

Article

# Exploring the Pedestrian Route Choice Behaviors by Machine Learning Models

Cheng-Jie Jin <sup>1,2,\*</sup>, Yuanwei Luo <sup>1,2</sup>, Chenyang Wu <sup>3,4</sup>, Yuchen Song <sup>1,2</sup> and Dawei Li <sup>1,2</sup>

<sup>1</sup> Jiangsu Key Laboratory of Urban ITS, Southeast University of China, Nanjing 210096, China

<sup>2</sup> Jiangsu Province Collaborative Innovation Center of Modern Urban Traffic Technologies, Nanjing 210096, China

<sup>3</sup> School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, China; cywu@nwpu.edu.cn

<sup>4</sup> Urban System Lab, Imperial College London, London SW7 2AZ, UK

\* Correspondence: 101011900@seu.edu.cn

**Abstract:** To investigate pedestrian route choice mechanisms from a perspective distinct from that employed in discrete choice models (DCMs), this study utilizes machine learning models and employs SHapley Additive exPlanations (SHAP) for model interpretation. The data used in this paper come from several pedestrian flow experiments with two routes, which were recorded by UAV. Our findings indicate that logistic regression (similar to a binary logit model) exhibits good computational efficiency but falls short in predictive accuracy when compared to other machine learning models. Among the 12 machine learning models assessed, by calculating the new indicator named OP, we find that eXtreme Gradient Boosting (XGB) and Light Gradient Boosting (LGB) strike the best balance between accuracy and computational efficiency. Regarding feature contribution, our analysis reveals that bottlenecks exert the most significant influence on pedestrian route choice behavior, followed by the time it takes pedestrians to return from the end of the route to the origin (reflecting pedestrian characteristics and attitudes). While the pedestrian density of the shorter route contributes less compared to bottlenecks and return time, it exhibits a threshold effect, meaning that once the density of the shorter route surpasses a certain threshold, most pedestrians opt for the longer route.

**Keywords:** route choice; pedestrian; machine learning; SHapley Additive exPlanations



**Citation:** Jin, C.-J.; Luo, Y.; Wu, C.; Song, Y.; Li, D. Exploring the Pedestrian Route Choice Behaviors by Machine Learning Models. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 146. <https://doi.org/10.3390/ijgi13050146>

Academic Editors: Maria Antonia Brovelli and Wolfgang Kainz

Received: 23 February 2024

Revised: 15 April 2024

Accepted: 26 April 2024

Published: 28 April 2024



**Copyright:** © 2024 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

Walking is not only a common human behavior, but also a travel mode. Complex transportation network nowadays provides diverse routes for pedestrians to choose, and they may opt for different routes based on their preferences and trip objectives. Exploring pedestrian route choice behavior and revealing their tactics may help improve transportation planning, traffic control, and the construction of walking facilities. As a result, the study of pedestrian route choice remains a significant research topic [1,2].

Many researchers have employed discrete choice models (DCMs) to analyze pedestrians' route choice behavior [3]. Despite the advantages of DCMs in analyzing factors driving a certain decision process, their low behavior prediction accuracy remains an evident drawback. With the development of computer technologies, machine learning models are widely used in various research fields because of their good prediction performance. In recent years, some researchers introduced several explainable machine learning frameworks, such as Local Interpretable Model-Agnostic Explanations (LIME) [4], Individual Conditional Expectation (ICE) [5], and SHapley Additive exPlanations (SHAP) [6] to improve the interpretability of machine learning models. Among them, SHAP is a game theory-based, model-agnostic, unified approach that can measure feature importance, thus making it widely used. Therefore, it is possible to explore the mechanisms of pedestrian route choice and predict pedestrian choices by explainable machine learning models.

To address the research gap, we conducted pedestrian route choice experiments to collect pedestrians' route choice behavior under different circumstances. We then employed several machine learning models to explore the underlying mechanisms of pedestrian route choice behaviors, and compare the predictive performance of the models. Finally, we utilized SHAP to understand feature importance and the mechanism of each feature's influence on pedestrian route choice. The results presented in this paper can offer a new approach for researchers and practitioners in analyzing the underlying mechanism of pedestrian route choice decisions. Although only two routes are considered in this study, such a scenario represents a fundamental condition for understanding the essence of route choice behavior. Therefore, studying this scenario can serve as a starting point for subsequent investigations.

The remainder of this article is structured as follows. Section 2 presents a review of the studies about pedestrian route choice behavior studied by machine learning models. Some related fields, including pedestrian behavior studied by DCMs and vehicle behavior studied by machine learning models, are also discussed. Section 3 describes the pedestrian route choice experiments and the data obtained from the experiments. Section 4 introduces the machine learning models used, including the features and the hyper-parameters. Section 5 discusses the results of multiple machine learning models in various datasets, including the prediction accuracies and speeds. Section 6 interprets the results of these machine learning models by SHAP. Finally, the conclusions and future works are given in Section 7.

## 2. Literature Review

### 2.1. Discrete Choice Models in Pedestrian Route Choice Analysis

Pedestrian route choice behavior has traditionally been analyzed through discrete choice models (DCMs), which typically assume that users are utility maximizers [7]. For example, King and Bode [8] found that pedestrians avoid busier and farther destinations. Haghani and Sarvi [9] focused on the structure of the multinomial logit (MNL) model, and found that the effect of decision rule specification was not as important as hypothetical bias. At the same time, other principles, such as random regret minimization, have also been used by some researchers [10].

The basic assumption of the MNL model is the independence of irrelevant alternatives (IIA) [11]. However, pedestrians do not always conform to this assumption. To address this issue, other models with more complex structures, such as mixed logit models [12] and latent class models [13], have been employed to account for the strong heterogeneity in pedestrian route choice behavior. Haghani and Sarvi [14] studied the egress behavior of pedestrians in crowded, complex, and confined spaces. They found that in nonemergency situations, the proximity of the exit leading to the passenger's destination emerged as the dominant factor influencing route choice. However, during emergency scenarios, pedestrians placed a much higher priority on avoiding crowded exits. Additionally, generalized linear models (GLMs) have also found applications in this field. Tong and Bode [15–17] utilized GLMs to analyze data collected from virtual reality-based pedestrian route choice experiments. After tracking pedestrians' continuous decision-making, they observed that pedestrians tend to assign diminishing value to environmental information [15], but put higher values on shorter distance, fewer turns, and fewer accumulated angle changes [16]. Additionally, they tend to become followers if other virtual pedestrians opt for a specific exit [17].

Generally speaking, researchers have identified multiple factors affecting pedestrian route choices, which can be categorized into route attributes, environment factors, and socio-demographic factors [18]. In terms of route attributes, studies have found that within limited spatiotemporal scales (such as inside a building or a metro station), pedestrians tend to favor routes that are short, wide, with low pedestrian density, and higher walking speeds [19–21].

## 2.2. Machine Learning Models in Pedestrian Route Choice Analysis

In recent years, the widespread adoption of machine learning models for route choice behavior analysis has gained traction due to their superior predictive capabilities. The earliest application of machine learning models in route choice analysis dates back to 2002, when Yamamoto et al. [22] studied the route choice behavior on Japanese highways using a decision tree approach. Barua et al. [23] explored the viability of using Support Vector Machine (SVM) and Neural Network (NN) for route choice modeling, with the findings indicating that SVM achieved an accuracy rate of 70.86%, surpassing NN, which scored 68.34%. Many researchers have also compared machine learning models with DCMs in this field, and consistently found that machine learning models outperform DCMs regarding both accuracy and computing efficiency [24–26].

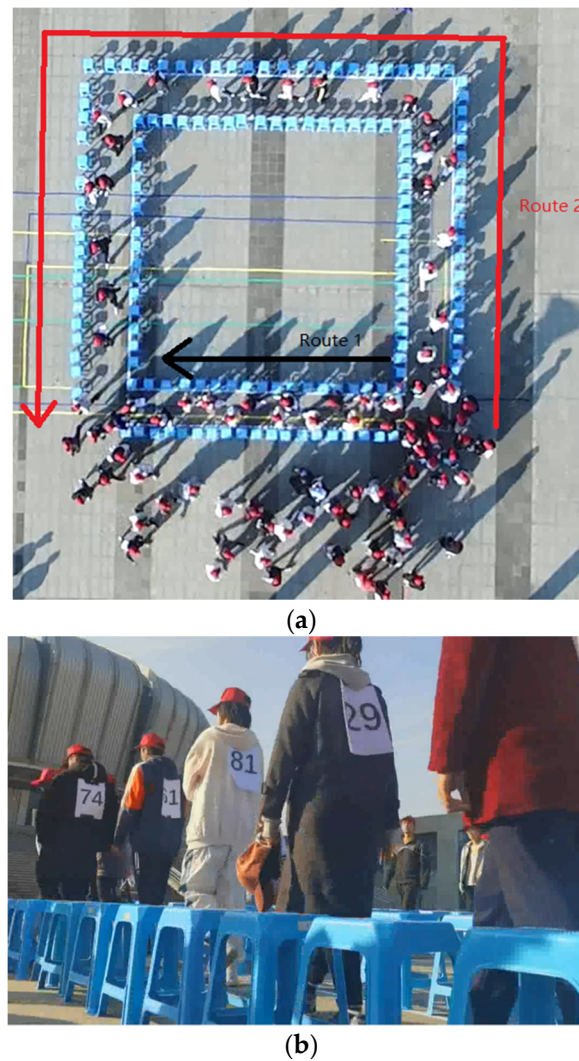
Nevertheless, in contrast to studies on vehicle route choice behavior, there are relatively fewer investigations into pedestrian route choice behavior employing machine learning models. Yuen et al. [27] demonstrated the feasibility of Multi-Layer Perception (MLP) in replicating passenger choice behavior between two escalators. Wang et al. [28] conducted a comparison of several machine learning models, with Artificial Neural Network (ANN) emerging as the top performer in terms of predictive accuracy and stability. SVM exhibited high accuracy but lacked training efficiency, while the predictive performance of K-Nearest Neighbors (KNN) was directly associated with data complexity. Building on these findings, Zhou et al. [29] employed machine learning models to simulate pedestrian movement during evacuations, where Gradient-Boosted Decision Trees (GBDT) and SVM demonstrated superior predictive performance among the available models [30].

While machine learning models excel in prediction, interpreting their results in a manner akin to DCMs remains a challenge for many machine learning models (e.g., MLP). To bridge this interpretability gap, the SHAP framework has gained prominence in recent transportation research, encompassing areas such as electric vehicle charging choice behavior [31], traffic safety analysis [32], vehicle route choice [33], etc. However, its application in the analysis of pedestrian route choice behavior remains relatively limited.

## 3. Data

There are two types of studies that focused on pedestrian route choice behavior at different spatiotemporal scales. Some focused on a large spatiotemporal scale, where pedestrians move in urban road network and can travel as far as several kilometers. Others center their attention on pedestrian movement within smaller spatiotemporal scales, such as inside buildings or metro stations, with a particular emphasis on evacuations during emergency situations [34,35]. Due to the difference in spatiotemporal scale, the data sources used in these studies often differ, with the former type of studies typically making use of GPS data [36–39] and the latter predominantly relying on video data. In this paper, we focus on pedestrians moving within a small spatiotemporal scale during non-emergency conditions, so video data is employed in this study.

To capture pedestrians' route choice behavior, we conducted two pedestrian route choice experiments in the Jiulonghu Campus of Southeast University in China in both 2020 and 2021. The participants in these experiments were exclusively university students, aged between 18 and 25 (Due to the pandemic of COVID-19, the access to our university campus was limited to only students and staff between 2020 and 2022. Therefore, it was very difficult to recruit enough non-student pedestrians at the time of conducting these experiments. In the future, we will try to conduct more experiments with elderly people.). The experimental routes were delineated using plastic stools, as illustrated in Figure 1a. To record the experiments comprehensively, we employed an unmanned aerial vehicle (UAV) equipped with a camera. The captured video footage was recorded at a rate of 25 frames per second, with a resolution of  $2704 \times 1520$  pixels. Additionally, we positioned two stationary cameras to monitor and track all participants, as shown in Figure 1b. Each participant was identifiable through unique numbers displayed on their attire. They all wore distinctive red caps to facilitate their identification in the experimental video recordings.



**Figure 1.** The configuration of Run 11. (a) The top view from the UAV; (b) The view from one fixed camera near the shorter route.

In each experiment session, participants were required to complete multiple runs. During each run, pedestrians were instructed to walk at their normal speed for 5 to 6 laps. Each pedestrian made a single route choice at the origin and was required to walk all the way to the destination. Upon reaching the destination, they retraced their path to the origin and commenced a new lap. The way from the origin to the destination was confined by the arrangement of plastic stools. However, the return journey was unconstrained, allowing participants to move freely when heading back from the destination.

In this research, we call the shorter route as Route 1 (indicated by the black arrow in Figure 1a) and the longer one as Route 2 (indicated by the red arrow in Figure 1a). The configurations for each run are outlined in Table 1. To set the experimental parameters, we utilized a D-efficient design approach. Here, L1 and L2 represent the lengths of Route 1 and Route 2, respectively, while W1 and W2 denote their respective widths. The variable DC stands for Distance Control: a value of 1 indicates that each pedestrian was required to maintain a social distance of 1 m from other pedestrians during the experiment, whereas a value of 0 signifies the absence of distance control. BT represents the average waiting time at bottleneck 1, which signifies the presence of a flow-limited bottleneck on Route 1, and each pedestrian needed to wait for a specific duration before passing the bottleneck. In most bottleneck experiments, the waiting time at the bottleneck remained constant (DBT = 0). However, in Run 17, the bottleneck waiting time was made a random variable

with a 50% probability of either 2 s or 8 s of waiting time. Consequently, DBT was set at 3 s for this particular run.

**Table 1.** Design variables of the pedestrian route choice experiments.

Run	Year	Lap	DC	BT (s)	DBT (s)	L <sub>1</sub> (m)	L <sub>2</sub> (m)	W <sub>1</sub> (m)	W <sub>2</sub> (m)
1	2020	6	0	0	0	12	24	0.5	0.5
2	2020	6	0	5	0	12	24	0.5	0.5
3	2020	6	0	10	0	12	24	0.5	0.5
4	2020	5	0	0	0	11	11	1.0	0.5
5	2020	6	0	0	0	8	14	1.0	1.0
6/7	2020	6	0	0	0	8	14	0.5	1.0
8	2020	20	0	0	0	8	14	0.5	1.0
11	2021	6	0	0	0	8	24	1.0	1.25
12	2021	6	0	5	0	8	24	1.0	1.25
13	2021	6	0	0	0	14	18	0.5	1.25
14	2021	6	0	0	0	14	18	0.75	1.25
15	2021	6	0	0	0	11	18	0.75	1.0
16	2021	6	0	0	0	11	21	0.5	0.75
17	2021	8	0	5	3	12	24	0.5	0.5
21	2020	6	1	0	0	12	24	0.5	0.5
22	2020	6	1	5	0	12	24	0.5	0.5
24	2020	5	1	0	0	11	11	1.0	0.5
25	2020	6	1	0	0	8	14	1.0	1.0
26	2020	6	1	0	0	8	14	0.5	1.0
31	2021	5	1	0	0	8	24	1.0	1.25
32	2021	5	1	5	0	8	24	1.0	1.25
33	2021	5	1	0	0	14	18	0.5	1.25
34	2021	5	1	0	0	14	18	0.75	1.25
35	2021	5	1	0	0	11	18	0.75	1.0
36	2021	5	1	0	0	11	21	0.5	0.75

We manually extracted several features from the video recordings, as detailed in Table 2. The average pedestrian density was calculated as the number of pedestrians on a route divided by the product of the route's length and width. The travel time of a pedestrian was determined by subtracting the time of entry into the origin from the time of departure from the route. Similarly, the return time of a pedestrian was calculated by subtracting the time of reaching the destination from the time of returning to the origin. The dummy variable 'Cro' was derived from observations indicating whether the origin was crowded, while the dummy variable 'Pair' was based on observations of whether the pedestrian was accompanied by others.

**Table 2.** Features extracted from the video recordings.

Features	Explanations
D <sub>1</sub> /D <sub>2</sub>	Average pedestrian density on Route 1/Route 2
T <sub>1</sub> /T <sub>2</sub>	Travel time from origin to destination on Route 1/Route 2
BN	Number of pedestrians waiting at the bottleneck
RT	Return Time from destination to origin
Cro	Whether the origin is crowded (Yes:1, No:0)
Pair	Whether the pedestrian is moving with others (Yes:1, No:0)

During the experiments, pedestrians were asked to choose the routes repeatedly. However, the data of the first lap were excluded from the following study, due to the absence of a stable state for pedestrian dynamics. We finally reached a sample size of 7697 runs, with 3290 collected in 2020 and 4407 collected in 2021. In this paper, three different datasets



are studied, including two datasets from different years and the combination of them, as presented in Table 3. It reveals a noteworthy observation that the proportion of choosing Route 1 in all the datasets are similar (<50%). The possible explanation is that when both routes were congested during the experiments, the lengths and widths of routes became not important for route choices.

**Table 3.** The basic information of three datasets.

Name of Datasets	Year	Runs Involved	Sample Size	Proportion of Choosing Route 1
21	2021	Run 11–17, Run 31–36	3290	43.1%
20	2020	Run 1–10, Run 21–26	4407	42.9%
ALL	2020, 2021	All runs	7697	43.0%

#### 4. Model Parameters and Features

##### 4.1. Models and Metrics

Firstly, we introduce the 12 machine learning models employed in this study to analyze pedestrian route choice behavior, which are K-Nearest Neighbors (KNN), Logistics Regression (LR), Support Vector Machine (SVM), Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), Bagging, Gradient Boosting Decision Tree (GBDT), Adaboost (ADB), eXtreme Gradient Boosting (XGB), Light Gradient Boosting (LGB), and Multi-Layer Perception (MLP). All of them are simple and easy to use for everyone.

Next, the performance metrics used to evaluate the performance of the models are:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

$$\text{F1\_score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

where  $TP$ ,  $FP$ , and  $FN$  refer to true positive, false positive, and false negative in the classification, respectively.

Except for the three performance metrics introduced above, we also used receiver operating characteristic (ROC) curve and area under curve (AUC) to evaluate the machine learning models in Section 5. We performed a random split of the data into training and testing sets at a ratio of 7:3. This process was repeated ten times to calculate the average prediction metrics, mitigating the impact of sampling on the prediction results.

##### 4.2. Feature Selection

The dataset utilized in this paper comprises 17 features, which are presented in Tables 1 and 2 in Section 3. However, a challenge arises due to the nature of the data, particularly in relation to the travel times of pedestrians on Route 2. During each lap of the experiment, since pedestrians who opted for Route 1 did not traverse Route 2, it was impossible to record their travel times on Route 2. To address this limitation, we opted to replace the travel time on Route 2 with the average travel time of all pedestrians who selected Route 2 in each run. Essentially, pedestrians who chose Route 1 in each run were attributed the same value for  $T_2$ .

Nevertheless, this approach can introduce complications, especially when applying machine learning models, particularly those based on decision trees. For instance, when both  $T_1$  and  $T_2$  are employed, the models can yield highly unrealistic predictions, achieving an accuracy level of about 100%, as  $T$  serves as a direct indicator of route choice (pedestrians with the same  $T_2$  chose Route 1). To mitigate this issue, we excluded  $T$  from subsequent analyses.

In the following sections, we consider two scenarios for each dataset: (1) *full scenario* and (2) *lightweight scenario* (we call it LW for short in the tables). For the former one, we try to realize the best prediction performance by using many variables. For the latter one, we try to get a lightweight solution with only two variables, while the results are decent. The variables of the lightweight scenario consist of the first two variables with the largest contributions in the full scenario, which will be further discussed in Section 5.

#### 4.3. The Determination of Hyper-Parameters

Because of the different structures of models, the way of determining hyper-parameters varies by models. Taking the lightweight scenario (Here, only two variables (RT and  $D_1$ ) are used. More details of this scenario can be found in Section 5.1.) of Run 11 as an example, the hyper-parameters of models are determined by:

- (1) **Models with only one hyper-parameter.** RF, Bagging, GBDT, ADB, XGB, LGB, KNN, and NB belong to this type. We present the relationship between the numbers of estimators/ neighbors and the corresponding F1 scores in Figure 2. It can be seen that for the five tree-based models (except XGB in Figure 2f) and KNN, when the number of estimators/neighbors gradually increases and reaches the critical value (marked by the dashed line), the result becomes stable. As a larger number of estimators makes the time needed to measure the models longer (Figure 3), we chose the critical values in Figure 2a–g as the optimal values for hyper-parameters of the values. For NB, as its hyper-parameter is the prior distribution, we tested four typical distributions. The results in Figure 2h demonstrate that Gaussian (normal distribution) is the optimal choice for this dataset.

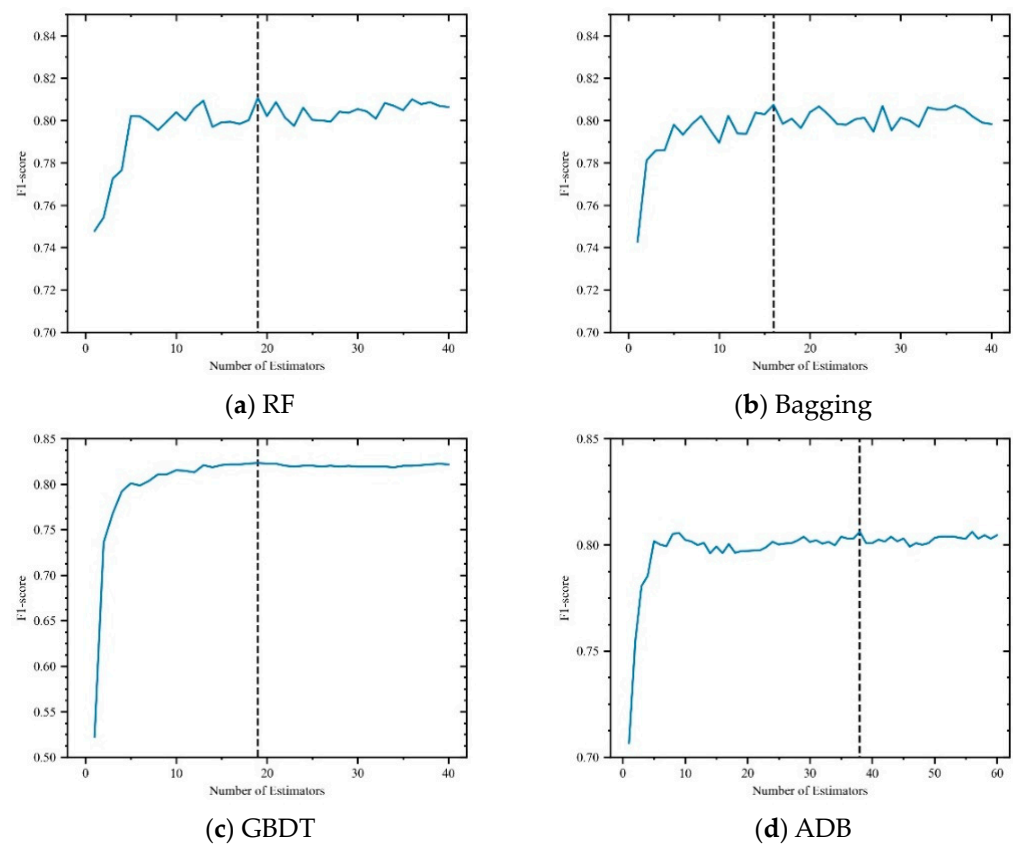
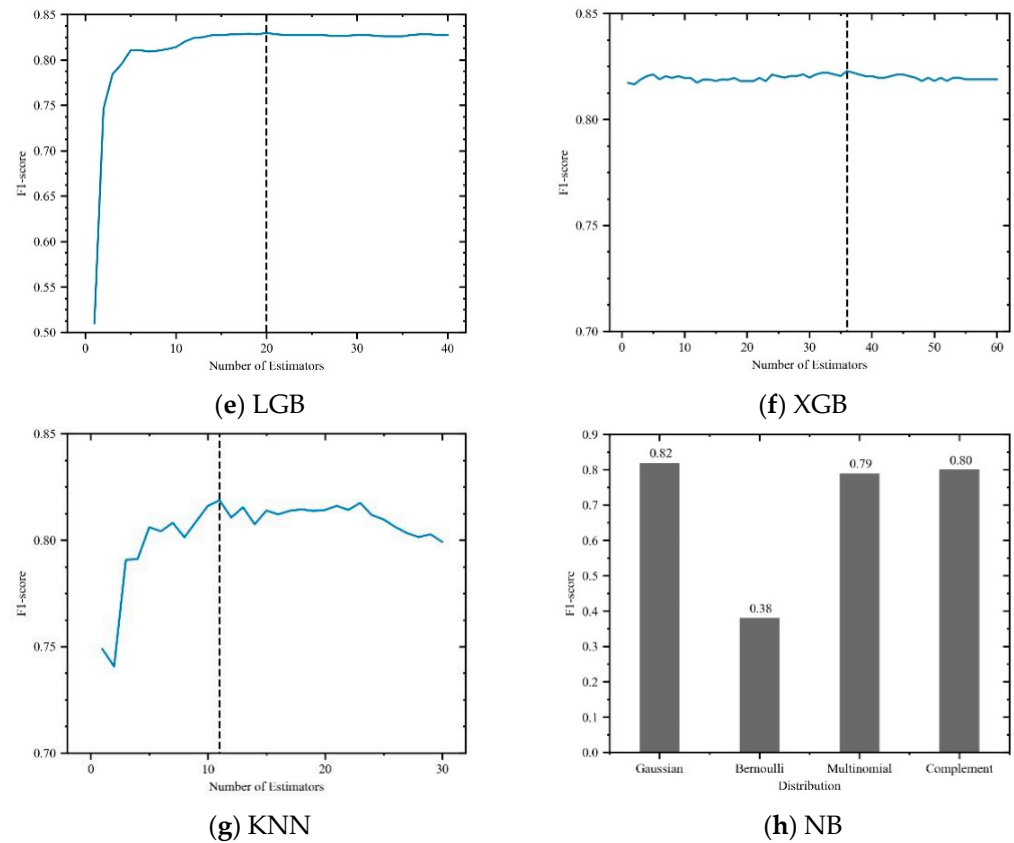
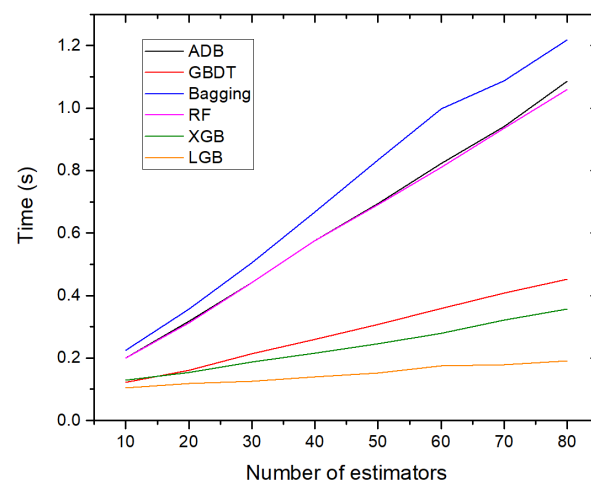


Figure 2. Cont.



**Figure 2.** The relationship between the possible parameters and the corresponding F1 scores in the lightweight scenario of Run 11.



**Figure 3.** The relationship between the time needed and the number of estimators in the lightweight scenario of Run 11.

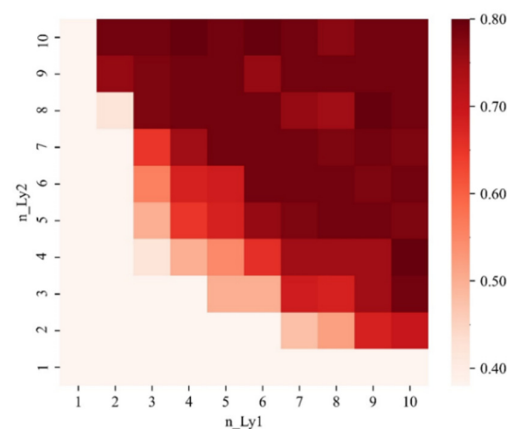
- (2) **Models with two hyper-parameters.** SVM and MLP models belong to this type. We compared the results from different combinations of hyperparameters to find the best combination. When we chose the kernel as “rbf”, the Gamma = 1 and C = 1 combination returns the best results (0.824) for the SVM model (see Table 4). The kernel “linear” is not chosen, as its F1 scores are always equal to 0.80 (lower than 0.824).



**Table 4.** The F1 scores of SVM in the lightweight scenario of Run 11.

Gamma/C	1	10	100	1000
1	0.824	0.816	0.822	0.812
0.1	0.817	0.822	0.823	0.823
0.01	0.799	0.815	0.816	0.813
0.001	0.432	0.801	0.802	0.816

For MLP, TensorFlow is used for network construction and Keras is used as the basic framework. The ReLU function is used as the activation function to ensure the non-linear fitting effect and fast convergence of the model. When two layers are considered, the effect of numbers of neurons is presented in Figure 4. We name the two hyper-parameters n\_Ly1 and n\_Ly2, respectively. Clearly, when the two hyper-parameters reach a critical value (~6), the prediction performance becomes relatively stable. Therefore, to balance the prediction performance and computational speed, we choose n\_Ly1 = 6 and n\_Ly2 = 6 in this case.

**Figure 4.** The F1 scores of MLP in the lightweight scenario of Run 11.

- (3) **Default values.** For LR and DT, the impact of the hyper-parameters is not significant in this study. Therefore, we opted for the default values.

In summary, all hyper-parameters used in this study are listed in Table 5, including the results of three full scenarios. It is easy to understand that the full scenarios are usually more complex than the corresponding lightweight scenarios. For example, for most tree-based models, the numbers of estimators in the full scenarios are larger. So are the values of n\_Ly1 and n\_Ly2 for MLP. On the contrary, for KNN, the numbers of neighbors in the full scenarios are smaller.

**Table 5.** The hyper-parameters used in this paper.

Model	Type	Hyper-Parameters	Run 11 (LW)	Run 11 (Full)	Dataset 21 (Full)	Dataset ALL (Full)
RF	1	n_estimator	19	59	109	136
Bagging	1	n_estimator	16	63	136	175
GBDT	1	n_estimator	19	90	138	49
ADB	1	n_estimator	38	76	128	71
LGB	1	n_estimator	20	21	51	42
XGB	1	n_estimator	36	30	29	11
KNN	1	n_neighbor	11	7	12	24
NB	1	prior distribution	Gaussian	Gaussian	Gaussian	Gaussian
SVM	2	C, gamma	1, 1	1, 1	1, 1	10, 1
MLP	2	n_Ly1, n_Ly2	6, 6	10, 10	10, 10	20, 20
LR	3	Penalty	L2	L2	L2	L2
DT	3	criterion, max depth	gini, 4	gini, 4	gini, 4	gini, 4

## 5. Results

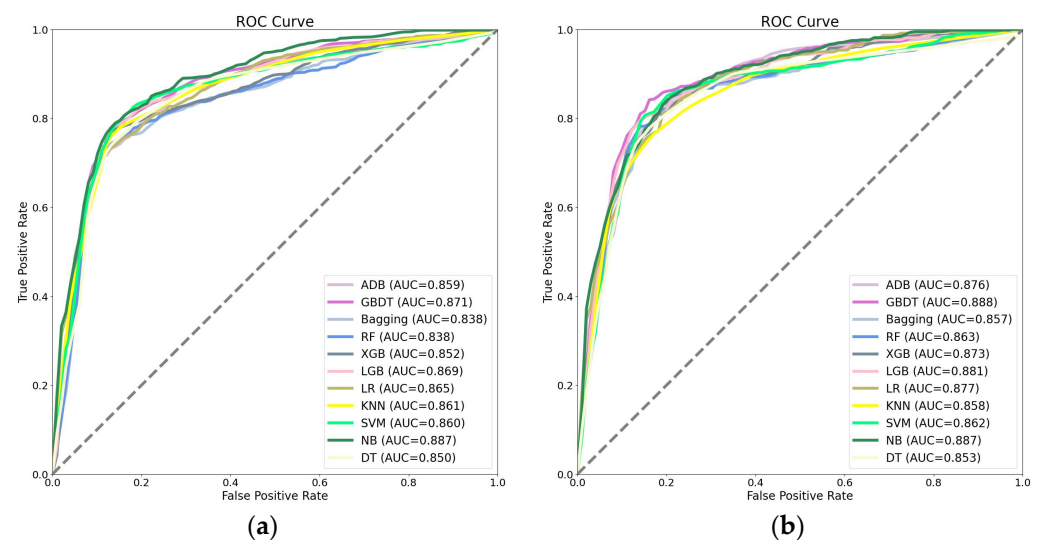
### 5.1. Results of Individual Runs

Firstly, we analyzed a representative experiment (Run 11) where no additional regulations were implemented. In this dataset, Year, DC, BT, DBT,  $L_1$ ,  $L_2$ ,  $W_1$ ,  $W_2$ , and BN are fixed. Consequently, only six available variables—Lap, Pair, Cro, RT,  $D_1$ , and  $D_2$ —are available.

Table 6 shows that all models achieved a prediction metric over 0.80, except NB. Interestingly, the models of the lightweight scenario employing only RT and  $D_1$  perform comparably to those of the full scenario. A possible explanation is that the models of full scenario may be overfitted due to the limited data in Run 11. The performance of GBDT and LGB rank top, but these two models have no significant advantage over others due to their complex structures. Additionally, the ROC curves for all models (shown in Figure 5) are similar, indicating that machine learning models have effectively utilized the features. Further enhancements in prediction performance may be difficult.

**Table 6.** The prediction metrics for Run 11.

Model	Precision (LW)	Precision (Full)	Recall (LW)	Recall (Full)	F1 Score (LW)	F1 Score (Full)
NB	0.82	0.79	0.82	0.76	0.82	0.75
LR	0.80	0.81	0.80	0.80	0.80	0.80
KNN	0.82	0.81	0.82	0.80	0.82	0.80
DT	0.81	0.81	0.81	0.81	0.80	0.81
Bagging	0.81	0.82	0.80	0.82	0.80	0.82
RF	0.81	0.82	0.81	0.82	0.81	0.82
ADB	0.81	0.82	0.81	0.82	0.81	0.82
XGB	0.83	0.83	0.83	0.83	0.83	0.82
MLP	0.80	0.83	0.80	0.83	0.80	0.83
SVM	0.83	0.83	0.82	0.83	0.82	0.83
GBDT	0.83	0.84	0.82	0.84	0.82	0.84
LGB	0.83	0.84	0.83	0.84	0.83	0.84



**Figure 5.** The ROC curves of Run 11. (a) Lightweight scenario; (b) Full scenario.

Next, we focus on the bottleneck experiments. In both Run 12 and Run 32, pedestrians were required to wait for 5 s at the bottleneck before proceeding. In Run 17, the waiting time was randomized to be either 2 or 8 s. The modeling results of these three experiments, as shown in Table 7, all demonstrate high accuracy. Nonetheless, the exceptional prediction performance observed may be misleading. In Run 12 and 32 (except for RF and Bagging),

the recalls of Route 2 are nearly 1.0, while those of Route 1 are nearly 0. This implies that these models predict that all the pedestrians would choose Route 2, which seems not reasonable. Such an unexpected outcome can be attributed to two factors:

- (1) The majority of pedestrians chose Route 2 in the three bottleneck experiments (84%, 77%, and 79% for Runs 12, 17, and 32, respectively). It is more difficult for the machine learning models to deal with such imbalanced datasets.
- (2) The feature dimension of the dataset used in this study is inadequate, with many features that are not operational (for instance, the width and length of routes in individual runs). This situation impedes the model's ability to interpret the pedestrian route choice principle in bottleneck experiments.

**Table 7.** The prediction metrics of bottleneck experiments in the lightweight scenarios.

Model	Run 12			Run 17			Run 32		
	Accuracy	Recall (Route 1)	Recall (Route 2)	Accuracy	Recall (Route 1)	Recall (Route 2)	Accuracy	Recall (Route 1)	Recall (Route 2)
Bagging	0.79	0.09	0.91	0.74	0.27	0.88	0.70	0.04	0.91
RF	0.79	0.10	0.92	0.74	0.28	0.88	0.70	0.01	0.92
DT	0.85	0.00	1.00	0.80	0.10	0.99	0.74	0.03	0.95
KNN	0.85	0.00	1.00	0.80	0.25	0.95	0.76	0.07	0.96
LR	0.85	0.02	0.99	0.78	0.25	0.93	0.75	0.05	0.97
XGB	0.84	0.02	0.99	0.79	0.22	0.95	0.75	0.01	0.97
ADB	0.84	0.02	0.99	0.79	0.22	0.94	0.77	0.00	1.00
SVM	0.85	0.00	1.00	0.81	0.18	0.98	0.77	0.00	1.00
NB	0.85	0.00	1.00	0.82	0.30	0.96	0.77	0.00	1.00
GBDT	0.85	0.02	0.99	0.80	0.16	0.98	0.77	0.00	1.00
LGB	0.85	0.00	1.00	0.80	0.20	0.97	0.77	0.00	1.00
MLP	0.85	0.00	1.00	0.82	0.18	0.98	0.77	0.00	1.00

In addition, it is worth noting that in many models for Run 17, the recalls of Route 1 are approximately 0.3, indicating that models accurately identify a few pedestrians choosing Route 1. This noticeable improvement can be attributed to the randomized wait time implemented at the bottleneck, which slightly impacts pedestrian behavior in the experiments, and enables models to extract valuable information from the dataset.

## 5.2. Results of Multiple Runs

In this section, we discuss the two datasets with multiple runs, including Dataset 21 and Dataset ALL. For simplicity, we only discuss the results of full scenarios, since those of lightweight scenarios are less predictive for larger datasets.

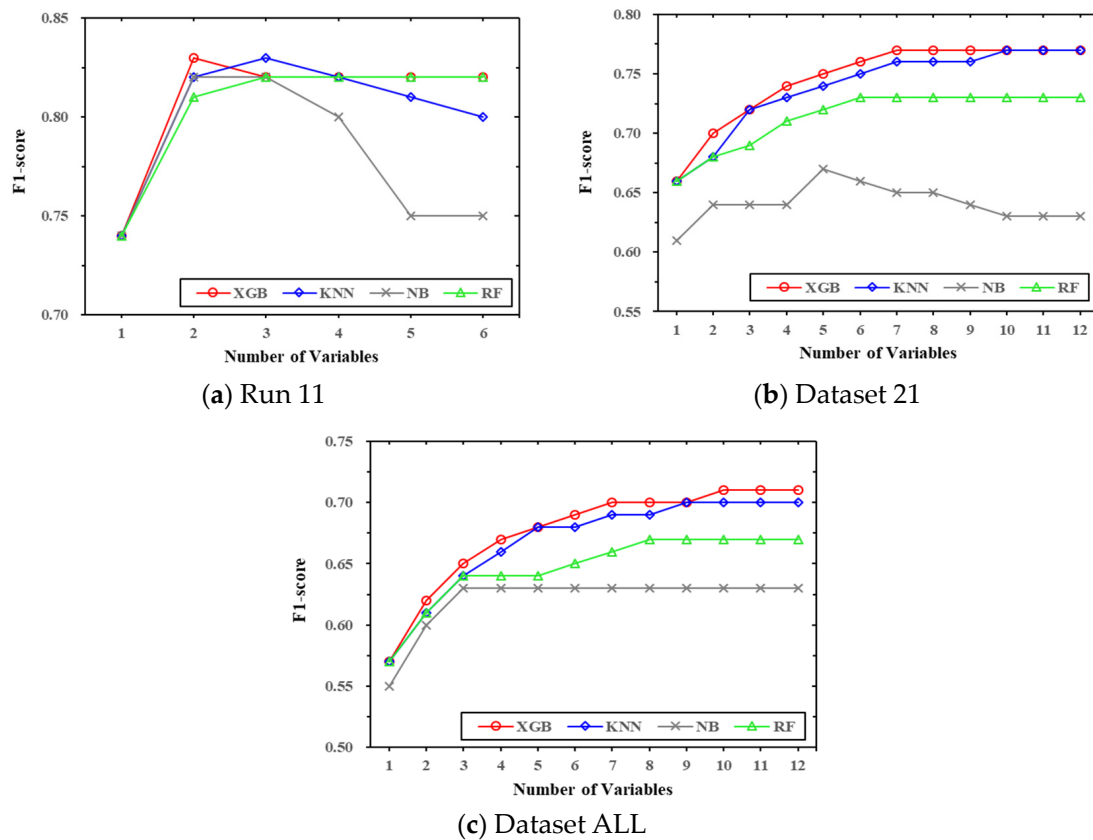
The relationship between the number of variables and the corresponding F1 scores in different datasets is shown in Figure 6. For Run 11 in Figure 6a, the results of all the typical models do not increase with the number of variables. On the contrary, except NB for Dataset 21, the other models for Dataset 21 and Dataset ALL have a clear growing trend in Figure 6b,c, which coincides with the empirical findings observed in the previous studies.

### (1) Dataset 21

Table 8 and Figure 7 present the prediction metrics of the models. Here, the full scenario includes RT, BT, D<sub>1</sub>, Cro, Pair, L<sub>1</sub>, L<sub>2</sub>, W<sub>1</sub>, and W<sub>2</sub>. In other words, the impacts of Lap, DC, BN, DBT, and D<sub>2</sub> are minor for this dataset. The possible explanations could be:

- BN indicates the queuing state on the shorter route. Its effect is similar to that of BT, which could be also replaced by BT.
- In most runs (except Run 17), there is always DBT = 0. In Run 17, the proportion of those choosing Route 1 is close to some other runs with the same BT, e.g., Run 2. From the statistical results, we can see that the influence of DBT is not significant.

- When making decisions, pedestrians usually focus on the situations of the shorter route (Route 1), rather than that of the longer one (Route 2). Such a tendency makes  $D_2$  not as important as  $D_1$ .

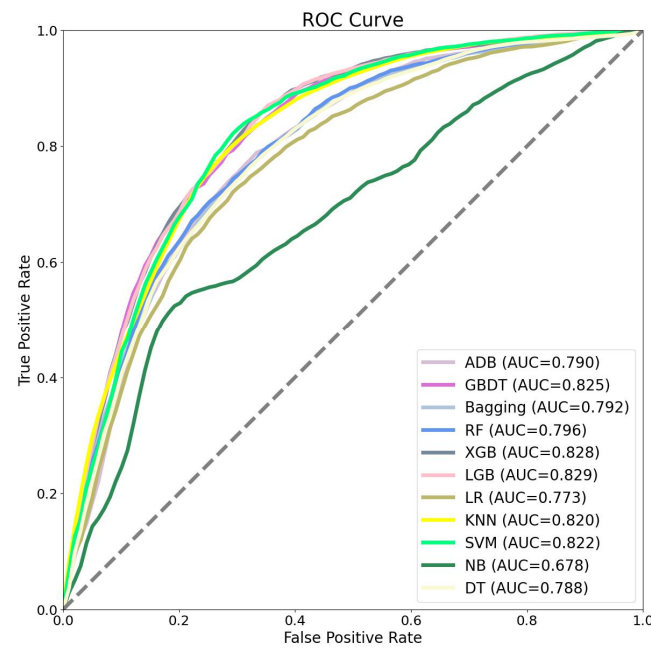


**Figure 6.** The relationship between the numbers of variables and the F1 scores in different datasets.

**Table 8.** The prediction metrics of Dataset 21.

Model	Precision	Recall	F1 Score
NB	0.66	0.64	0.64
LR	0.72	0.72	0.72
Bagging	0.73	0.73	0.73
RF	0.73	0.73	0.73
DT	0.73	0.73	0.73
ADB	0.73	0.73	0.73
KNN	0.76	0.76	0.76
MLP	0.77	0.77	0.76
GBDT	0.76	0.76	0.76
XGB	0.77	0.77	0.77
LGB	0.77	0.77	0.77
SVM	0.77	0.77	0.77

In Table 8, the top-performing models for this dataset are XGB, LGB, and SVM, achieving F1 scores of 0.75. In contrast, RF and Bagging exhibit relatively weaker performance. This is as expected, given that RF and Bagging models are good at extracting useful information and ensuring model stability rather than making accurate predictions. Meanwhile, NB is the worst-performance model, with an F1 score of only 0.64. This is due to its assumption that the data should adhere to normal distribution, which is different from our data where the distribution of BT is discrete.



**Figure 7.** The ROC curves of Dataset 21.

## (2) Dataset ALL

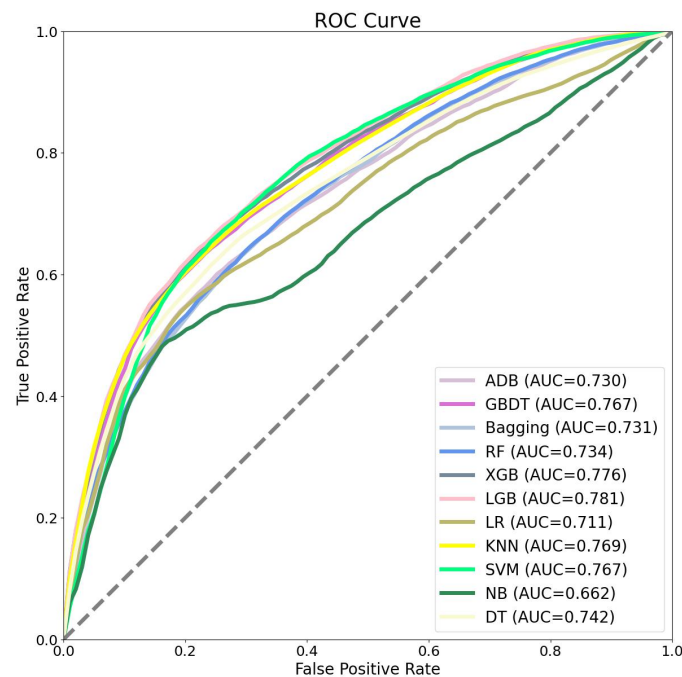
To further explore the generalization performance of the models, we expanded the data sample size by using Dataset ALL, which comprises experimental data from 2020 to 2021. The prediction metrics in Dataset ALL are shown in Table 9 and Figure 8. The results agree with those of Dataset 21 (Table 8 and Figure 7), where NB, Bagging, RF, and LR have lower metrics, while MLP, XGB, LGB, GBDT, and SVM perform better.

**Table 9.** The prediction metrics of Dataset ALL.

Model	Precision	Recall	F1 Score
NB	0.69	0.64	0.63
LR	0.65	0.65	0.65
RF	0.67	0.67	0.67
Bagging	0.67	0.67	0.67
DT	0.69	0.68	0.68
ADB	0.69	0.69	0.69
MLP	0.69	0.69	0.69
KNN	0.70	0.69	0.70
SVM	0.70	0.70	0.70
GBDT	0.70	0.69	0.70
XGB	0.71	0.70	0.70
LGB	0.71	0.71	0.71

However, it is worth noting that the prediction performance of all machine learning models has dropped compared to those in Dataset 21. For example, the top F1 score for all models in the full scenarios is 0.77 on Dataset 21. But in Dataset ALL, it decreases to 0.71. As Dataset ALL is composed of data collected over two years with more pedestrians involved, the increased heterogeneity of pedestrians may explain the drop in the prediction performance. Further research on how to incorporate pedestrian heterogeneity is worth investigating in future studies.





**Figure 8.** The ROC curves of Dataset ALL.

### 5.3. Model Validation and Evaluation

In Section 5.2, the training and testing data are sourced from the same dataset. The performance of these models on unknown datasets, however, is unclear. Therefore, Dataset 20 was utilized as a validation set in this section, and predictions are made using models trained on Dataset 21.

The prediction metrics in Table 10 align with the results based on Dataset 21 (Table 8): RF and Bagging exhibit lower metrics, and NB demonstrates the lowest performance. The prediction metrics decrease around 0.1 compared to the results based on Dataset 21, which may be also attributed to the absence of attributes reflecting an individual's heterogeneity in our models, as discussed in Section 5.2.

**Table 10.** The prediction metrics on validation dataset.

Model	Precision	Recall	F1 Score
NB	0.66	0.60	0.59
RF	0.62	0.62	0.62
Bagging	0.62	0.62	0.62
LR	0.62	0.62	0.62
DT	0.66	0.64	0.64
KNN	0.66	0.65	0.65
ADB	0.65	0.65	0.65
LGB	0.66	0.65	0.65
MLP	0.66	0.65	0.65
XGB	0.67	0.66	0.66
SVM	0.66	0.66	0.66
GBDT	0.66	0.65	0.66

Next, the speeds of different models are compared in Table 11, and the unit used is the second. We summed the training time and predicting time of the models. The observations are:

- (1) For most models, the time required is typically determined by the hyper-parameters used, such as the number of estimators for tree-based models.
- (2) The differences between models are clear: NB, DT, XGB, and LGB are faster than the other models, with NB being the fastest due to its simplicity.

- (3) RF and Bagging have longer computational times, while SVM and MLP require an extremely long time. SVM has a time complexity of about  $O(N^3)$ , where  $N$  is the size of the dataset [40]. For MLP, the time required is primarily determined by the structure of the considered neural networks. Therefore, the latter two models are not recommended for large datasets.

**Table 11.** The time used in different models and different scenarios.

Model	Run 11 (LW)	Run 11 (Full)	Dataset 21 (LW)	Dataset 21 (Full)	Dataset ALL (LW)	Dataset ALL (Full)
NB	0.10	0.10	0.11	0.13	0.14	0.15
DT	0.09	0.10	0.13	0.13	0.16	0.25
LGB	0.22	0.27	0.45	0.74	0.65	0.83
XGB	0.45	0.42	0.36	1.06	0.41	0.67
LR	0.16	0.24	0.22	2.01	0.53	1.73
KNN	0.18	0.20	0.78	1.01	1.93	2.87
GBDT	0.31	0.57	1.64	4.73	2.24	3.38
ADB	1.09	2.39	3.19	6.35	4.18	5.03
RF	0.59	1.78	2.31	7.48	4.36	14.06
Bagging	0.58	3.79	5.03	5.41	9.06	27.66
SVM	0.16	0.19	7.67	13.17	40.25	49.36
MLP	44.96	55.62	117.77	129.36	170.25	242.07

We evaluated the models based on a combined consideration of accuracy and efficiency. Initially, the time taken by each model was standardized to an efficiency value ranging from 0 to 1, with a higher value indicating greater efficiency. Subsequently, both the F1 score and the efficiency value were assigned equal weight, resulting in a new indicator named OP (Overall Performance). Table 12 reveals that XGB and LGB emerged as the top-performing models, effectively balancing accuracy and efficiency. In contrast, the performance of MLP and SVM was suboptimal, primarily attributed to their slower processing speeds.

**Table 12.** The OP results when considering both accuracy and efficiency.

Model	Run 11 (Full)	Dataset 21 (Full)	Dataset ALL (Full)
MLP	0.42	0.38	0.35
SVM	0.91	0.83	0.75
Bagging	0.88	0.84	0.78
RF	0.89	0.84	0.81
NB	0.87	0.82	0.81
LR	0.90	0.85	0.82
ADB	0.89	0.84	0.83
DT	0.90	0.86	0.84
GBDT	0.91	0.86	0.84
KNN	0.90	0.88	0.84
XGB	0.91	0.88	0.85
LGB	0.92	0.88	0.85

## 6. SHAP Explanations

To interpret the models introduced in Section 5, we employed the SHAP (Shapley Additive exPlanations) framework in this section. SHAP has three key properties: local accuracy, missingness, and consistency. The SHAP model is defined as follows:

$$f(x) = g(z') = \phi_0 + \sum_{i=1}^M \phi_i z'_i \quad (4)$$

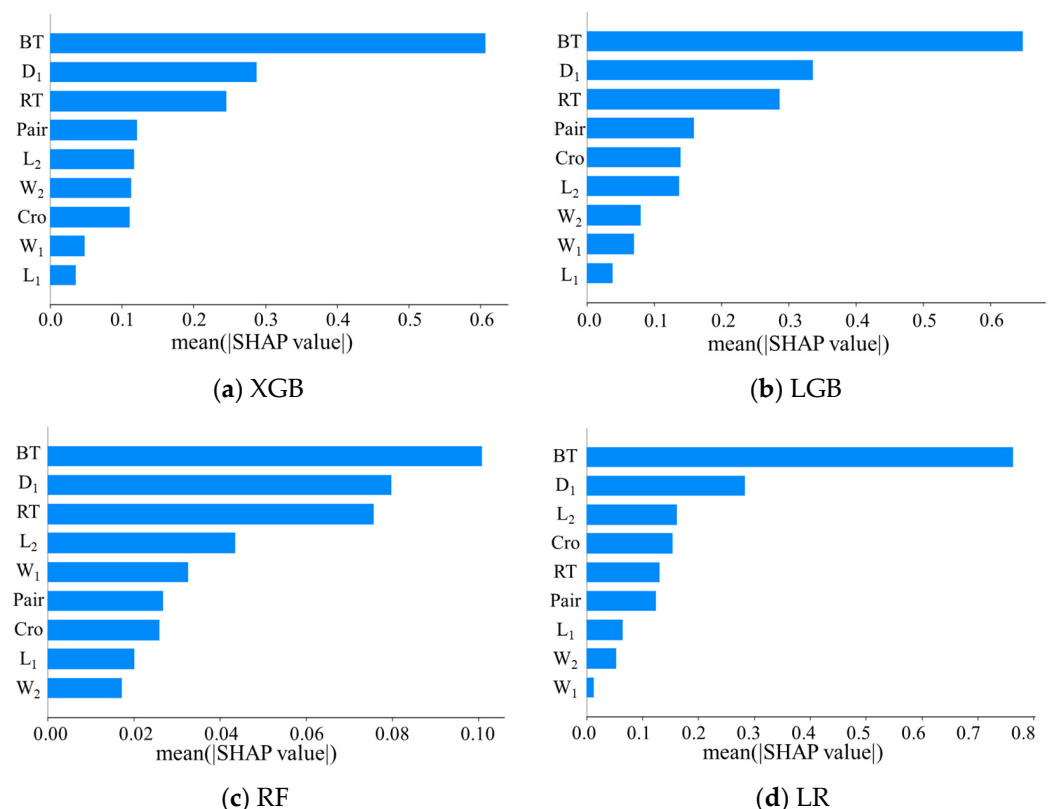
where  $g(z')$  denotes the SHAP explanation model,  $\phi_i$  represents the SHAP value, and  $z'_i \in \{0, 1\}$  is a binary variable. Two methods are commonly used to calculate SHAP values,

namely Kernel SHAP and Tree SHAP. Tree SHAP is specifically designed for tree-based models [41]. Therefore, in this paper, Tree SHAP was used for the calculation of GBDT, XGB, LGB, RF, Bagging, and DT, while the results of other models were estimated by Kernel SHAP.

### 6.1. The Global Importance of Features

In the SHAP framework, the contribution of each feature to the model is represented by the magnitude of the average absolute SHAP value. Typically, this information is visualized using bar charts to facilitate a comparison of the global importance of individual features.

Figure 9 shows the global importance of each feature for four representative models from Dataset ALL, including two models with good overall performance from Section 5 (XGB and LGB) and two others with lower prediction accuracy (The results of some other typical models (e.g., SVM, KNN, or MLP) are not presented in this section, since the speed of Kernel SHAP is extremely slow, needing several days to get the final results.). Across all models, BT emerges as the variable with the highest contribution, suggesting that the presence of a bottleneck on a route exerts the most significant impact on pedestrian route choices. The contribution of  $D_1$  ranks second, following by RT (except in the results of LR). Although the global importance of route attributes ( $L_1$ ,  $L_2$ ,  $W_1$  and  $W_2$ ) varies among different models, they generally contribute the least to the model. This again validates that the route attributes in our experiments do not align with the actual attractiveness of routes, as we mentioned in Section 3.



**Figure 9.** The global importance of each feature of Dataset ALL.

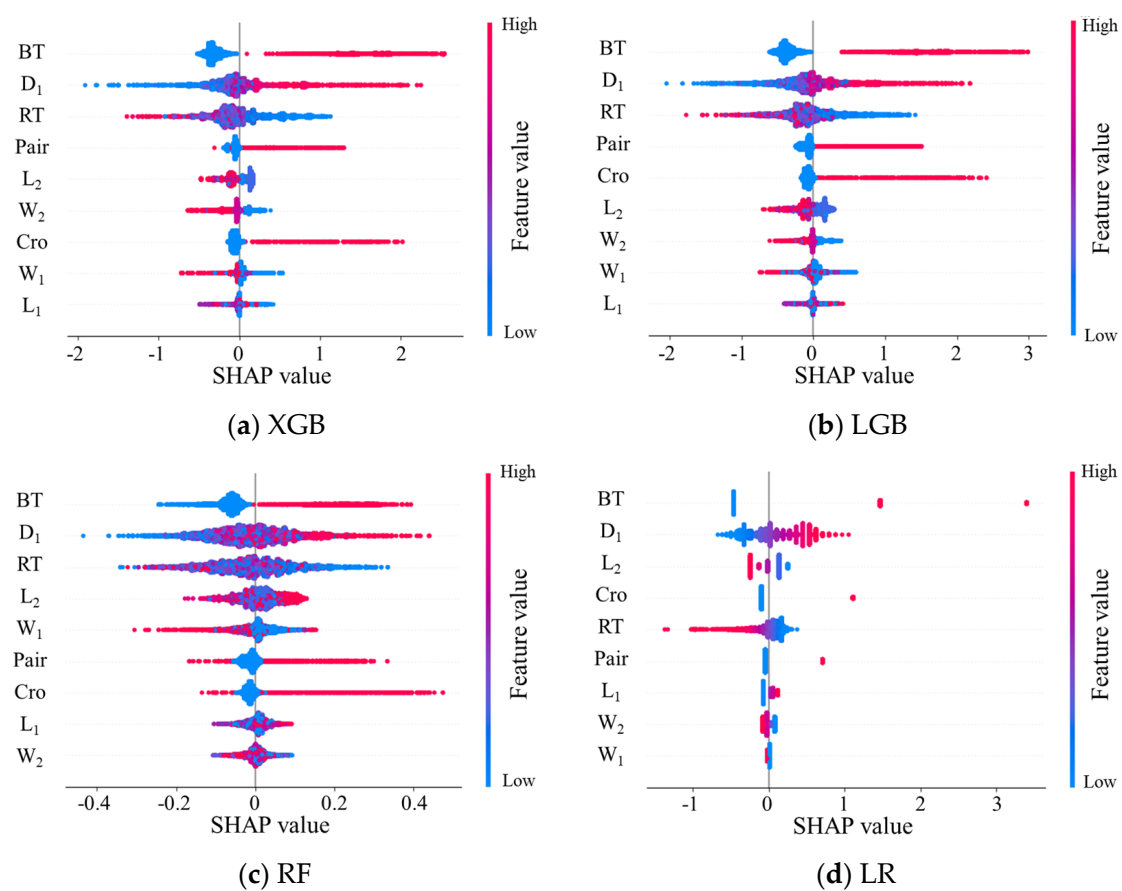
It is interesting to note how different models deal with these variables. XGB and LGB, which have comparable structures, assign similar contributions to most variables. For RF, the differences between the contributions of various variables are not large. However, LR places particularly high importance on BT compared to the other models, which could lead to relatively poor performance.

## 6.2. The Influence of Features

In Section 6.1, we discussed the magnitude of each feature's contribution to the models. However, feature importance alone does not reveal the specific influence of each feature on model predictions. In this section, we employ summary plots and dependence plots in SHAP to gain a deeper understanding of the mechanism underlying each feature's influence.

Figure 10 illustrates the impacts of key features on individual predictions. Each point on the summary plot represents an observation in the dataset, with the X-value indicating the SHAP value of a feature. The color of each point corresponds to the feature value of the observation. A higher, positive SHAP value indicates a greater likelihood of a pedestrian choosing Route 2, while a smaller, negative SHAP value indicates a higher possibility of choosing Route 1. The transition from blue to red signifies an increase in feature value. We can find that:

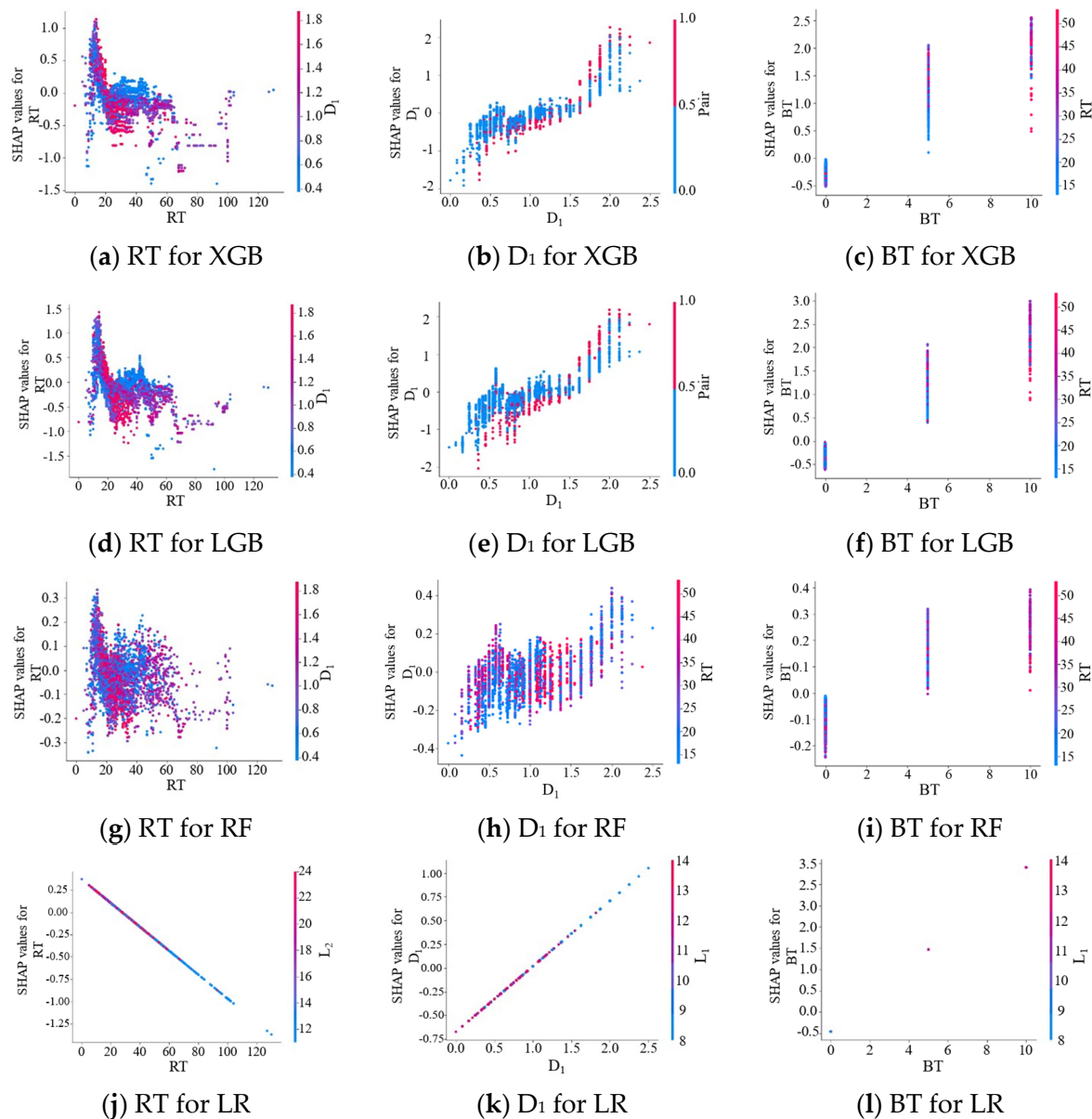
- (1) Higher BT leads to more pedestrians choosing Route 2, as it is directly related to the increased congestion on Route 1.
- (2) Increasing  $D_1$  prompts pedestrians to choose Route 2, indicating rising congestion on Route 1.
- (3) RT has a positive impact on the choice of Route 1, as a shorter RT suggests pedestrians wanting to finish the experiment quickly, and Route 1 is shorter.
- (4) Pedestrians walking with others (with Pair = 1) and facing congestion at the starting point (with Cro = 1) prefer to choose Route 2, since Route 2 is usually wider in many runs and they can choose to avoid the congestion.



**Figure 10.** The impact of each feature in Dataset ALL.

In addition, the results of LR in Figure 10d differ from those of tree-based models. Although many variables do not have obvious influences on the route choices, the contributions of  $D_1$  and RT are similar to those of tree-based models.

Next, in the dependence plots (Figure 11), we examine the intricate relationship between feature values and SHAP values. Our discussion focuses on the three most influential variables in Dataset ALL, namely RT, D<sub>1</sub>, and BT. Each point in the graph represents an observation, with the *x*-axis representing the feature value and the *y*-axis representing the SHAP value. Similar to Figure 10, the plots of LR (Figure 11j–l) clearly differ from the other models (The possible reason may be that in the results of LR, the SHAP values are approximately represented by the coefficients of the generalized linear models.), so our attention is directed towards the results of the remaining three models.



**Figure 11.** The dependence plots of Dataset ALL.

Given that BT has only three discrete values (0, 5, and 10) in our experiments, the dependence plots for BT appear straightforward, reinforcing the conclusions drawn from the summary plots. Regarding the results for RT in Figure 11a,d,g, it is evident that as RT increases from 0 to 20, the SHAP value consistently increases. However, beyond an RT value of 20, the SHAP value fluctuates around 0, indicating that larger RT values have little impact on the pedestrians' route choices. The possible explanation is that larger RT mainly



results from objective factors rather than subjective ones: it may be due to the congestion around the origin in some runs, which could not be changed by the attitude of a pedestrian.

For  $D_1$ , the results presented in Figure 11b,e,h indicate an inverse relationship: higher  $D_1$  values are associated with a reduced likelihood of pedestrians choosing Route 1. The threshold of  $D_1 = 1.8$  emerges as crucial, with all SHAP values turning positive when  $D_1$  surpasses 1.8. This suggests that pedestrians tolerate a maximum density of 1.8 ped/m on Route 1. This critical value aligns with our previous findings from single-file pedestrian flow experiments [42,43]. If  $D_1$  exceeds this threshold, pedestrians are more inclined to select Route 2 to avoid congestion. Therefore, our machine learning models establish a connection between the results of pedestrian experiments at two different levels (operational and tactical).

Lastly, as previously discussed, the predictive performance of DT is slightly inferior to that of XGB and LGB. We speculate that the more scattered distribution in Figure 11g,h might contribute to this, making the relationship between variables and pedestrian route choices less clear.

## 7. Discussion and Conclusions

### 7.1. Discussion

Based on the quantitative analysis of the experimental results, we would like to briefly discuss two important topics:

#### (1) Comparisons between machine learning models and DCMs

While the primary focus of this paper revolves around machine learning models, it is worth briefly mentioning the results obtained from DCMs. Theoretically, the estimation results of coefficients and the prediction outcomes of LR are akin to those of the Binomial Logit Model. As previously discussed, LR models require short training and predicting times, but their predicting accuracy ranks among the lowest. This finding is consistent with previous studies [27–30].

#### (2) Further investigation on the SHAP values

In DCMs, the coefficients of each variable are meaningful and can be transformed into other important statistics, such as odds ratios and willingness to pay, etc. However, the potential for further discussion on SHAP values remains unclear. While the coefficient of a variable (e.g., RT) in the Binomial Logit Model may differ from its SHAP value of some machine learning model (e.g., XGB or LGB), comparing these results can be challenging. Future research should aim to explore the relationship between these two different theories more comprehensively.

### 7.2. Conclusions

To explore pedestrians' route choice behavior from a perspective distinct from discrete choice models, we applied machine learning models to analyze pedestrian route choice behaviors. The dataset is extracted from videos of route choice experiments conducted in 2020 and 2021. In these experiments, participants were required to choose between two routes—one shorter but subject to distance control or bottleneck effects, and the other free from such constraints but longer. A total of 12 machine learning models were tested. Based on their prediction accuracy and computational efficiency, we calculated a new indicator named OP. We found that XGB and LGB emerged as the two models that strike a balance between computational speed and accuracy across all tested datasets, which also shows the advantages of ensemble learning algorithms in this field. On the other hand, NB, LR, and DT exhibited speed but lacked predictive accuracy. SVM and MLP demonstrated quite good accuracy, but their computational time was notably slower, particularly when dealing with large datasets.

Next, we employed the SHAP framework to interpret the model results. In terms of feature contribution, the most influential factor in pedestrian route choice was BT (average waiting time at the bottleneck), followed by RT (return time from destination) and  $D_1$

(restrain density on Route 1). The width and length of the routes, however, made little contribution to pedestrian choice behavior when both routes were congested. The critical values found in the SHAP plots are close to what we observed in the previous single-file pedestrian experiments, which also implies that we build a bridge between different levels of pedestrian movement.

Several aspects could be enhanced in future research. Firstly, individual features, particularly those representing pedestrian heterogeneity, are currently absent from the datasets and the corresponding analysis. To advance the quality of research in this area, it is imperative to gather additional experimental data that can encompass a broader range of relevant features. Secondly, this study focused on a simple scenario with two possible route choices. In future investigations, to enhance the practicality and applicability of the results, the experimental study of pedestrians' choices among multiple routes is necessary. The structure of the road network should also be considered in future studies where the number of routes increases. Finally, the results of this work are based on experimental data. Future studies should compare real pedestrian choice data and evaluate whether some differences exist between the experimental data and real-world data.

**Author Contributions:** Conceptualization, Cheng-Jie Jin and Dawei Li; Methodology, Cheng-Jie Jin, Chenyang Wu and Yuchen Song; Software, Yuanwei Luo; Validation, Yuanwei Luo; Formal analysis, Yuanwei Luo; Resources, Cheng-Jie Jin; Data curation, Cheng-Jie Jin; Writing—original draft, Cheng-Jie Jin and Yuanwei Luo; Writing—review & editing, Cheng-Jie Jin, Chenyang Wu and Yuchen Song; Visualization, Yuanwei Luo; Supervision, Dawei Li. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China (No. 52102389, 71801036, 71971056) and the Northwestern Polytechnical University Start-Up Funding (D5000230159).

**Data Availability Statement:** Data will be made available on request.

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

## References

- Guo, R.-Y.; Huang, H.-J. Modelling and solving dynamic entry pedestrian flow assignment problem. *Transp. B-Transp. Dyn.* **2023**, *11*, 1560–1590. [[CrossRef](#)]
- Haghani, M.; Sarvi, M. Crowd behaviour and motion: Empirical methods. *Transp. Res. Part B Methodol.* **2018**, *107*, 253–294. [[CrossRef](#)]
- Manski, C.F. The structure of random utility models. *Theory Decis.* **1977**, *8*, 229–254. [[CrossRef](#)]
- Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why Should I Trust You?” Explaining the Predictions of Any Classifier. In Proceedings of the NAACL-HLT 2016—2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Demonstrations Session, San Diego, CA, USA, 12–17 June 2016; pp. 97–101.
- Goldstein, A.; Kapelner, A.; Bleich, J.; Pitkin, E. Peeking inside the black box: Visualizing statistical learning with plots of individual conditional expectation. *J. Comput. Graph. Stat.* **2015**, *24*, 44–65. [[CrossRef](#)]
- Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Volume 30.
- Koppelman, F.S.; Wen, C.H. Alternative nested logit models: Structure, properties and estimation. *Transp. Res. Part B Methodol.* **1998**, *32*, 289–298. [[CrossRef](#)]
- King, C.; Bode, N.W.F. A virtual experiment on pedestrian destination choice: The role of schedules, the environment and behavioural categories. *R. Soc. Open Sci.* **2022**, *9*, 211982. [[CrossRef](#)] [[PubMed](#)]
- Haghani, M.; Sarvi, M. Laboratory experimentation and simulation of discrete direction choices: Investigating hypothetical bias, decision-rule effect and external validity based on aggregate prediction measures. *Transp. Res. Part A Policy Pract.* **2019**, *130*, 134–157. [[CrossRef](#)]
- Haghani, M.; Sarvi, M. Hypothetical bias and decision-rule effect in modelling discrete directional choices. *Transp. Res. Part A Policy Pract.* **2018**, *116*, 361–388. [[CrossRef](#)]
- Ray, P. Independence of Irrelevant Alternatives. *Econometrica* **1973**, *41*, 987–991. [[CrossRef](#)]
- McFadden, D.; Train, K. Mixed MNL models for discrete response. *J. Appl. Econom.* **2000**, *15*, 447–470. [[CrossRef](#)]
- Greene, W.H.; Hensher, D.A. A latent class model for discrete choice analysis: Contrasts with mixed logit. *Transp. Res. Part B Methodol.* **2003**, *37*, 681–698. [[CrossRef](#)]

14. Haghani, M.; Sarvi, M. Human exit choice in crowded built environments: Investigating underlying behavioural differences between normal egress and emergency evacuations. *Fire Saf. J.* **2016**, *85*, 1–9. [\[CrossRef\]](#)
15. Tong, Y.; Bode, N.W.F. The value pedestrians attribute to environmental information diminishes in route choice sequences. *Transp. Res. Part C Emerg. Technol.* **2021**, *124*, 102909. [\[CrossRef\]](#)
16. Tong, Y.; Bode, N.W.F. How building layout properties influence pedestrian route choice and route recall. *Transp. A Transp. Sci.* **2022**, *20*, 2143249. [\[CrossRef\]](#)
17. Tong, Y.; Bode, N.W.F. An investigation of how context affects the response of pedestrians to the movement of others. *Saf. Sci.* **2023**, *157*, 105919. [\[CrossRef\]](#)
18. Basu, N.; Haque, M.M.; King, M.; Kamruzzaman, M.; Oviedo-Trespalacios, O. A systematic review of the factors associated with pedestrian route choice. *Transp. Rev.* **2022**, *42*, 672–694. [\[CrossRef\]](#)
19. Li, H.; Zhang, J.; Xia, L.; Song, W.; Bode, N.W.F. Comparing the route-choice behavior of pedestrians around obstacles in a virtual experiment and a field study. *Transp. Res. Part C Emerg. Technol.* **2019**, *107*, 120–136. [\[CrossRef\]](#)
20. Zhang, D.; Huang, G.; Ji, C.; Liu, H.; Tang, Y. Pedestrian evacuation modeling and simulation in multi-exit scenarios. *Phys. A Stat. Mech. Its Appl.* **2021**, *582*, 126272. [\[CrossRef\]](#)
21. Liao, W.; Wagoum, A.U.K.; Bode, N.W.F. Route choice in pedestrians: Determinants for initial choices and revising decisions. *J. R. Soc. Interface* **2017**, *14*, 20160684. [\[CrossRef\]](#)
22. Yamamoto, T.; Kitamura, R.; Fujii, J. Driver's route choice behavior: Analysis by data mining algorithms. *Transp. Res. Rec.* **2002**, *1807*, 59–66. [\[CrossRef\]](#)
23. Barua, S. A Discrete Route Choice Model Using Support Vector Machine in Context of Dhaka City. *Daffodil Int. Univ. J. Sci. Technol.* **2019**, *14*, 3–8.
24. Yao, R.; Bekhor, S. Data-driven choice set generation and estimation of route choice models. *Transp. Res. Part C Emerg. Technol.* **2020**, *121*, 102832. [\[CrossRef\]](#)
25. Lai, X.; Fu, H.; Li, J.; Sha, Z. Understanding drivers' route choice behaviours in the urban network with machine learning models. *IET Intell. Transp. Syst.* **2019**, *13*, 427–434. [\[CrossRef\]](#)
26. Tan, S.K.; Hu, N.; Cai, W. A data-driven path planning model for crowd capacity analysis. *J. Comput. Sci.* **2019**, *34*, 66–79. [\[CrossRef\]](#)
27. Yuen, J.K.K.; Lee, E.W.M.; Lam, W.W.H. An intelligence-based route choice model for pedestrian flow in a transportation station. *Appl. Soft Comput.* **2014**, *24*, 31–39. [\[CrossRef\]](#)
28. Wang, K.; Shi, X.; Goh, A.P.X.; Qian, S. A machine learning based study on pedestrian movement dynamics under emergency evacuation. *Fire Saf. J.* **2019**, *106*, 163–176. [\[CrossRef\]](#)
29. Zhou, Z.X.; Nakanishi, W.; Asakura, Y. Route choice in the pedestrian evacuation: Microscopic formulation based on visual information. *Phys. A Stat. Mech. Its Appl.* **2021**, *562*, 125313. [\[CrossRef\]](#)
30. Zhou, Z.X.; Nakanishi, W.; Asakura, Y. Data-driven framework for the adaptive exit selection problem in pedestrian flow: Visual information based heuristics approach. *Phys. A Stat. Mech. Its Appl.* **2021**, *583*, 126289. [\[CrossRef\]](#)
31. Ullah, I.; Liu, K.; Yamamoto, T.; Zahid, M.; Jamal, A. Modeling of machine learning with SHAP approach for electric vehicle charging station choice behavior prediction. *Travel Behav. Soc.* **2023**, *31*, 78–92. [\[CrossRef\]](#)
32. Hasan, A.S.; Jalayer, M.; Das, S.; Kabir, M.A.B. Application of machine learning models and SHAP to examine crashes involving young drivers in New Jersey. *Int. J. Transp. Sci. Technol.* **2023**, *in press*. [\[CrossRef\]](#)
33. Kong, X.; Zhang, Y.; Eisele, W.L.; Xiao, X. Using an Interpretable Machine Learning Framework to Understand the Relationship of Mobility and Reliability Indices on Truck Drivers' Route Choices. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 13419–13428. [\[CrossRef\]](#)
34. Dong, H.; Zhou, M.; Wang, Q.; Yang, X.; Wang, F.Y. State-of-the-Art Pedestrian and Evacuation Dynamics. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 1849–1866. [\[CrossRef\]](#)
35. Li, Y.; Chen, M.; Dou, Z.; Zheng, X.; Cheng, Y.; Mebarki, A. A review of cellular automata models for crowd evacuation. *Phys. A Stat. Mech. Its Appl.* **2019**, *526*, 120752. [\[CrossRef\]](#)
36. Lue, G.; Miller, E.J. Estimating a Toronto pedestrian route choice model using smartphone GPS data. *Travel Behav. Soc.* **2019**, *14*, 34–42. [\[CrossRef\]](#)
37. Sevtsuk, A.; Basu, R.; Li, X.; Kalvo, R. A big data approach to understanding pedestrian route choice preferences: Evidence from San Francisco. *Travel Behav. Soc.* **2021**, *25*, 41–51. [\[CrossRef\]](#)
38. Yamamoto, T.; Takamura, S.; Morikawa, T. Structured random walk parameter for heterogeneity in trip distance on modeling pedestrian route choice behavior at downtown area. *Travel Behav. Soc.* **2018**, *11*, 93–100. [\[CrossRef\]](#)
39. Oyama, Y.; Hato, E. Link-based measurement model to estimate route choice parameters in urban pedestrian networks. *Transp. Res. Part C Emerg. Technol.* **2018**, *93*, 62–78. [\[CrossRef\]](#)
40. Feng, C.; Liao, S. Scalable Gaussian Kernel Support Vector Machines with Sublinear Training Time Complexity. *Inf. Sci.* **2017**, *418–419*, 480–494. [\[CrossRef\]](#)
41. Lundberg, S.M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J.M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.I. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2020**, *2*, 56–67. [\[CrossRef\]](#)

42. Jin, C.J.; Jiang, R.; Li, R.; Li, D. Single-file pedestrian flow experiments under high-density conditions. *Phys. A Stat. Mech. Its Appl.* **2019**, *531*, 121718. [[CrossRef](#)]
43. Jin, C.J.; Jiang, R.; Wong, S.C.; Xie, S.; Li, D.; Guo, N.; Wang, W. Observational characteristics of pedestrian flows under high-density conditions based on controlled experiments. *Transp. Res. Part C Emerg. Technol.* **2019**, *109*, 137–154. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.