Traffic & Transportation

# Analysis and Dynamic Prediction of Bus Dwell Time Under Rainfall Conditions

Baoyun SUN[1], Yaping YANG[2], Lei DONG[3], Honglin LU[4], Zimin WANG[5]

Original Scientific Paper
Submitted: 8 Jan 2024
Accepted: 20 June 2024

[1] cebysun@sjzu.edu.cn, Shenyang Jianzhu University, School of Transportation and Surveying Engineering
[2] Ameliaping.y@gmail.com, Shenyang Jianzhu University, School of Transportation and Surveying Engineering
[3] Corresponding author, donglei@sjzu.edu.cn, Shenyang Jianzhu University, School of Architecture and Urban Planning
[4] 18841998932@163.com, Shenyang Jianzhu University, School of Transportation and Surveying Engineering
[5] Min970409@163.com, Shenyang Jianzhu University, School of Transportation and Surveying Engineering



Publisher:
Faculty of Transport and Traffic Sciences, University of Zagreb

**ABSTRACT**
Exploring the degree to which bus stop times are affected by rainfall is necessary for a reasonable formulation of bus-scheduling management schemes under rainy conditions. Although numerous mathematical models have been proposed, the predictive accuracy of existing models is insufficient for the precise formulation of bus policies. This study considered linear bus stops in Shenyang as research targets, and based on field survey data, we analysed the bus dwell time and its influencing factors under varying degrees of rainfall. The Pearson correlation analysis method and SPSS software were used to reveal the degree of influence of parameters, such as the number of passengers boarding and alighting buses, rainfall level, number of berthing spaces, load rate and presence of signalised intersections, on the bus stop time under rainfall conditions. Support vector machine, k-nearest neighbour and backpropagation (BP) prediction models were established, and the BP neural network model, having the best prediction effect, was optimised using a genetic algorithm (GA). The constructed GA-BP prediction model was more realistic than the BP prediction model and can be used to predict bus dwell times under rainfall conditions. The study findings will facilitate bus punctuality and improve customer appeal for bus services.

**KEYWORDS**
rainfall conditions; bus dwell time; influencing factors; correlation; GA-BP model.

## 1. INTRODUCTION

As urban traffic problems worsen, public transportation systems have garnered widespread attention worldwide because they effectively alleviate urban traffic congestion, reduce traffic accidents, ensure daily mobility for low-income groups and enhance travel quality [1]. Bus stops, as fundamental infrastructures of public transportation systems, serve as vital links between passengers and vehicles, fulfilling the basic functions of vehicle stopping and passenger boarding/alighting, thus constituting an indispensable part of public transportation systems [2]. However, when buses stop at stations without dedicated bus lanes, they often create traffic bottlenecks, impeding the normal flow of other vehicles, non-motorised vehicles and pedestrians, consequently affecting the operational capacity and overall service quality of bus transit.

Insufficient capacity at both linear and bay-type bus stops can lead to urban traffic congestion and accidents and also result in chaotic vehicle parking, causing adverse effects on the overall transportation system of cities, especially during peak commuting hours. Bus dwell time, which refers to the time during which a bus serves passengers at a bus stop, is a major factor influencing the throughput capacity of bus stops. Accurately determining the impact of bus dwell time on the efficiency and reliability of the bus system can produce significant effects and help formulate more effective strategies that ultimately improve passenger satisfaction. Bus dwell time is constrained by various factors, including the number of passengers getting on and off, fare payment method, bus congestion level, use of bus doors and external environmental factors [3].

As a common external condition, rainy weather, with reduced visibility, slippery roads and slower vehicle speeds, can exacerbate bus-stop queues. This decreases the capacity of bus stops, contributes to increased urban traffic congestion, and is a potential traffic safety hazard. In addition to affecting the visibility of bus drivers in rainy conditions, vehicle deceleration rates, passenger boarding durations, bus operational efficiency and traffic flow vary compared to clear weather. Furthermore, puddles at bus stations and splashing rainwater pose safety risks to bus operations, leading to a decline in service quality. These factors collectively impact the travel experiences of passengers, diminishing their overall comfort and convenience during daily commutes [4, 5]. Therefore, scientifically and reasonably predicting the bus stop time under rainy conditions is of great practical significance and application value in improving the efficiency of bus operations and the overall service quality of public transportation.

## 2. LITERATURE REVIEW

Several studies have extensively investigated the various factors influencing bus stop times across different scales, such as people, vehicles, roads and traffic organisations. When buses halt at stations, their operational characteristics change significantly, resulting in additional bus stops and stoppage loss. Previous studies have emphasised the critical role of passenger volume among the factors influencing bus stoppage and stoppage loss times [4, 6]. Milkovits [7] used data from automatic passenger counting and vehicle positioning systems on Chicago buses to develop and implement pre-processing techniques, estimated a dwell-time model and found that the impact of crowding on bus dwell times was significant. Al-Jumailey [8] found that high congestion at bus rapid transit (BRT) stations increases passenger resistance to movement, obstructing passenger boarding, ultimately leading to increased boarding and bus dwell times. Ji et al. [9] studied the movement characteristics of boarding passengers and bus dwell-time prediction and found that enlarging the platform area and installing guardrails can reduce bus dwell-time fluctuations. Bladikas et al. [10] examined bus stop times during adverse weather conditions, revealing a substantial increase in boarding and alighting times during rainy and snowy weather. Stover et al. [11] used the least squares regression equation to study the impact of weather on bus ridership in Prince George's County, Maryland. They found that unfavourable weather conditions, such as rain and snow, negatively impact bus ridership.

Regarding predictive models, Glick et al. [12] investigated the interactions between buses at stops and bus delays, employing a log-linear regression model to forecast bus arrival times. They found that bus interactions significantly impact dwell times. Ding [13] used data from the Changzhou BRT route and applied an autoregressive integrated moving average model (ARIMA)–support vector machine (SVM) and a hybrid model to predict bus stop times. The ARIMA–SVM combination model demonstrated effective predictive performance. However, the study did not incorporate the impact of weather factors into its analysis. Rashidi [14] employed machine-learning models, including k-nearest neighbours (KNN), gradient boosting regressor, random forest regressor, XGBoost and one-dimensional convolutional neural networks, to predict the stopping times of a bus at eight bus stations within the Tu Makuru to Kyathasandra road sections in India over two weeks. Their study revealed that the XGBoost model was more accurate than the other models. However, the prediction error exceeded 20% because they only used location data with no boarding and alighting passenger quantity data. Similarly, backpropagation (BP) neural network prediction models have found widespread applications in public transportation. For example, Liu et al. [15] proposed an enhanced analytic hierarchy process-backpropagation (AHP–BP) neural network method based on passenger survey data to evaluate and analyse quality factors. The results revealed that the improved AHP–BP neural network method was quite promising for future urban public transportation service quality assessments. However, AHP-BP is more suitable for situations where the problem structure is well defined and the related factors are clear. Additionally, this method has high computational complexity, a long training time and relies on the subjective judgment of decision-makers, which could lead to biased results.

In summary, existing studies on the factors influencing bus dwell times and their predictions have the following shortcomings: First, they do not consider rainfall conditions, which makes it challenging to comprehensively support the rational formulation of bus-scheduling management plans in scenarios with increased rainfall frequency and intensity. Second, urban public transportation systems are complex systems influenced by numerous factors, resulting in significant fluctuations in bus dwell times under rainy conditions. The predictive accuracy of traditional mathematical models is insufficient to support the precise formulation of bus policies and scheduling management.

Therefore, this study builds on existing research and considers rainfall factors using the actual operational data of buses operating under varying rainfall intensities. This study employed the Pearson correlation analysis and SPSS software to investigate the impact of rainfall levels on parameters such as boarding and alighting passenger numbers, boarding loss time, parking space availability, passenger load factors and station entry and exit times. The analysis revealed the extent to which various factors influence bus stop times under different rainfall conditions. Considering the significant variability of bus dwell times in rainy conditions and the nonlinear mapping capability of the BP neural network model, this study established a BP neural network prediction model. This study also used the SVM and KNN prediction models and compared the predictive accuracy and sample generalisation abilities of the three models. A BP neural network model with optimal predictive performance was further optimised using a genetic algorithm (GA), resulting in the development of a higher-precision GA-BP prediction model. Thus, this provides a theoretical basis for the rational formulation of bus-scheduling management plans and the enhancement of conventional public transportation service levels under rainy weather conditions in urban environments.

## 3. MATERIALS AND METHODS

The bus-stopping process at a station occurs in four stages: waiting outside the station, decelerating to enter the station, stopping within the station and accelerating to leave the station. Consequently, the bus dwell time at a station encompasses four components: waiting time outside the station, deceleration to enter time, stopping within the station time and acceleration to leave time (including exit-waiting time) [16]. *Figure 1* shows the composition of the bus stop time.

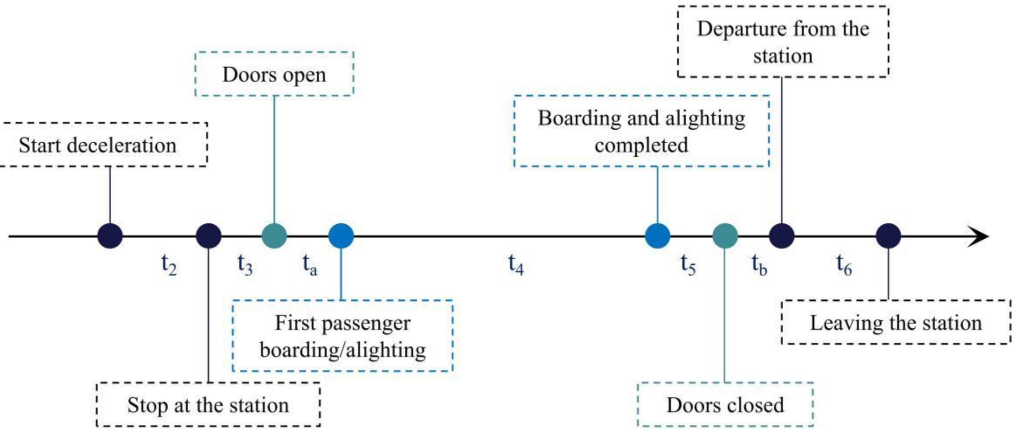



*Figure 1 – Bus time definition*

The bus stop time is calculated using *Equations 1 and 2*

$$T = t_1 + t_2 + t_3 + t_4 + t_5 + t_6 + t_7 \tag{1}$$

$$t_7 = t_a + t_b \tag{2}$$

where $T$ denotes the bus dwell time(s), $t_1$ is the waiting time outside the station(s), $t_2$ is the deceleration time(s) to enter the station(s), $t_3$ is the door opening time(s), $t_4$ is the in-station stopping time, taking the maximum of boarding and alighting times(s), $t_5$ is the door closing time(s), $t_6$ the accelerating to leave time(s), $t_7$ the loss time(s), $t_a$ the boarding loss time(s), and $t_b$ the leaving station loss time(s).

Crowding conditions inside the bus significantly influence passenger boarding and alighting times [6]. The existing studies commonly use the passenger load factor to quantify the level of crowding on a bus. The load factor refers to the ratio of the actual passenger load to the rated passenger capacity of the bus [1]. When the load factor is in the (0,1) range, passengers can find seats and ample space remains within the bus. For a load factor between (1, 1.5), some passengers may have to stand, leading to physical contact among passengers and a perceived sense of crowding. When the load factor exceeds 1.5, passengers experience significant discomfort, and movement within the bus becomes difficult.

The loss time is the sum of the boarding and leaving station loss times; the boarding loss time is the time difference between the opening of the front door of the bus and the first passenger boarding; and the leaving station loss time refers to the time difference between the closure of vehicle doors and the initiation of departure from the station. These times are associated with the starting performance of the vehicle and driver operation.

### 3.1 Data source and analysis

*Classification of rainfall levels*

Meteorological departments primarily use the amount of rainfall as a measure of rainfall intensity, with the precipitation intensity characterised by the rainfall amount at various time intervals such as 1-h, 12-h and 24-h durations. This study focused on the short-term impact of rainfall on bus stops and, thus, adopted a 1-h rainfall intensity. The classification of rainfall levels [17] is presented in *Table 1*.

*Table 1 – Classification of rainfall levels*

| Precipitation (mm/h) | Rainfall magnitude | Rainfall level |
|---|---|---|
| $<0.5$ | No rain | 0 |
| 0.5–2 | Light rain | 1 |
| 2–6 | Moderate rain | 2 |
| $>6$ | Heavy rain and storm | 3 |

*Survey period*

This study assumed that the impact of hourly precipitation on the bus-stopping process remains consistent within an hour, owing to the hourly updating and statistical nature of the precipitation data. The analysis focused on precipitation during the operational hours of bus services, specifically examining the influence of rainfall on the bus-stopping process within this timeframe. Precipitation data were obtained from the rainfall monitoring system at Nanta Park, Shenyang, from April to September 2023 (easy m. cloud). Rainfall data were recorded at 5-minute intervals. Nanta Park was approximately 5.5 km from the survey location and had similar weather conditions. The precipitation data are summarised in *Table 2*.

*Table 2 – Statistics of rainfall data during the investigation period from September to June in Shenyang*

| Date | 1-hour Precipitation in mm | Rainfall Level |
|---|---|---|
| 2023.4.25 | 3.0 | 2 |
| 2023.4.28 | 2.0 | 1 |
| 2023.4.29 | 1.0 | 1 |
| 2023.5.12 | 3.5 | 2 |
| 2023.6.10 | 5.5 | 2 |
| 2023.6.26 | 9.0 | 3 |
| 2023.7.22 | 2.0 | 1 |
| 2023.8.12 | 12.0 | 3 |
| 2023.9.22 | 1.0 | 1 |

*Dwell-time data*

Eight linear bus stations in Hunnan new district, Shenyang City, were selected as the investigation targets: Hengda Jiangwan Station, Yangguan Station, Hunnan East Road Zhuke Third Street Station, Shenyang Export Processing Zone Station, Hunnan East Road Wenshuo Street Station, East Asia International City Station, Hunnan East Road Wenhua Street Station and Shenyang Jianzhu University Bus Station. All eight bus stations had two berths and no dedicated bus lanes. Following the observational method proposed by González [18], this study employed two observers during data collection. The bus docking process was captured through video observation with Observer 1 stationed near the front bus door. The camera recorded the entire sequence from the bus arrival to departure, focusing on boarding passenger conditions and the deceleration process as the bus approached the station. In addition, unexpected events were documented. Observer 2 was positioned near the rear bus door and captured the entire bus sequence from arrival to departure. The primary focus was on passenger alighting conditions, acceleration upon leaving the bus station and the situation of passengers waiting to exit. Each stopping process corresponded to one sample; 409 samples were collected, including 288 samples collected on rainy days, as summarised in *Table 3*.

*Table 3 – Sample size*

| **Sample size/Rainfall level** | **Rainy conditions** | **Non-rainy conditions** | | |
|---|---|---|---|---|
| | 0 | 1 | 2 | 3 |
| Peak workday samples | 48 | 14 | 14 | 4 |
| Peak non-workday samples | 52 | 14 | 0 | 28 |
| Off-peak workday samples | 11 | 113 | 40 | 3 |
| Off-peak non-workday samples | 10 | 51 | 0 | 7 |
| Total samples | 121 | 288 | | |

The following data were extracted from the research materials:
- Bus information
- Stopping station information
- Boarding and alighting passenger numbers
- Onboard crowding level
- Bus loss time
- Bus dwell time

The rated passenger capacity of buses was obtained through on-site surveys, and the nature of stopping stations and the number of berths were directly observed on-site. The boarding and alighting passenger numbers and actual passenger loads were obtained through video observations. The onboard crowding level was calculated as the ratio of the actual passenger load to the rated passenger capacity. The loss and bus stop times were calculated from the time differences observed in the videos. As rainfall does not affect the opening and closing of bus doors, relevant data for door opening and closing times were not extracted in this study. The records for the Shenyang Jianzhu University Bus Station provided the largest sample size and served as the training set for subsequent model establishment, whereas data from other bus stations were used as the test set to validate the predictive performance of the model.

### 3.2 Correlation analysis

This study employed the Pearson correlation analysis method to investigate the correlation between various parameters during bus stops under different rainfall intensities [19], analysing all sample data using the SPSS software. The correlation coefficient heat map is presented in *Figure 2*, where the colour depth represents the magnitude of the values.

## 4. RESULTS

A scatter plot of the dwell time under different rainfall levels is illustrated in *Figure 3*. As observed from the graph, under non-rainy conditions, bus dwell times were concentrated in the 10–30 second range, whereas under rainy conditions, bus dwell times were distributed from 10 to 40 seconds, with a higher concentration of samples under light rain conditions. Compared with non-rainy conditions, the data under rainy conditions exhibited greater dispersion; the distribution of dwell times did not reveal distinct linear characteristics.

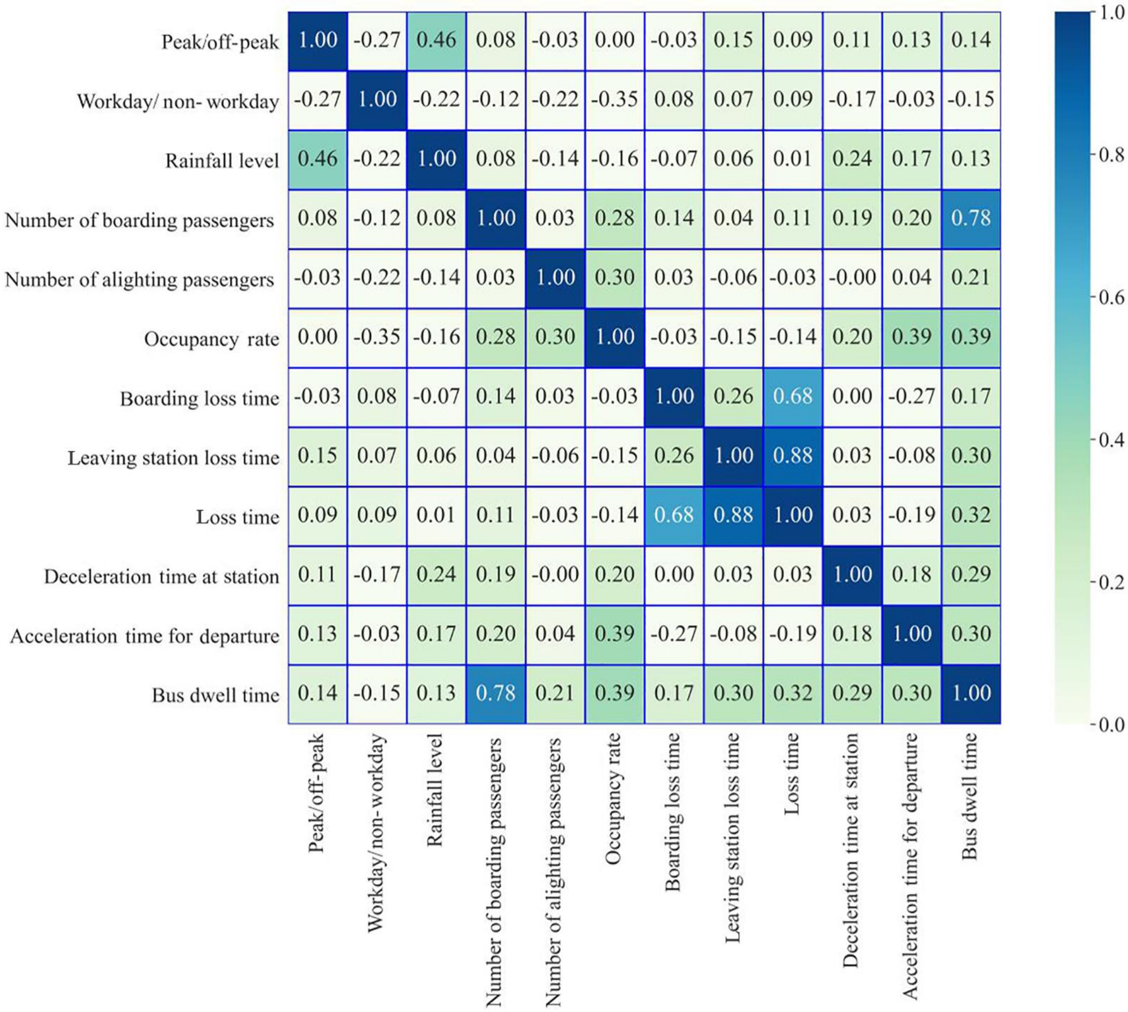



*Figure 2 – Correlation coefficient heatmap*

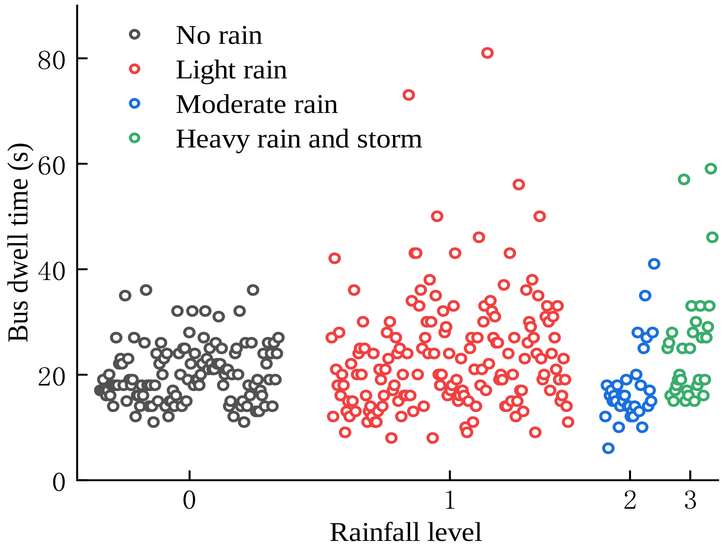



*Figure 3 – Scatter plot of bus stop time under varying degrees of rainfall*

The frequencies of boarding passengers during peak workdays, off-peak workdays, peak non-workdays and off-peak non-workdays are illustrated in *Figure 4*. From the graph, the frequency of boarding passengers ranging from 0 to 5 individuals was the highest across all samples, with a decreased frequency of boarding passengers in the 6–17 individual range. A comparison revealed that passengers commuting by bus were more prevalent on off-peak workdays.

The correlation between bus dwell time under rainy conditions and the frequency of boarding passengers was notably positive, with a correlation coefficient of 0.78. The bar chart in *Figure 5* depicts the frequency of passenger boarding under different rainfall levels. Boarding passengers were concentrated in the 0–11 individual range during rainfall Level 0 (no rain), 0–6 individuals during rainfall Level 1 (light rain) and rainfall Level 3 (heavy rain and storm) and 0–5 individuals during rainfall Level 2 (moderate rain). Rainfall influences travel behaviour and choices.

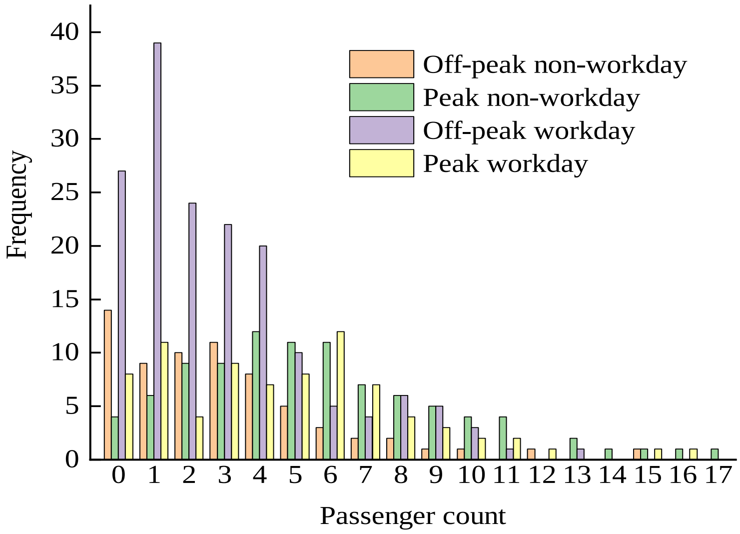



*Figure 4 – Frequency of boarding passengers at different time intervals*

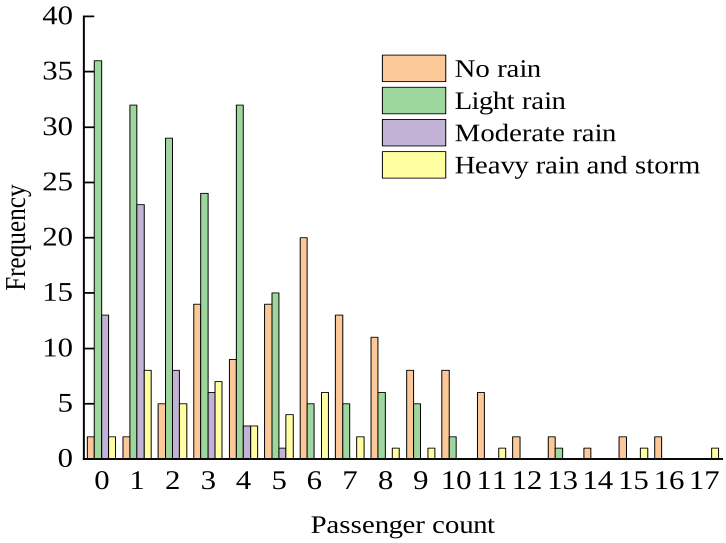



*Figure 5 – Frequency of the number of passengers under varying rainfall levels*

The box plots in *Figure 6*, representing boarding passengers in the 1–5 individual range under different rainfall levels, indicate that the greater the number of boarding passengers, the longer the bus dwell time. For the same number of boarding passengers, higher rainfall levels correspond to longer bus dwell times. When the number of boarding passengers is 1–3 individuals, bus dwell times under different rainfall levels are concentrated in the 12–25 s range, with samples featuring longer dwell times being more prevalent under rainy conditions.

A positive correlation exists between the boarding loss time and the rainfall level. Generally, during rainfall, passengers wait with umbrellas, and closing the umbrellas upon boarding extends the boarding loss time. A positive correlation also exists between the boarding loss time and occupancy rate; as the bus occupancy rate increases, the bus becomes more crowded, rendering boarding more challenging and resulting in longer boarding loss times. Under light rain conditions, with occupancy rates in the 1–1.5 range, the longest boarding loss time occurred, reaching 10 seconds. Box plots depicting boarding loss times under different rainfall conditions are shown in *Figure 7*.

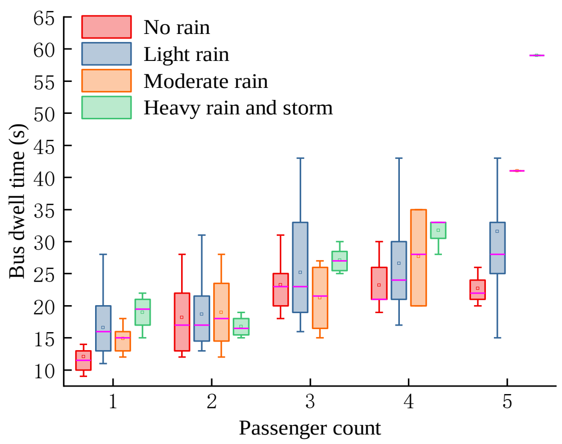



*Figure 6 – Parking time box diagram*

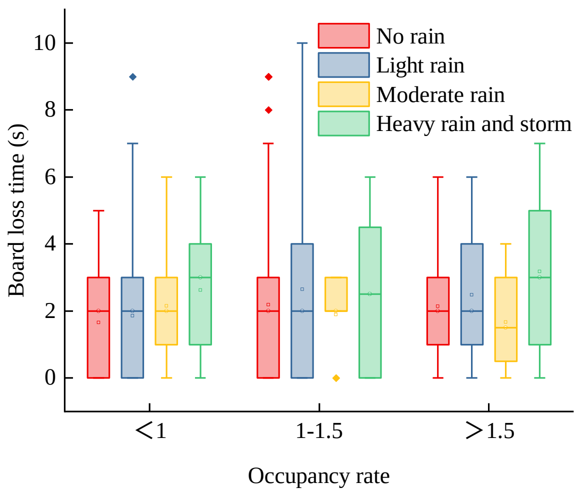


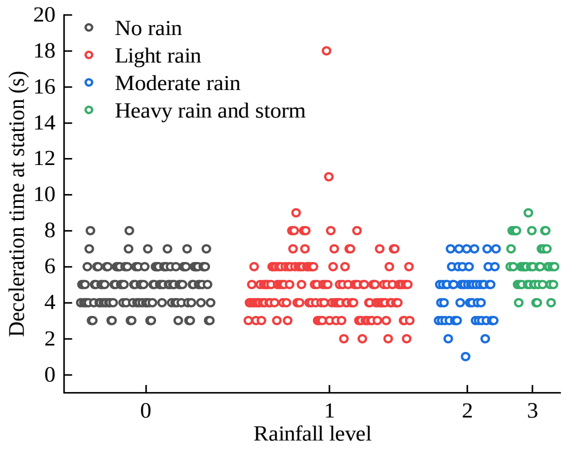

*Figure 7 – Box plot of boarding loss time and load rate under different rainfall levels*

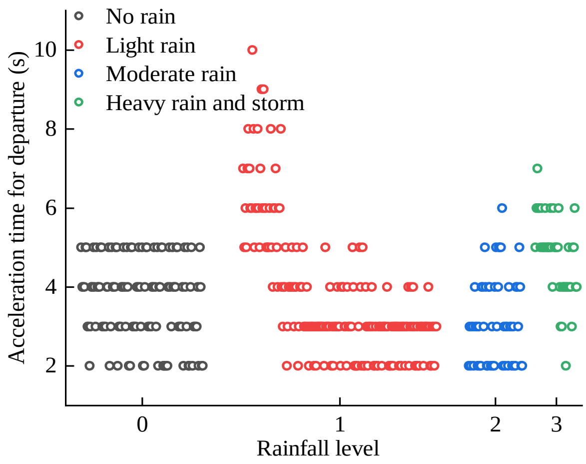
Bus stops with a uniform capacity of two berths were selected, excluding samples with boarding loss times of 0 seconds. The descriptive statistical analyses of boarding loss times under different rainfall conditions are presented in *Table 4*. The mean boarding loss times for each rainfall level aligned closely with the existing research findings [20]. The coefficients of variation ranged from 0.54 to 0.64, indicating moderate variation with no significant dispersion.

*Table 4 – Descriptive statistics of the boarding loss time*

| Rainfall level | Sample size | Mean | Max | Min | Standard deviation | Med | CV |
|---|---|---|---|---|---|---|---|
| 1 | 102 | 2.92 | 7 | 1 | 1.59 | 2 | 0.54 |
| 2 | 46 | 2.59 | 6 | 1 | 1.53 | 2 | 0.59 |
| 3 | 18 | 2.72 | 8 | 1 | 1.74 | 2 | 0.64 |

The acceleration departure time and deceleration arrival time both exhibited positive correlations of 0.17 and 0.24, respectively, with the rainfall level. As the rainfall level increased, bus speeds during arrival and departure decreased, leading to longer acceleration departure and deceleration arrival times. A scatter plot depicting the deceleration arrival and acceleration departure times under different rainfall levels is shown in *Figure 8*. Under no-rain conditions, the deceleration arrival times were concentrated in the 3–6 seconds range, whereas the acceleration departure times were concentrated in the 2–5 seconds range. Under light rain conditions, deceleration arrival times were cantered in the 3–6 seconds range, with acceleration departure times concentrated in the 2–8 seconds range. Under moderate rain conditions, the deceleration arrival times were in the 3–7 seconds range, and the acceleration departure times were concentrated in the 2–6 seconds range. Under heavy rain and storm conditions, deceleration arrival times were cantered at approximately 5–6 seconds, with acceleration departure times concentrated in the 4–6 seconds range. The deceleration arrival and acceleration departure times increased as the rainfall level increased.




*(a)* *(b)*

*Figure 8 – Scatter plot of deceleration arrival time and acceleration departure time under different degrees of rainfall: a) scatter plot of deceleration arrival times; b) scatter plot of acceleration departure times*

*Test of differences*

A one-way ANOVA was conducted to test the differences in bus dwell times under various rainfall conditions and determine whether bus dwell times differ under different rainfall levels. The sample sizes for different rainfall conditions and the number of boarding passengers are listed in *Table 5*. The p-value was used to assess statistical significance, with a significance level of 0.05. P-values of lower than this threshold indicate a significant difference; otherwise, no significant difference exists [18].

First, a difference test was performed on the dwell times between rainy and non-rainy conditions, considering the potential variations owing to different boarding passenger counts. Outliers in the samples were removed before testing. As some sample sizes for specific boarding passenger counts were small, the test was not conducted for these cases. *Table 6* presents the results. The p-values for boarding passenger counts of 0, 3, 4, 5, 6, 7, 8 and 9 were all less than 0.05, indicating significant differences in dwell times between the rainy and non-rainy conditions.

Next, a difference test was conducted for rainy conditions under three rainfall levels (Level 1: light rain, Level 2: moderate rain, and Level 3: heavy rain). Outliers were removed and the test was performed only for boarding passenger counts in the 0–5 individual range because of the limited sample size for Level 2 (moderate rain) with boarding passenger counts in the 0–5 individual range. *Table 7* presents the results. For boarding passenger counts of 3–5 individuals, no significant difference in dwell times was observed among the three rainfall levels ($p > 0.05$). However, a significant difference was observed ($p < 0.05$) for boarding passenger counts in the 0–2 individual range. This difference was attributed to the dispersed nature of passengers. The arrival of the first passenger at the bus stop and the time taken to open the umbrellas during boarding contributed to the extended boarding time.

## 5. Establishing prediction models

This study established a historical database of 288 data samples from bus stops under rainfall conditions in Shenyang City from April to September. Factors affecting the dwell time were taken as input features, including rainfall level, boarding passengers, alighting passengers, occupancy rate, peak/off-peak status and workday/non-workday status, constituting a total of six features. Peaks were denoted as “1” and off-peaks as “0”; workdays as “1” and non-workdays as “0”; weather conditions were represented by 0, 1, 2 and 3 for clear weather, light rain, moderate rain and heavy rain, respectively. Occupancy rates less than 1 were denoted as “1”, between 1 and 1.5 as “2” and greater than 1.5 as “3”. The target object for the prediction was dwell time.

*Table 5 – Number of samples of passengers in different rainfall levels*

| **Boarding passenger count** | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| 0 | 2 | 36 | 13 | 2 |
| 1 | 2 | 32 | 23 | 8 |
| 2 | 5 | 29 | 8 | 5 |
| 3 | 14 | 24 | 6 | 7 |
| 4 | 9 | 32 | 3 | 3 |
| 5 | 14 | 15 | 1 | 4 |
| 6 | 20 | 5 | 0 | 6 |
| 7 | 13 | 5 | 0 | 2 |
| 8 | 11 | 6 | 0 | 1 |
| 9 | 8 | 5 | 0 | 1 |
| 10 | 8 | 2 | 0 | 0 |
| 11 | 6 | 0 | 0 | 1 |
| 12 | 2 | 0 | 0 | 0 |
| 13 | 2 | 1 | 0 | 0 |
| 14 | 1 | 0 | 0 | 0 |
| 15 | 2 | 0 | 0 | 1 |
| 16 | 2 | 0 | 0 | 0 |
| 17 | 0 | 0 | 0 | 1 |

*Table 6 – Bus dwell-time difference test based on two weather conditions: no rain and rain*

| **Boarding passenger count** | ***P*** |
|---|---|
| 0 | 0.04 |
| 3 | 0.01 |
| 4 | 0.05 |
| 5 | 0.04 |
| 6 | 0.04 |
| 7 | 0.03 |
| 8 | 0.004 |
| 9 | 0.002 |

*Table 7 – Bus dwell-time difference test based on three rainfall classes (1, 2, 3)*

| **Boarding passenger count** | ***P*** |
|---|---|
| 0 | 0.03 |
| 1 | 0.01 |
| 2 | 0.09 |
| 3 | 0.13 |
| 4 | 0.21 |
| 5 | 0.21 |

The dataset from the Shenyang Architecture University Bus Station, with the highest sample quantity (202), was used as the training set, whereas data from the other bus stations (86 samples) served as the testing set. The influence of different feature dimensions was eliminated, the determination of the optimal solution was expedited and the data were normalised according to *Equation 3* [21] before training to ensure prediction accuracy:

$$x_{new} = \frac{x - Min}{Max - Min} \tag{3}$$

where, $x_{new}$ represents the normalised x data; *Max* and *Min* denote the maximum and minimum values in the original dataset.

Owing to significant fluctuations in the survey data, the BP neural network model, known for its nonlinear mapping capabilities [15], was employed in this study to predict bus stop times under rainfall conditions. Additionally, the SVM and KNN prediction models were used for the comparative analysis.

## 5.1 Configuration of the prediction models

*Configuration of BP neural network model*

— Input and output-layer configuration: This study established a historical database using 288 data samples from bus stops during rainfall events in Shenyang from April to September. The dwell time was selected as the prediction target. Factors affecting the dwell time, including rainfall level, boarding passengers, alighting passengers, occupancy rate, peak hours and working days, were taken as input features. Therefore, the number of input-layer neural nodes was set to six. The encoding for different factors was as follows: "1" for peak hours, "0" for non-peak; "1" for working days, "0" for non-working days; the weather was coded as 0, 1, 2 and 3 for good weather, light rain, moderate rain and heavy rain, respectively; "1" for occupancy rates less than 1, "2" for occupancy rates between 1 and 1.5, and "3" for occupancy rates greater than 1.5 [22]. The prediction target was the dwell time and the output-layer neural nodes were set to one.

— Function and parameter settings: The BP neural network employed a rectified linear unit (ReLU) function, which is known for its nonlinearity, computational efficiency and gradient-vanishing alleviation advantages. The expression is given by Equation 4 [23]. The training employed a backpropagation algorithm for weight adjustment. The hyperparameters were set as follows: alpha = 3 and learning rate = 0.01. These values were iteratively refined for optimal penalty parameters, learning rates and number of iterations, with a maximum of 1,000 iterations.

$$f(x) = max(0, x) \tag{4}$$

— Hidden layer configuration: A small-scale hidden layer (one or two layers) was chosen owing to the limited sample size. The number of nodes in a single hidden layer was estimated using empirical formulas (*Equations 5, 6, 7, and 8*) [22], resulting in an estimated value between 4 and 13:

$$H = \sqrt{MN} \tag{5}$$

$$H = \sqrt{M + N} + \alpha \tag{6}$$

$$H = \sqrt{0.43MN + 0.12N^2 + 2.54M + 0.77N + 0.35} + 0.51 \tag{7}$$

*H* should satisfy the following condition:

$$\sum_{i=0}^{H} C_H^i > P \tag{8}$$

where $M$ denotes the number of nodes in the input layer, $H$ is the number of nodes in the hidden layer, $N$ is the number of nodes in the output layer, $\alpha$ is a constant in the 1–10 range and $P$ is the number of samples.

The experiments revealed that when the number of nodes was 13, the mean squared error was minimised, resulting in a value of 0.67, as shown in *Figure 9*.

However, the effectiveness based on the graph was still not optimal. Therefore, an incremental method was employed for the setup and experimentation with the double hidden layer. The incremental method gradually increases the number of new nodes to optimise the network structure [24]. After multiple experiments, the optimal number of nodes for the double hidden layers was determined to be 30 and 25, with an $R^2$ of 0.71.

L2 regularisation was introduced to prevent overfitting and improve the generalisation ability of the BP neural network model. L2 regularisation involves adding the sum of the squares of the weight parameters (L2 norm) to the loss function. This is done to balance the fit of the model to the training data and its generalisation to the prediction data, thus reducing the risk of overfitting. The L2 regularisation weight was set to 0.01.

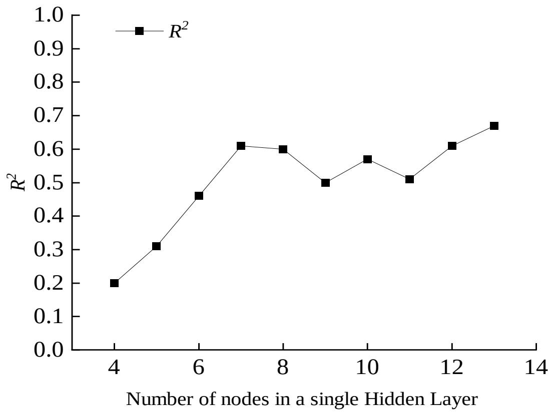



*Figure 9 – Mean square error of each node*

*Configuration of the SVM model*

The SVM algorithm selects the kernel function by setting the kernel parameter to Rbf. The expression for the Rbf function is given in *Equation 9* [25].

$$K(x,z) = (y < x,z > +r)^d \tag{9}$$

The penalty coefficient $C$ was set to 1.0, the degree (degree of the polynomial kernel) was set to 3 and gamma was set to “scale” so that the gamma values are automatically scaled according to the data. The default values for tol and epsilon were 1e-3 and 0.1, respectively.

*Configuration of the KNN model*

For the KNN model settings [26], the distance metric uses the Euclidean distance: neighbours were set to five, leaf size was set to 30, and p was set to two. The Euclidean distance expression is given in *Equation 10*.

$$d(X,Y) = \sqrt{\sum_{i=1}^{|X|} (X_i - Y_i)^2} \tag{10}$$

where $X$ is the training set samples, $Y$ is the sample to be predicted, $|X|$ is the number of training set samples, $X_i$ is the $i^{th}$ feature for sample $X$ and $Y_i$ is the $i^{th}$ feature for sample $Y$.

### 5.2 Model prediction results and evaluation

Four metrics were employed to evaluate the predictive performance of the three models: the mean squared error (*MSE*), mean absolute error (*MAE*), mean bias error (*MBE*) and coefficient of determination ($R^2$) [27]. $R^2$ reflects the relative contribution of the regression, with a value closer to one indicating a better-fitting regression model. Lower values of *MSE*, *MAE* and *MBE* imply lower prediction errors. *Table 8* presents a comparison of the prediction results of the two models. As summarised in the table, the BP prediction model exhibited lower values for *MSE*, *MAE* and *MBE* than the SVM and KNN models, with a larger $R^2$ value under the same conditions of training and testing samples, indicating superior predictive performance and fewer errors. The comparative plots of the predicted values for each sample are shown in *Figure 10*.

*Table 8 – Comparison of evaluation indicators between the three models*

| | SVM | KNN | BP |
|---|---|---|---|
| *MSE* | 109.49 | 66.74 | 43.46 |
| *MAE* | 6.56 | 5.67 | 4.47 |
| *MBE* | 3.71 | 2.22 | 0.39 |
| $R^2$ | 0.32 | 0.58 | 0.73 |

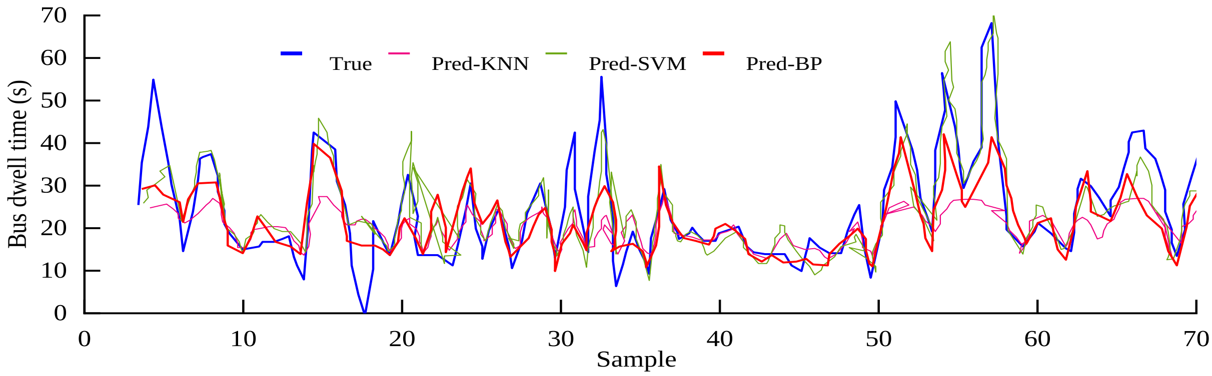



*Figure 10 – Comparison of the SVM, KNN, and BP models*

### 5.3 Model optimisation

To enhance the performance of the BP neural network model and identify the optimal parameters, the GA was integrated into the model, resulting in a GA-BP prediction model. The BP neural network optimised by the GA can effectively avoid the local minimum defects of the original BP neural network and has a high convergence speed and high accuracy [28]. The GA evaluates the fitness of individuals using a BP neural network to calculate the fitness of each individual in the training dataset. Fitness, measured by using an error function, reflects the accuracy of an individual's prediction of the target output [29]. Based on the fitness evaluation, a subset is selected as the parent generation for the next population. Parent chromosomes are selected for crossover to generate a new offspring. Mutation, crossover and other genetic processing methods are used to evolve the offspring chromosomes. The process is repeated until a new population is generated. The optimised weights and thresholds calculated by the GA are used as the initial parameters for the BP neural network [30, 31].

By using this approach, GA can search for the optimal solution in the parameter space of the BP neural network, identifying the neural network structure and weight configuration that best address the specific problem. After multiple experiments, optimal results were achieved with a population size of 30, a crossover probability of 0.8, a mutation probability of 0.2, and an iteration period of 200, as illustrated in the technical roadmap in *Figure 11* [32].

### 5.4 GA-BP model prediction results and evaluation

Following the same methodology as in the previous section, the dataset from the Shenyang Jianzhu University Bus Stop, with the most samples (202 samples), was employed as the training set. Data from other bus stops, totalling 86 samples, constituted the test set. A comparative analysis of the prediction results of the BP and GA-BP models is shown in *Figure 12*.

Evaluation metrics for both models is presented in *Table 9*, which shows the following:

1) The $R^2$ value for the GA-BP neural network was 0.77, exhibiting the highest $R^2$ value between the two models.
2) The *MSE*, *MAE* and *MBE* metrics were comparatively smaller for the GA-BP model, indicating that the prediction performance was better aligned with real-world scenarios.

Consequently, the GA-BP model effectively forecasted bus dwell times under rainy conditions. The loss function curve for the GA-BP model is shown in *Figure 13*. The loss curve rapidly declined during the early training stages and gradually stabilised as the training progressed, indicating that model convergence with minimal loss was achieved on the training data, reflecting a favourable performance.

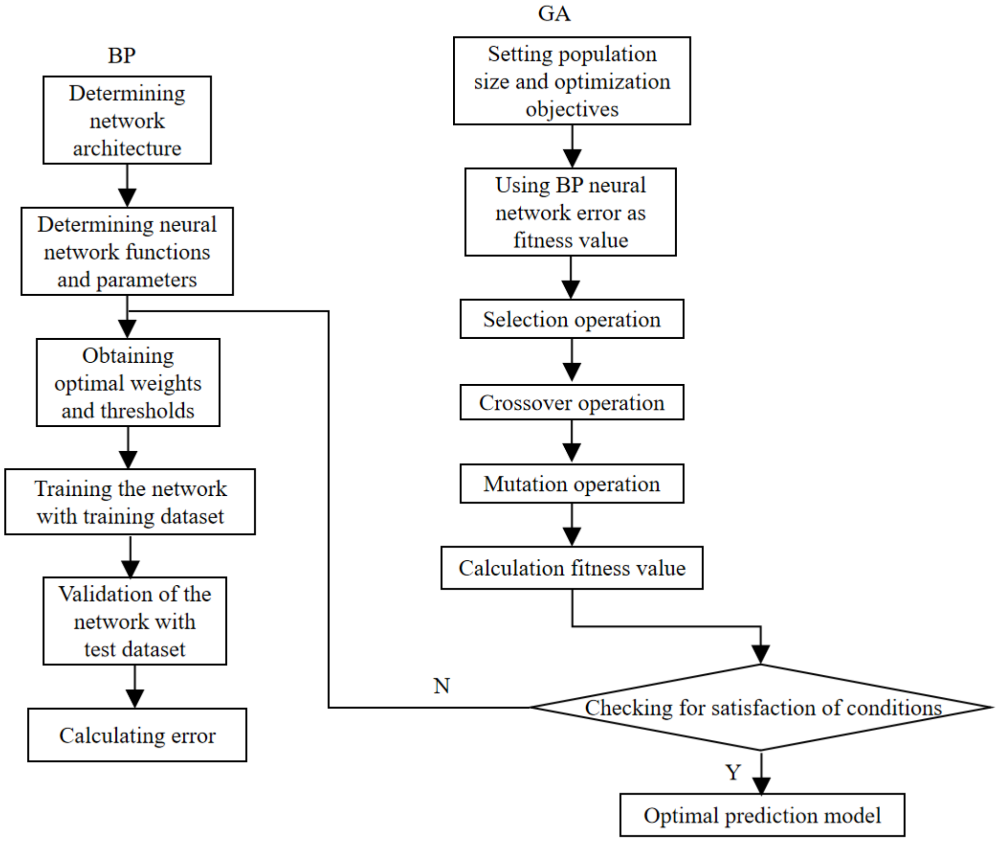



*Figure 11 – GA-BP prediction model roadmap*

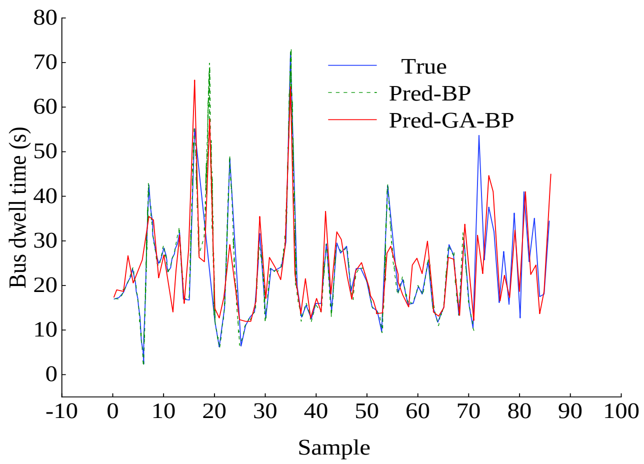



*Figure 12 – Prediction results of the BP and GA-BP models*

*Table 9 – Comparison of evaluation indicators between the BP and GA-BP models*

| Evaluation metrics | BP | GA-BP |
|---|---|---|
| *MSE* | 45.20 | 38.21 |
| *MAE* | 4.81 | 4.32 |
| *MBE* | -0.84 | -0.94 |
| $R^2$ | 0.71 | 0.77 |

This study did not explore the influence of factors such as different bus stop types and berth capacities owing to data limitations.

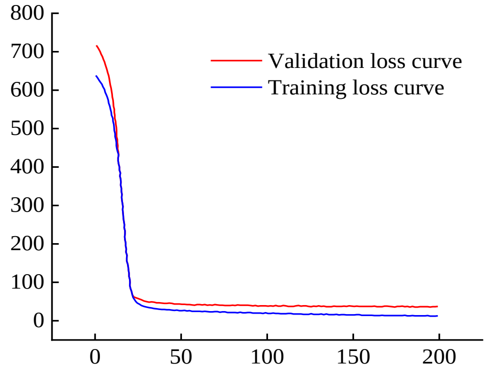



*Figure 13 – GA-BP model fitness map*

## 6. CONCLUSIONS

This study investigates the impact of rainfall levels on various parameters during bus dwell times, including boarding and alighting passenger counts, boarding loss times, occupancy rates and station approach and departure times. This study revealed the influence of various factors on bus dwell times under varying degrees of rainfall, leveraging observational surveys. A predictive model for bus dwell times under rainy conditions was established and optimised. The key findings are summarised as follows:

1) A positive correlation exists between bus dwell times under rainy conditions and rainfall levels. Significant differences in dwell times between rain and no-rain conditions were observed for the same boarding passenger counts. As rainfall levels increased, the probability of passengers using umbrellas also increased. For passenger counts in the 0–2 individual range, the increased dispersion of passengers led to longer bus dwell times for the same boarding counts.
2) The bus dwell time under rainy conditions was most influenced by boarding and alighting passenger counts. Rainfall influences travel behaviour and choices, resulting in reduced passengers as rainfall intensifies. The bus dwell times increased with higher rainfall levels for the same boarding passenger count. Boarding loss time correlated positively with rainfall levels and internal occupancy rates, increasing with higher rainfall levels and occupancy rates. Accelerated departure and decelerated approach times exhibited positive correlations with rainfall levels, with longer times observed under heavy rain conditions.
3) A BP neural network prediction model was constructed to forecast bus dwell times under rainy conditions. SVM and KNN prediction models were employed for comparison. Under identical training and testing conditions, the BP prediction model demonstrated superior performance with minimal errors. Building upon this, a GA-BP prediction model was developed by optimising the BP neural network parameters using the GA. This enhanced model more accurately predicted the bus dwell times under rainy conditions, thereby facilitating improved scheduling during inclement weather conditions.

The primary contribution of this work is considering rainfall levels in the study of bus dwell times and the development of predictive models. The study findings will facilitate bus punctuality and improve customer appeal for bus services. Owing to data limitations, this study did not explore the influence of factors such as different bus stop types and berth capacities. Future research can include more in-depth analyses of bus dwell times under rainy conditions as the data becomes more comprehensive.

## ACKNOWLEDGMENTS

This research was supported by the Foundation of the Liaoning Province Education Administration (grant numbers LJKMZ20220934 and LJKZZ20220080).

## REFERENCES

[1] KFH Group. Transit Capacity and Quality of Service Manual. *Transportation Research Board*. Washington DC, 2013.

[2] Fernandez R. Modelling public transport stops by microscopic simulation. *Transportation Research Part C-Emerging Technologies*. 2010;18(6):856-868. DOI: 10.1016/j.trc.2010.02.002.

[3] Kostyniuk LP, D'Souza CR. Effect of passenger encumbrance and mobility aid use on dwell time variability in low-floor transit vehicles. *Transportation Research Part A-Policy and Practice*. 2020;132:872-881. DOI: 10.1016/j.tra.2020.01.002.

[4] Bergel-Hayat R, Debbarh M, Antoniou C, Yannis G. Explaining the road accident risk: Weather effects. *Accident Analysis & Prevention.* 2013;60:456-465. DOI: 10.1016/j.aap.2013.03.006.

[5] Bocker L, Dijst M, Prillwitz J. Impact of everyday weather on individual daily travel behaviours in perspective: A literature review. *Transport Reviews*. 2013;33(1):71-91. DOI: 10.1080/01441647.2012.747114 .

[6] Tirachini A. Bus dwell time: The effect of different fare collection systems, bus floor level and age of passengers. *Transportmetrica A: Transport Science*. 2013;9(1):28-49. DOI: 10.1080/18128602.2010.520277.

[7] Milkovits MN. Modeling the factors affecting bus stop dwell time: Use of automatic passenger counting, automatic fare counting, and automatic vehicle location Data. *Transportation Research Record*. 2008;2072(1):125-130. DOI: 10.3141/2072-13.

[8] Al-Jumailey SI. Effect of operation policies on dwell time. *2011 International Conference on Electric Technology and Civil Engineering (ICETCE)*. 2011:1194-1196. DOI: 10.1109/ICETCE.2011.5776416.

[9] Ji Y, et al. How does a static measure influence passengers' boarding behaviors and bus dwell time? Simulated evidence from Nanjing bus stations. *Transportation Research Part A-Policy and Practice*. 2018;110:13-25. DOI: 10.1016/j.tra.2018.02.003.

[10] Bladikas A, Tsai FM, Chien SI. Evaluation of bus travel time and schedule adherence under adverse weather. *The 90th Transportation Research Board Annual Meeting*, 2009.

[11] Stover V, McCormack E. The impact of weather on bus ridership in Pierce County, Washington. *Journal of Public Transportation*. 2012, 15(1): 95-110.

[12] Glick TB, Figliozzi MA. Analysis and application of log-linear and quantile regression models to predict bus dwell times. *Transportation Research Record*. 2019;2673(10):118-128. DOI: 10.1177/0361198119848701.

[13] Ding J, Yang M, Cao Y, Kong SL. Dwell time prediction of bus rapid transit using arima-svm hybrid model. *Applied Mechanics and Materials*. 2014;587-589:1993-1997. DOI: 10.4028/www.scientific.net/amm.587-589.1993.

[14] Rashidi S, Ranjitkar P, Hadas Y. Modeling bus dwell time with decision tree-based methods. *Transportation Research Record*. 2014;2418(1):74-83. DOI: 10.3141/2418-09.

[15] Liu Y. An improved AHP and BP neural network method for service quality evaluation of city bus. *International Journal of Computer Applications in Technology*. 2018;58(1):37-44. DOI: 10.1504/ijcat.2018.094060.

[16] Femandez R. An expert system for the preliminary design and location of high capacity bus-stops. *Traffic Engineering & Control*. 1993;34(11):533-539.

[17] Avanzato R, Beritelli F, Raspanti A, Russo M. Assessment of multimodal rainfall classification systems based on an audio/video dataset. *International Journal on Advanced Science, Engineering and Information Technology*. 2020;10:1163-1168. DOI: 10.18517/ijaseit.10.3.12130.

[18] González EM, Romana MG, Álvaro OM. Bus dwell-time model of main urban route stops: Case study in Madrid, Spain. *Transportation Research Record*. 2012;2274(1):126-134. DOI: 10.3141/2274-14.

[19] Yue Y, Yeh AG. Spatiotemporal traffic-flow dependency and short-term traffic forecasting. *Environment and Planning B: Planning and Design*. 2008;35(5):762-771. DOI: 10.1068/b33090.

[20] Novales M, et al. Increasing boarding lost time at regular bus stops during rainy conditions: A case study. *Journal of Public Transportation*. 2021;23(1):63-80. DOI: 10.5038/2375-0901.23.1.4.

[21] Panda SK, Jana PK. A multi-objective task scheduling algorithm for heterogeneous multi-cloud environment. *International Conference on Electronic Design, Computer Networks & Automated Verification*, 2015. DOI: 10.1109/EDCAV.2015.7060544.

[22] Khanlou HM, et al. Prediction and optimization of electrospinning parameters for polymethyl methacrylate nanofiber fabrication using response surface methodology and artificial neural networks. *App Neural Computing and Applications*. 2014;25:767-777. DOI: 10.1007/s00521-014-1554-8.

[23] Leshno M, Lin VY, Pinkus A, Schocken S. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Network*. 1993;6(6):861-867. DOI: 10.1016/s0893-6080(05)80131-5.

[24] Islam MM, et al. A new adaptive merging and growing algorithm for designing artificial neural networks. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 2009;39(3):705-722. DOI: 10.1109/ijcnn.2008.4634073.

[25] Hao PY, Chiang JH, Chen YD. Possibilistic classification by support vector networks. *Neural Network*. 2022;149:40-56. DOI: 10.1016/j.neunet.2022.02.007.

[26] Zhang S, Li J. KNN classification with one-step computation. *IEEE Transactions on Knowledge and Data Engineering*. 2021:1-1. DOI: 10.1109/tkde.2021.3119140.

[27] Wei W, et al. Prediction of tool wear based on GA-BP neural network. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*. 2022;236(12):1564-1573. DOI: 10.1177/09544054221078144.

[28] Qi A. Research on prediction model of improved BP neural network optimized by genetic algorithm. *Journal of Engineering Research*. 2018;150:764-767. DOI: 10.2991/macmc-17.2018.145.

[29] Liu CY, et al. Application of GA-BP neural network optimized by grey Verhulst model around settlement prediction of foundation pit. *Geofluids*. 2021:1-16. DOI: 10.1155/2021/5595277.

[30] Yi XD. Selection of initial weights and thresholds based on the genetic algorithm with the optimized back-propagation neural network. In: *12th international conference on fuzzy systems and knowledge discovery (FSKD)*, Zhangjiajie, China, 15–17 Aug 2015; IEEE Publications: New York, 2015; p. 173-177. DOI: 10.1109/fskd.2015.7381935.

[31] Zhang S, Huo Z, Zhai C. Building carbon emission scenario prediction using STIRPAT and GA-BP neural network model. *Sustainability*. 2022;14(15):9369. DOI: 10.3390/su14159369.

[32] Liu D, Liu C, Tang Y, Gong C. A GA-BP neural network regression model for predicting soil moisture in slope ecological protection. *Sustainability*. 2022;14(3):1386-1386. DOI: 10.3390/su14031386.

孙宝芸，杨亚平，董雷，鲁泓麟，汪子皿

**降雨条件下公交停靠时间特征与动态预测**

**摘 要：**

探究公交停靠时间受降雨天气的影响程度有利于合理制定雨天情况下的公交调度管理方案。尽管已经提出了许多数学模型，但现有模型无法满足公交政策制定的精确性。本文以沈阳市直线型公交停靠站为研究对象，根据现场调查数据，对不同降雨程度下公交停靠特征及影响因素进行分析。用 Pearson 相关分析方法和 SPSS 软件，揭示上下车人数、降雨等级、泊位数、承载率和信号交叉口的存在等参数对降雨条件下公交停靠时间的影响程度。建立支持向量机、k 最近邻算法和反向传播预测模型，并利用遗传算法对预测效果最好的 BP 神经网络模型进行优化。GA-BP 预测模型比 BP 神经网络模型预测结果精度更高，可用于预测降雨条件下的公交停靠时间。这项研究的结果将有助于提高公交车的准点率，并提高公交服务的吸引力。

**关键词：**

降雨条件；公交停靠特征；影响因素；相关性；GA-BP 模型