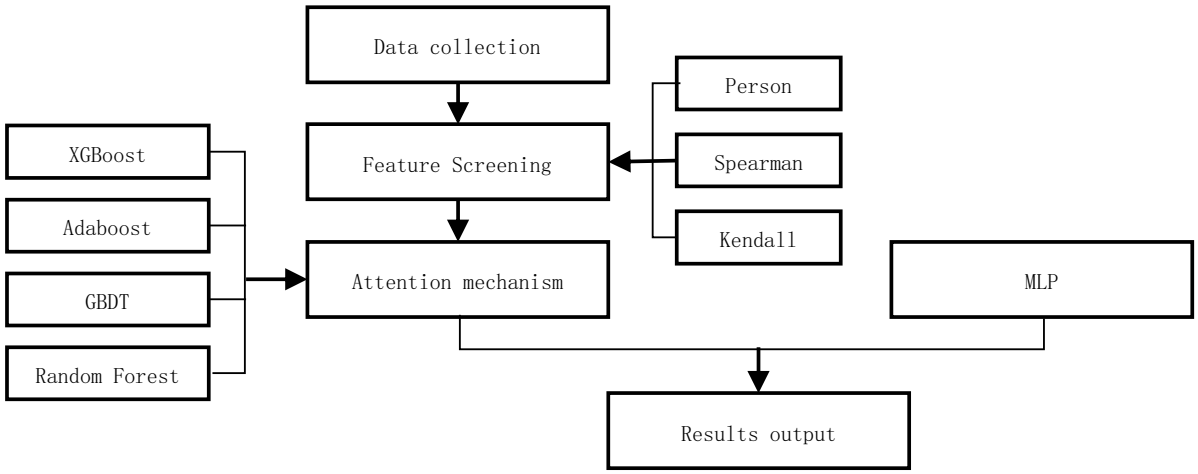


Figure 5. Algorithm flow chart.

4.2. The Attention Mechanism

The attention mechanism is a core technique proposed based on the mechanism of human vision [30]. Its core is to find a small amount of valuable information from a large amount of redundant information, that is, to make the more important information more easily noticed. Therefore, the essential problem of the attention mechanism is to give more weight to the valuable information [31]. Compared with other similar methods, the attention mechanism has the following characteristics: the structure can solve the most advanced multi-task model and does not use the loop model structure, and completely relies on the attention mechanism to establish the global dependence between the input and output. Therefore, the attention mechanism has the advantages of interpretation, higher parallelism, and simple structure.

Each period of COVID-19 is unique, and it is difficult to obtain sufficient data, and the variability of OD flows under COVID-19 makes it difficult for a single model to fit all OD types. Therefore, it is necessary to introduce the attention mechanism so that a prediction model that specializes in a certain OD type can be assigned a heavier task to obtain better prediction results. Thus, the algorithm can improve the generalization ability of the model while preventing model overfitting.

5. Model Validation

5.1. Variable Filtering

By analyzing the trend of passenger flow, it can be seen that the objective factors affecting the OD passenger flow of the metro are mainly divided into three categories: climate, holidays, and COVID-19. The climate factor includes two indicators of weather types and minimum temperature, and the holiday factor includes two indicators of whether it is a holiday and whether it is a weekend. The factor of COVID-19 includes three indicators: number of confirmed cases, government measures, and the level of COVID-19. In COVID-19, government policy was to close recreational venues, and the level index of COVID-19 included the number of low risks, medium risks, and high areas. The digitization of the influencing factors is shown in Table 3.

Table 3. Digitization of influencing factors.

Factors	Elaboration
Weather types	Sunny: 0; Cloudy: 1; Rain: 2; Overcast: 3; Light rain to moderate rain: 4
Minimum temperature	19, 20, 21, 22, 23, 24
Weekend	No: 0; Yes: 1
Holiday	Yes: 1; No: 0
Diagnosed number	0, 1, 2, 3, 5, 6, 8, 9, 11, 12, 14
Recreation site closure	No: 0; Yes: 1
Number of outbreak areas	0, 5
Low risk	Yes: 0; No: 1
Medium risk	0, 4
High risk	0, 1

The above influencing factors were analyzed by Pearson, Spearman, and Kendall correlation measure, and their obtained correlation coefficients are shown in Table 4.

Table 4. Correlation coefficient of characteristic indicators.

	Weather Types	Minimum Temperature	Weekend	Holiday	Diagnosed Number	Recreation Site Closure	Number of Outbreak Areas	Low Risk	Medium Risk	High Risk
Pearson	−0.045	−0.525	−0.134	−0.018	−0.400	−0.752	−0.751	0.751	−0.751	−0.751
Spearman	−0.051	−0.352	−0.120	−0.014	−0.344	−0.609	−0.630	−0.630	−0.630	−0.630
Kendall	−0.060	−0.506	−0.144	−0.017	−0.437	−0.733	−0.759	0.759	−0.759	−0.759

As shown in Table 4, the correlation between the two indicators of weather and whether it was a holiday on the impact of epidemic travel was poor. Therefore, the screening index of influencing factors includes: minimum temperature, whether it is a weekend, diagnosed number of people, whether to close recreational venues, number of epidemic areas, whether it is a low risk, and the number of medium and high risk.

A chi-square test was used to test the significance of the eight influencing factors, and the results showed that these factors were all significant, as shown in Table 5.

Table 5. Significant results of influencing factors.

Factors	Significant
Minimum temperature	0.015
Weekend	0.001
Diagnosed number	0.000
Recreation site closure	0.007
Number of outbreak areas	0.001
Low risk	0.001
Medium risk	0.001
High risk	0.001

Asymptotic significance is shown, and the level is 0.05.

Since the multi-model attention+ MLP model proposed in this paper is proposed on the basis of five sub-models, the above influencing factors also affect the combined model prediction.

5.2. Analysis of Prediction Results

In order to objectively analyze and evaluate the generalization ability and prediction ability of the combined model, the results of all sub-models in the prediction process are compared and analyzed with those of the combined model. Their hyperparameters are shown in Table 6.

Table 6. Hyperparameters of the model.

Model Name	Parameter Setting
Random forest	n-estimators: 200
Adaboost	Base_estimator: decision tree, learning rate: 1, n-estimators: 50, algorithm: SAMME.R
GBDT	Learning rate: 1, n-estimators: 100, sub-sample: 1
XGBoost	Learning rate: 0.1, n-estimators: 100, lambda: 0, alpha: 0
MLP	Batch size: 7, learning rate: 0.01, optimizer: Adam, loss function: MSELoss, Number of iterations: 200
Attention mechanism	Batch size: 8, learning rate: 0.02, optimizer: Adam, loss function: MSELoss, Number of iterations: 100

The subway passenger flow from 15 August to 30 August from Hua'nan Square to 2nd Hospital of Dalian Medical University of Dalian Metro Line 1 was selected as an experimental test with 78 sets of OD. 17 August was an important point in COVID-19 because entertainment venues officially opened on that day and the passenger flow accelerated the following day, and the predicted results are shown in Figure 6.

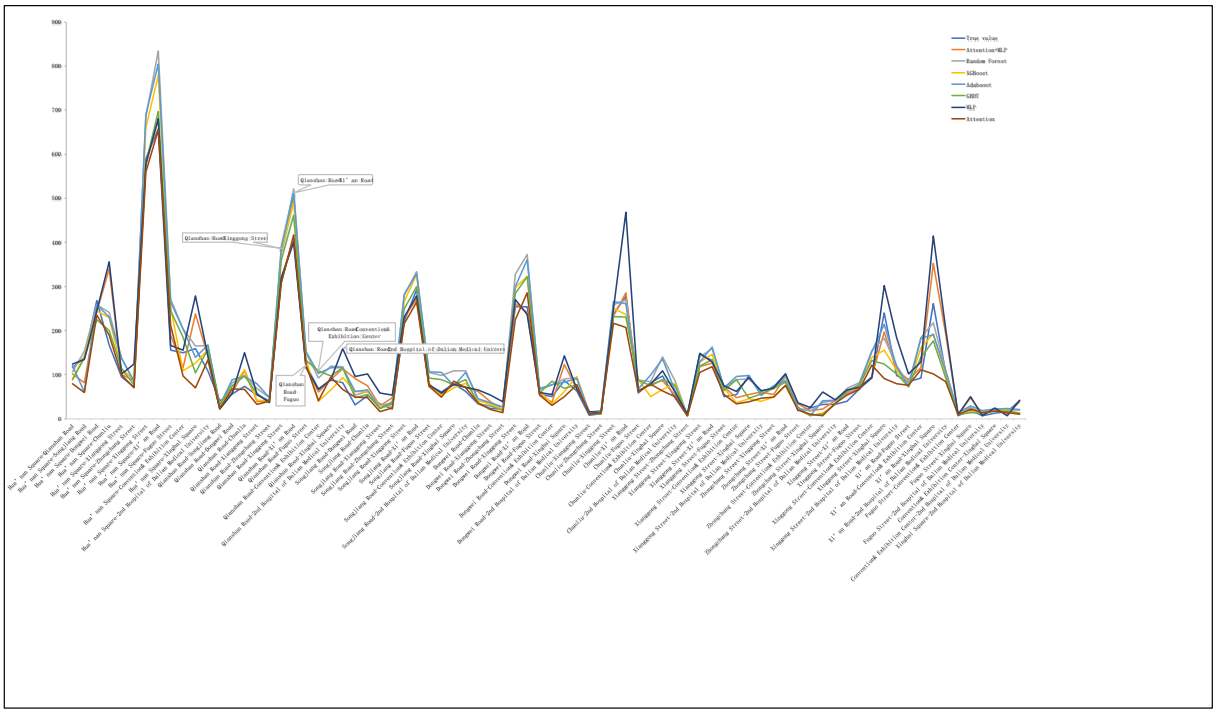


Figure 6. OD volume forecast on 16 August.

Qianshan Road is surrounded by heavy ground traffic and has a high passenger flow, so the six OD types are Residential–Employment, Residential–Transfer, Residential–Entertainment, Residential–Employment, Residential–Employment, and Residential–

Employment, taking Qianshan Road to Dalian Maritime University, Xi'an Road, Xinggong Street, Convention & Exhibition Center, Fuguo Street, and 2nd Hospital of Dalian Medical University as examples. A comparison of the predicted results is shown in Figure 7.

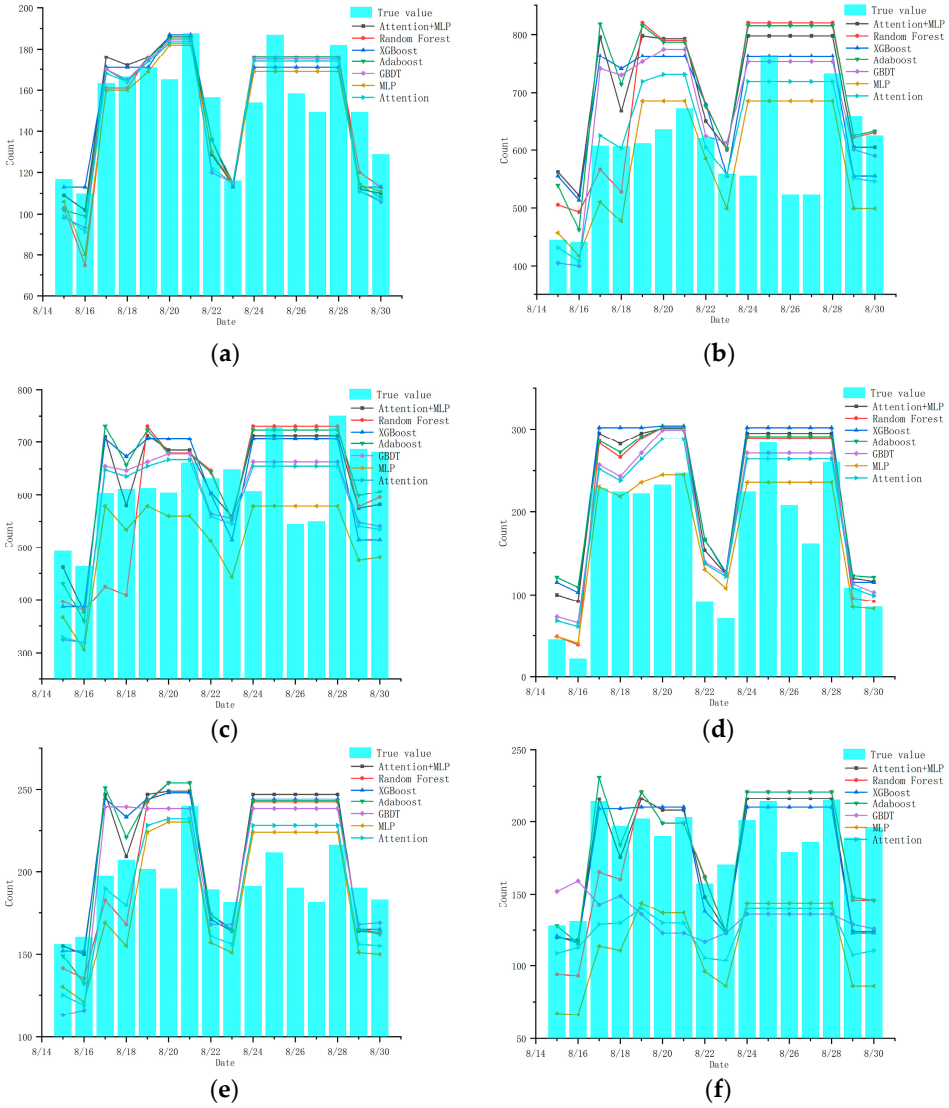


Figure 7. OD passenger flow prediction results. (**a**) Qianshan Road–Dalian Maritime University. (**b**) Qianshan Road–Xi'an Road. (**c**) Qianshan Road–Xinggong Stree. (**d**) Qianshan Road–Convention & Exhibition Center. (**e**) Qianshan Road–Fuguo Street. (**f**) Qianshan Road–2nd Hospital of Dalian Medical University.

As can be seen from Figure 7, only in the prediction of Qianshan Road–Dalian Maritime University Station did these prediction models have little difference, and the overall prediction effect is better than other OD types. In the prediction of Qianshan Road–Xi'an Road and Qianshan Road–Xianggong Street, the model predicts similar trends. This is because both types of stations have similar ridership. In the case of the Qianshan Road–Convention & Exhibition Center, the predictions were generally lower than the real values. The performance of GBDT, MLP, and attention mechanism on the Qianshan Road–2nd Hospital of Dalian Medical University was not as consistent as that on the Qianshan Road–Fuguo Street.

The mean absolute percent error (MAPE), root mean squared error (RMSE), and mean absolute error (MAE) were used as prediction evaluation indices, as shown in Equations (10)–(12).

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i - \hat{y}_i}{y_i} \tag{10}$$

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

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

The error results are shown in Figure 7.

It can be seen from Figure 8 that the single model only performs well in predicting the passenger flow of one or several types of OD pairs. For example, the RMSE of the XGBoost model of Qianshan Road–Dalian Maritime University reaches 18.835, and the MAE of the MLP model of Qianshan Road–Fuguo Street reaches 26.746. However, the multi-model attention+ MLP model with added attention mechanism is more suitable for all OD types of prediction work. The minimum value of MAPE was 9.535%, and the maximum value was 16.609%. This is because the attention mechanism picks out the characteristics of different basic models to adapt to the data so as to make prediction work and improve the accuracy of prediction.

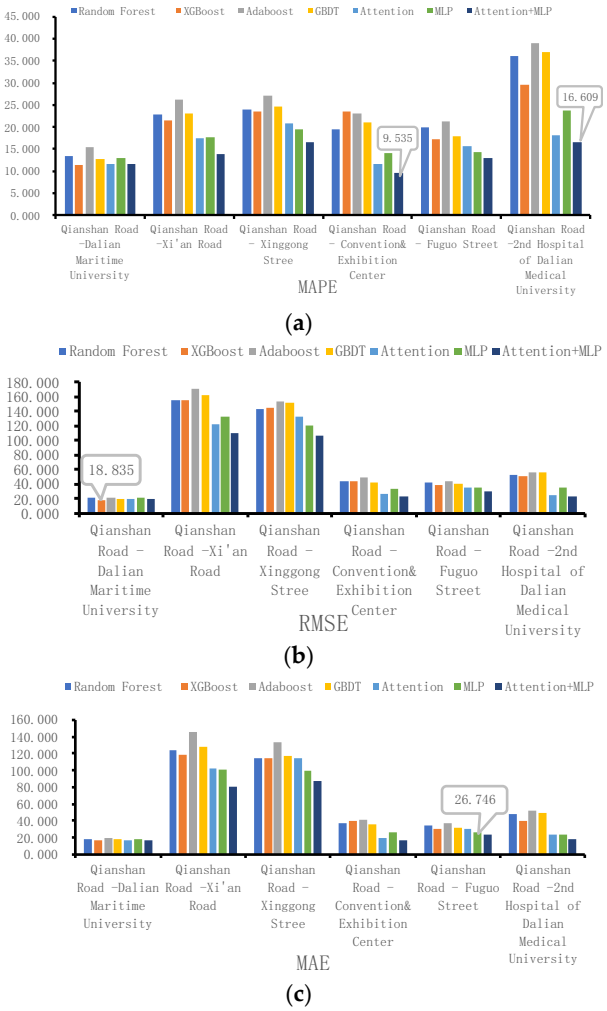
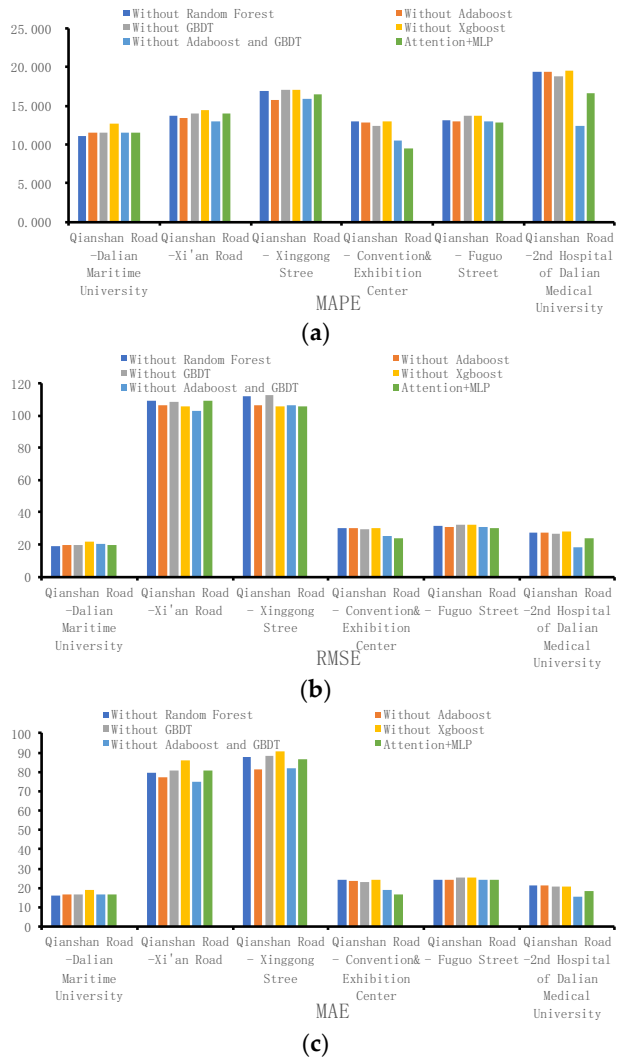


Figure 8. Error of OD passenger flow prediction. (**a**) MPE of OD passenger flow prediction. (**b**) RMSE of OD passenger flow prediction. (**c**) MAE of OD passenger flow prediction.

In order to further verify the scientificity of this paper and further explore and analyze the role of each module in the mixed model, we removed the four sub-models from the attention mechanism respectively as the second comparative experiment. At the same time, it can be clearly seen from Figure 8 that the Adaboost model and GBDT have poor prediction effects, so the two models will be removed from the attention mechanism at the same time also as this comparative experiment. Because the attention mechanism acts as a "magnifying glass", there are many ways to replace the attention mechanism, such as the entropy method, CRITIC weight method, and so on. However, due to the large number of methods and the difficulty in determining the evaluation criteria, we will not make too many comparisons here. (The comparative experiment on removing MLP is shown in Figure 8).

As can be seen from Figure 9 that the model with the smallest error is the one that excludes the Adaboost and GBDT, which is the fusion model of the attention mechanism composed of random forest and XGBoost model, followed by the one that integrates the attention mechanism composed of four sub-models. This is mainly because the attention mechanism amplifies their strengths as well as weaknesses. Therefore, the fusion model of attention mechanism composed of random forest and XGBoost and MLP is the best choice for the prediction of subway OD passenger flow under COVID-19.



(**a**)

(**b**)

(**c**)

Figure 9. Error of the removing some model in attention. (**a**) MPE of the removing some model in attention. (**b**) RMSE of the removing some model in attention. (**c**) MAE of the removing some model in attention.

6. Discussion

According to the findings, the multi-model attention+ MLP model can be successfully applied to predict metro OD passenger flows under COVID-19. In addition, COVID-19 was divided into three phases, and the impact of different OD flows was discussed separately.

It was found that during the burst stage of COVID-19, the metro passenger flow dropped sharply, indicating that the impact of COVID-19 on public transport was significant, which validated the studies in Refs. [17,20,21,32]. Furthermore, we found that the effect of COVID-19 on metro OD passenger flow was delayed in China, a finding that complements previous studies. During the stable stage of COVID-19, OD flows reach their lowest value at weekends. Residents prefer to stay at home rather than go out at this time, which is the same as the findings of Ref. [33]. During the recovery stage of COVID-19, residents begin to rely on the metro again for travel, but the passenger flow remains below normal for the same period, similar to the findings of Ref. [34]. In addition, the study found that at the end of the stage and residents were traveling more frequently than before as a retaliatory measure to alleviate the isolation period, similar to Ref. [35].

Residents are becoming increasingly dependent on the metro, which is why metro patrols are a hot topic of current research. Various forecasting methods have been developed and applied in order to predict passenger flows more accurately. However, so far, no scholars have studied the OD passenger flow under COVID-19. Similar to Ref. [36], this study divides stations into different types and forecasts them separately. Similar to Refs. [37,38], this paper also predicts passenger flow by using weather and day of the week as one of the influencing factors. Unlike the references [39,40], this study uses an attention mechanism, which ensures more accurate results and improves the generalization of the model. However, due to difficulties in data collection, additional influencing factors were not considered. In subsequent studies, more factors affecting OD can be filled in and combined with multi-models for a comprehensive study.

7. Conclusions

The prediction of passenger flow is a traditional subject that aims at providing a theoretical basis for operation management, planning, and design of the metro. The research results of this paper have made a very important contribution to the study of the subway during COVID-19. This paper divided an epidemic cycle into three stages and analyzed the trend of passenger flow at each stage. The metro OD passenger flow prediction model constructed in this paper fully considers the influence of various influencing factors on passenger flow. Through a series of comparative tests, this paper found the multi-model attention+ MLP model prediction algorithm conducive to the OD passenger flow of subway under COVID-19, that is, random forest and XGBoost as a sub-model of attention mechanism, and then integrated with MLP. The experimental results show that the multi-model attention+ MLP model improves the precision and accuracy of subway passenger flow prediction.

The difference between this study and subway passenger flow prediction is that we add the popularity factor and adopt some machine learning models as sub-models of the attention mechanism for the first time to weigh the output results of the attention mechanism and the output results of MLP. Multi-model attention+ MLP model improves the results of the comprehensive prediction of OD categories. According to the use characteristics of the different land of the surrounding areas, the subway stations are divided into five categories, and the OD passenger flow is predicted according to the temperature, time characteristics, and epidemic characteristics. In the current study, we found that the main epidemic factors affecting passenger flow were the number of confirmed cases, whether entertainment venues were closed, the number of affected areas, and whether they were low-risk, medium-risk, or high-risk areas.

We also studied passenger flow during COVID-19. We have divided the epidemic cycle into three phases. During the COVID-19 burst stage, OD flow dropped sharply within 3–4 days, with the impact of COVID-19 on flow lagging behind. In the COVID-19 stage, the

volatility of the value of passenger flow is low, and the overall trend is slowly increasing. During the recovery stage, which lasts about two weeks, there is a tendency for residents to take a "revenge" trip at the end of the first week, with flow slightly higher than normal.

Due to the instability of COVID-19 and limited data acquisition, the prediction accuracy of the model needs to be improved. Future work can improve the accuracy by adding other epidemic features and data. Nevertheless, the prediction methods and general laws proposed in this paper can still be used for similar studies.

Author Contributions: Conceptualization, Y.C. and X.L.; methodology, Y.C. and X.L.; software, X.L.; investigation, X.L.; validation, X.L.; visualization, Y.C. and X.L.; writing—original draft preparation, Y.C. and X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Scientific Research Funding Project of the Liaoning Provincial Education Department in 2020 (funder: Department of Education of Liaoning Province, grant no. JDL2020017), the Project of Liaoning Provincial Social Science Planning Fund in 2022 (funder: Social Science Planning Fund Office of Liaoning Province, grant no. L22BSH003), the 2022 Project of Dalian Academy of Social Sciences (funder: Dalian Academy of Social Sciences grant no. 2022dlsky078), the Education Quality Improvement Project for Post-graduate of Dalian Jiaotong University and the Teaching Reform Research Project for Undergraduate of Dalian Jiaotong University.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Dalian Metro Group for providing way passenger flow, Dalian Municipal Health Commission for epidemic data, and historical weather information data for Dalian City conditions provided by Dalian Meteorological Bureau as experimental materials for this study. We are also grateful to the editors and anonymous reviewers for their suggestions and comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. McKibbin, W.; Fernando, R. The global macroeconomic impacts of COVID-19: Seven scenarios. *Asian Econ. Pap.* **2021**, *20*, 1–30. [CrossRef]
2. Meng, P.C.; Li, X.Y.; Jia, H.F.; Li, Y.Z. Short-time rail transit passenger flow real-time prediction based on moving average. *J. Jilin Univ. Eng. Technol. Ed.* **2018**, *48*, 448–453.
3. Kang, J.; Duan, Z.T.; Tang, L.; Liu, Y.; Wang, C. A short term traffic flow prediction method based on gaussian processes regression. *J. Transp. Syst. Eng. Inf. Technol.* **2015**, *15*, 51.
4. Mohammed, O.; Kianfar, J. A machine learning approach to short-term traffic flow prediction: A case study of interstate 64 in Missouri. In Proceedings of the 2018 IEEE International Smart Cities Conference, Kansas City, MO, USA, 16–19 September 2018; pp. 1–7.
5. Sun, X.; Zhu, C.; Ma, C. Urban rail transit passenger flow forecasting—XGBoost. In Proceedings of the CICTP 2022, Changsha, China, 8–11 July 2022; pp. 1142–1150.
6. Huang, H.; Huang, L.; Song, R.; Jiao, F.; Ai, T. Bus single-trip time prediction based on ensemble learning. *Comput. Intel. Neurosc.* **2022**, *2022*, 6831167. [CrossRef]
7. Xu, Z.; Zhu, R.; Yang, Q.; Wang, L.; Wang, R.; Li, T. Short-term bus passenger flow forecast based on the multi-feature gradient boosting decision tree. In Proceedings of the International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, Kunming, China, 20–22 July 2019; pp. 660–673.
8. Gallo, M.; De Luca, G.; D' Acierno, L.; Botte, M. Artificial neural networks for forecasting passenger flows on metro lines. *Sensors* **2019**, *19*, 3424. [CrossRef]
9. Yang, D.; Chen, K.; Yang, M.; Zhao, X.C. Urban rail transit passenger flow forecast based on LSTM with enhanced long-term features. *IET Intell. Transp. Syst.* **2019**, *13*, 1475–1482. [CrossRef]
10. Li, M.; Li, J.; Wei, Z.J.; Wang, S.D.; Chen, L.J. Short-time passenger flow forecasting at subway station based on deep learning LSTM structure. *Urban Mass Transit* **2018**, *21*, 42–46.
11. Li, Y.; Wang, X.; Sun, S.; Ma, X.; Lu, G. Forecasting short-term subway passenger flow under special events scenarios using multiscale radial basis function networks. *Transp. Res. Part C Emerg. Technol.* **2017**, *77*, 306–328. [CrossRef]
12. Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wan, H. Attention based spatial-temporal graph convolutional networks for traffic flow forecasting. In Proceedings of the AAAI conference on artificial intelligence, Honolulu, HI, USA, 27 January–1 February 2019; pp. 922–929.

13. Wang, F.J.; Yu, J.H.; Zhao, J.H.; Mei, Z.Y. Short-term public traffic passenger volume forecasting method based on real-time relevance of stations. *Transp. Syst. Eng. Inf. Technol.* **2021**, *21*, 131.
14. Bai, L. Urban rail transit normal and abnormal short-term passenger flow forecasting method. *Transp. Syst. Eng. Inf. Technol.* **2017**, *17*, 127–135.
15. Sun, Y.; Leng, B.; Guan, W. A novel wavelet-SVM short-time passenger flow prediction in Beijing subway system. *Neurocomputing* **2015**, *166*, 109–121. [CrossRef]
16. Zhao, Y.Y.; Liang, X.; Jiang, X.G. Short-term metro passenger flow prediction based on EMD-LSTM. *Transp. Syst. Eng. Inf. Technol.* **2020**, *20*, 194–204.
17. Yang, Y.Z.; Tang, H.D. Residents' travel mode choice behavior in post-COVID-19 era considering preference differences. *J. Transp. Syst. Eng. Inf. Technol.* **2022**, *22*, 15.
18. Shi, J.; Long, Y.X. Research on the impacts of the COVID-19 on individual's leisure travel. *China J. Highw. Transp.* **2022**, *35*, 238.
19. Yang, J.; Wu, K.; Zhang, H.L.; Dai, S.X.; Wang, Y.L. Classification of subway stations based on land use and passenger flow characteristics. *J. Transp. Syst. Eng. Inf. Technol.* **2021**, *21*, 228.
20. Jiang, N.; Li, S.; Cao, S.Z.; Wei, J.; Wang, B.; Qin, N.; Duan, X. Transportation activity patterns of Chinese population during the COVID-19 epidemic. *Res. Environ. Sci.* **2020**, *33*, 1675–1682.
21. Wang, H.; Noland, R.B. Bikeshare and subway ridership changes during the COVID-19 pandemic in New York City. *Transp. Policy* **2021**, *106*, 262–270. [CrossRef]
22. Biau, G.; Scornet, E. A random forest guided tour. *Test* **2016**, *25*, 197–227. [CrossRef]
23. Li, Y.; Ngom, A. Data integration in machine learning. In Proceedings of the 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Washington, DC, USA, 9–12 November 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1665–1671.
24. Wang, Y.H.; Lv, J.; Wu, J.; Wang, C. Prediction method of restoring force based on online AdaBoost regression tree algorithm in hybrid test Share. *J. Southeast Univ. Engl. Ed.* **2020**, *36*, 181–187.
25. Chen, T.; Zhu, L.; Niu, R.Q.; Tringer, C.J.; Peng, L.; Lei, T. Mapping landslide susceptibility at the Three Gorges Reservoir, China, using gradient boosting decision tree, random forest and information value models. *J. Mt. Sci.* **2020**, *17*, 670–685. [CrossRef]
26. Guarino, A.; Grilli, L.; Santoro, D.; Messina, F.; Zaccagnino, R. To learn or not to learn? Evaluating autonomous, adaptive, automated traders in cryptocurrencies financial bubbles. *Neural. Comput. Appl.* **2022**, 1–42. [CrossRef]
27. Guarino, A.; Lettieri, N.; Malandrino, D.; Zaccagnino, R.; Capo, C. Adam or eve? Automatic users' gender classification via gestures analysis on touch devices. *Neural. Comput. Appl.* **2022**, *34*, 18473–18495. [CrossRef]
28. Masci, J.; Meier, U.; Cireşan, D.; Schmidhuber, J. Stacked convolutional auto-encoders for hierarchical feature extraction. In Proceedings of the International Conference on Artificial Neural Networks, Espoo, Finland, 14–17 June 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 52–59.
29. Seeland, M.; Mäder, P. Multi-view classification with convolutional neural networks. *PLoS ONE* **2021**, *16*, e0245230. [CrossRef]
30. Liu, Y.C.; Li, Z.P.; Lv, C.P.; Zhang, T.; Liu, Y. Network-wide traffic flow prediction research based on DTW algorithm spatial-temporal graph convolution. *J. Transp. Syst. Eng. Inf. Technol.* **2022**, *22*, 147.
31. Huang, T.Y.; Yang, Y.L.; Yang, X.J. A survey of deep learning-based visual question answering. *J. Cent. South Univ.* **2021**, *28*, 728–746. [CrossRef]
32. Huang, J.; Wang, H.; Fan, M.; Zhuo, A.; Sun, Y.; Li, Y. Understanding the impact of the COVID-19 pandemic on transportation-related behaviors with human mobility data. In Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, New York, NY, USA, 6–10 July 2020; pp. 3443–3450.
33. Pan, Y.; Darzi, A.; Kabiri, A.; Zhao, G.; Luo, W.; Xiong, C.; Zhang, L. Quantifying human mobility behaviour changes during the COVID-19 outbreak in the United States. *Sci. Rep.* **2020**, *10*, 20742. [CrossRef]
34. Currie, G.; Jain, T.; Aston, L. Evidence of a post-COVID change in travel behaviour—Self-reported expectations of commuting in Melbourne. *Transp. Res. Part A Policy Pract.* **2021**, *153*, 218–234. [CrossRef]
35. Ramires, A.; Carvalho, I.; Correia, A. Pre-and post-pandemic travel behaviour and intentions: Clustering Portuguese generations. *Anatolia* **2022**, 1–17. [CrossRef]
36. Yu, L.; Chen, Q.; Chen, K. Deviation of peak hours for urban rail transit stations: A case study in Xi'an, China. *Sustainability* **2019**, *11*, 2733. [CrossRef]
37. Hui, Y.; Wang, Y.G.; Peng, H.; Hou, S.Q. Subway passenger flow prediction based on optimized PSO-BP algorithm with coupled spatial-temporal characteristics. *J. Transp. Syst. Eng. Inf. Technol.* **2021**, *21*, 210–222.
38. Zhu, H.; Yang, X.; Wang, Y. Prediction of daily entrance and exit passenger flow of rail transit stations by deep learning method. *J. Adv. Transp.* **2018**, *2018*, 6142724. [CrossRef]
39. Yang, J.; Dong, X.; Jin, S. Metro passenger flow prediction model using attention-based neural network. *IEEE Access* **2020**, *8*, 30953–30959. [CrossRef]
40. Zhang, G.; Liu, W.; Zheng, H.; Ma, T. Short-term passenger flow prediction for urban rail transit based on time-space attention graph convolutional network. In Proceedings of the 2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT), Shenyang, China, 13–15 November 2020; pp. 548–553.