# Urban Street Networks and Sustainable Transportation

Edited by

Moeinaddini Mehdi

Printed Edition of the Topical Collection Published in *Sustainability*

MDPI

# Urban Street Networks and Sustainable Transportation

# Urban Street Networks and Sustainable Transportation

Editor

**Moeinaddini Mehdi**

MDPI

*Editor*
Moeinaddini Mehdi
University of Liège
Belgium

This is a reprint of articles from the Topical Collection published online in the open access journal *Sustainability* (ISSN 2071-1050) (available at: https://www.mdpi.com/journal/sustainability/ special_issues/urban_street_networks_sustainable_transportation).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

# Contents

# About the Editor

**Moeinaddini Mehdi**  Before he started his academic activities, he was the head of the Design Department in his hometown municipality, where he led many urban mobility projects. He was involved in various research projects related to integrated land-use and transportation planning, with a focus on sustainable mobility. The results of these studies have been published in high-quality journals. He has written and lectured widely on urban design and sustainable mobility and has taught courses on methods of planning analysis, sustainable transportation, public transport, and quantitative analysis.

*Article*

# Hierarchical Longitudinal Control for Connected and Automated Vehicles in Mixed Traffic on a Signalized Arterial

Xiao Xiao [1,*], Yunlong Zhang [1], Xiubin Bruce Wang [1], Shu Yang [2] and Tianyi Chen [3]

[1]  Zachry Department of Civil and Environmental Engineering, Texas A&M University,
    College Station, TX 77840, USA; yzhang@civil.tamu.edu (Y.Z.); bwang@civil.tamu.edu (X.B.W.)
[2]  Department of Geography Information Systems, School of Transportation, Southeast University,
    Nanjing 210096, China; shuyang@seu.edu.cn
[3]  Department of Civil and Environment Engineering, University of Wisconsin-Madison,
    Madison, WI 53706, USA; tchen224@wisc.edu
*  Correspondence: xx1991@tamu.edu

**Abstract:** This paper proposes a two-layer hierarchical longitudinal control approach that optimizes travel time and trajectories along multiple intersections on an arterial under mixed traffic of connected automated vehicles (CAV) and human-driven vehicles (HV). The upper layer optimizes the travel time in an optimization loop, and the lower layer formulates a longitudinal controller to optimize the movement of CAVs in each block of an urban arterial by applying optimal control. Four scenarios are considered for optimal control based on the physical constraints of vehicles and the relationship between estimated arrival times and traffic signal timing. In each scenario, the estimated minimized travel time is systematically obtained from the upper layer. As the results indicate, the proposed method significantly improves the mobility of the signalized corridor with mixed traffic by minimizing stops and smoothing trajectories, and the travel time reduction is up to 29.33% compared to the baseline when no control is applied.

**Keywords:** consecutive signalized arterials; urban street; hierarchical longitudinal control; optimal control; connected and automated vehicles

## 1. Introduction

Sustainable transportation in an urban area has become an important topic attracting researchers' attention [1]. In the research of sustainable transportation, there have been studies from policy aspects such as promoting public transport, demand and supply controlling, integrated land use, and transport planning [2]. Other studies include developing design methods to solve technical problems operating transport means and facilities in a more efficient way [3]. The research on pedestrians and cycling is a major part of studying sustainable transportation [4]. As for motorized trips, on one hand, controlling demand is a concern [5]. On the other hand, the movements of vehicles on urban street networks and their effects on sustainable transportation is also an important component. More efficient movement of vehicles on urban street networks means a safer, faster, and more environmentally friendly urban network. Therefore, improving mobility is crucial in building up sustainable transportation.

However, drivers often experience stop-and-go shockwaves traveling through signalized intersections when most of the surrounding vehicles are driven by humans. Traffic oscillation and queue backpropagation may result in a capacity drop, leading to an increase in travel time and a decrease in mobility [6]. On an urban street, even when the signals are well-coordinated, the travel time increases for drivers traveling through consecutive signalized intersections [7]. Systematic methods for controlling vehicles on an urban arterial are essential.

The applications of CAVs in a traffic system have been studied in the last few years. CAVs can react to, communicate with, or make cooperative decisions considering the

environment such as surrounding vehicles and traffic facilities with the help of vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I) communication technologies. Adaptive Cruise Control (ACC) and Cooperative Adaptive Cruise Control (CACC) take advantage of the V2V communications so that vehicles can drive at a harmonized speed with short headways, addressing some issues that may occur for HVs in mobility, fuel efficiency, and safety issues [8]. When only considering the longitudinal direction, the design of a CACC system is usually based on a vehicle dynamics control strategy. To achieve ACC, vehicle dynamics are modeled by an optimal control framework to maintain speed while reducing emissions. When it comes to CACC, constant longitudinal spacing or headway should also be maintained [9]. Among all the objectives, the mobility, fuel efficiency, and stability of the traffic are the major concerns [8].

The longitudinal control strategies have been developed to improve mobility to mitigate the stop-and-go waves and other adverse traffic effects on freeways [10–12]. The stability problem of the longitudinal control of a CACC system in a CAV environment has also been well studied in previous studies [13–18]. Although longitudinal control strategies in the freeway environment have been well studied, the existence of traffic signals in an urban area makes the longitudinal control strategies of CAV significantly different from those in the freeway environment. The traffic signals cut traffic streams into interrupted flows and vehicle platoons which will be cut off and reformulated.

Many previous studies concerned the strategies for vehicles approaching an isolated intersection. For instance, Rakha and Kamalanathsharma developed eco-driving strategies for vehicles at an isolated intersection by integrating microscopic fuel consumption models in objective functions to minimize environmental adverse effects [19]. They also proposed a dynamic programming-based method to control the speed of a vehicle by splitting the process of approaching a signalized intersection into three states, showing that the method can save fuel and travel time significantly for an individual vehicle [20]. Chen et al. developed an eco-driving model that achieves the minimization of a linear combination of emissions and travel time [21]. Yang et al. developed an eco-CACC system to improve the fuel efficiency of CAVs at an isolated intersection considering the existing queues. Optimal control is used to design trajectories for leading CAVs of platoons to lead vehicles smoothly approaching an isolated intersection. The performances under different market penetration rates are demonstrated, showing a throughput benefit ranging from 0.88% to 10.80% [22]. A shooting heuristic (SH) is proposed for optimal control solutions for vehicle trajectories at intersections [23,24]. Individual Variable Speed Limits with location optimization are designed to smooth the trajectories of CAVs to improve mobility at an intersection [25].

In some studies, the platoon of CAVs is usually cooperatively considered. For example, a mixed-integer linear programming (MILP) based model is used to optimize vehicle trajectories as well as the traffic signal at isolated signalized intersections. The trajectories are generated by optimal control, car-following models, and lane choice models [26]. A Predictive Cruise Control method is used to control vehicles when traveling through multiple consecutive intersections to save fuel and $CO_2$ emissions [27]. A nonlinear-programming-based method to control a CAV platoon is designed to pass multiple intersections to maximize throughput and comfort [28].

In addition to only considering one intersection model, more pieces of the literature studied control strategies for consecutive traffic signals since the traffic signals are usually configured consecutively along the roadway in urban areas. Mandava et al. applied a dynamic speed-advise method to drive a CAV smoothly along consecutive intersections when no surrounding vehicles are concerned [29]. The method reduced fuel consumption and $CO_2$ emissions significantly and reduced travel time slightly (1.06%) for a single vehicle. Barth et al. developed an optimal control for a single vehicle to drive along consecutive signalized intersections, with a reduction in fuel consumption and $CO_2$ emissions. Other than the reduction in environmental adverse effects, queue minimization is considered in the development of the optimal trajectory of one single vehicle along consecutive intersections, which leads to an additional delay for the following vehicles [30]. A mixed-

integer programming sequential convex optimization is used to design an optimized speed plan of a vehicle when traveling along signalized intersections, saving travel time up to 6.00% [31]. Tang et al. incorporated a speed strategy into a car-following model for multiple vehicles to pass through multiple intersections [32].

Since the traffic stream will be in a state of having both CAVs and HVs for a long time, the control strategies for mixed traffic conditions become an important research direction. Specifically, HVs are concerned in some of the previous studies when developing the longitudinal control strategies of CAVs. The interaction of HVs and CAVs is modeled to optimize mobility [33] and emissions [34]. Wei et al. tested HVs as moving obstacles to validate their integer programming and dynamic programming models [35]. Recently, some studies also focus on the evaluation of the performance of mixed traffic. For example, the performance of lane choice for the mixed traffic with CAVs is analyzed [36]. Speed estimation is conducted in a mixed traffic condition [37]. When HVs are considered, the sequence of the mixed traffic needs to be assumed; for example, Zhao et al. used scenarios in the experiment to show the possible combination of HVs and CAVs [34].

The operation strategy of connected and automated vehicles at intersections can either be modeled in a centralized way, as the studies using dynamic programming or cooperative control mentioned before, or a decentralized way. For example, Du et al. developed a multi-layer coordination strategy for CAVs at intersections without the help of signals [38]. Yao and Li proposed a decentralized control method for CAVs at an intersection to optimize their own travel time, fuel consumption, and safety risks and showed that it is more computationally efficient than a centralized control [39]. Mahbub et al. developed a coordination method for CAVs at a corridor considering multiple traffic scenarios using a two-level optimization [40].

Although the problem of the longitudinal control of connected automated vehicles has been widely studied, the control for CAVs in mixed traffic is hard when considering consecutive signalized arterials, which can lead to a problem of variable control horizon. In addition, the synchronization of the calculation of CAV travel time and trajectory is a difficulty in the proposed problem. To fill in the gap, this paper provides a new approach of hierarchical longitudinal control that can address mixed traffic, tackle the variable horizon of CAVs, and give insight into the scenarios of CAV control on a signalized corridor. A centralized method is unable to model HVs, which are uncontrolled. To tackle this issue, this paper introduces an efficient decentralized method [41]. While the studies about single lanes focus on longitudinal control, CAV-related control on multilane scenarios is also a research direction concerning lane changing and lane assignment. For example, a cooperative sorting strategy is developed for the platooning of CAVs along multiple lanes [42]. Formation controls are used for the lane assignment for CAVs [43,44]. Therefore, focusing on the longitudinal control in this paper, a dedicated lane is considered to maximize the benefits of controlling CAVs and in showing how the methods influence traffic dynamics. In addition, due to the low MPR for a long period of time, HVs should also be allowed in the "dedicated" lane. In this setting, lane changing, and overtaking are not considered. Therefore, this paper models a single lane of mixed traffic. The contributions of the paper are highlighted below:

- Propose a systematic method to analyze CAVs at signals based on split scenarios according to preceding vehicle and signal conditions.
- Develop a hierarchical longitudinal control for CAVs considering variable horizon optimal control in urban streets.

## 2. Problem Statement

The problem aims to control the microscopic longitudinal behaviors of CAVs by minimizing the travel time given a fixed signal timing on an urban signalized arterial corridor. As shown in Figure 1, the mixed traffic travels through consecutive intersections on the urban street from upstream intersection 1 to intersection $i$ at downstream. The traffic is a mixture of HVs and CAVs. Communication devices are installed on CAVs to ensure real-time information exchange via V2V and V2I.
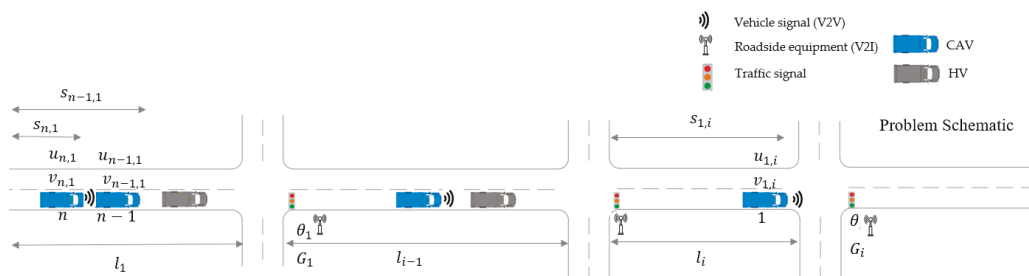


**Figure 1.** Schematic representation of the problem longitudinal control of connected and automated vehicles along a signalized arterial.

The assumptions of the paper are listed as follows. The V2V communication is assumed to be active once a vehicle entering the block. Information related to the timing plan such as offset $\theta_i$, the duration of green $G_i$, green elapse time $G_{n,i}$, and geometrical variable block length $l_i$ can be received by CAVs with no delay. The overtaking behavior of a vehicle is not in the scope of concerns. The car following behaviors of HVs are assumed as known, and HVs slow down and stop in front of a signal when they cannot pass within the current green interval.

In Figure 1, the vehicles move forward in their longitudinal direction. The travel time of a vehicle within a block is defined as the duration between the time instant when it passes the intersection $i - 1$ and the time instant it passes intersection $i$.

The vehicle dynamics within a block for a CAV are expressed by a state-space representation, indexed by the number of vehicles and intersections. On an urban street, the vehicles are not allowed to move backward. A CAV can obtain information of vehicle status such as position, acceleration, and speed from the preceding vehicle, no matter whether the preceding vehicle is a CAV or an HV.

The research question is how to reduce the travel time for all vehicles when they are traveling from the first intersection to the final intersection and provide a suitable trajectory for each vehicle. The difficulties of this problem are that traffic signals exist along consecutive intersections, cutting off the traffic. Multiple states exist for a vehicle, in which varying control horizons can appear; HVs are uncontrolled, and HVs and CAVs are mixed with arbitrary sequences, so an integrated centralized optimization is not applicable. In addition, the control horizon for each vehicle is different.

## 3. Methodology

The longitudinal control for CAVs follows a hierarchical structure: at the upper level, the travel time is calculated; at the lower level, the optimal control is applied to generate the trajectories.

### 3.1. Lower-Level Control: Mathematical Formulation of Optimal Control

When an individual vehicle is traveling within one block between two intersections, its state including position and speed is known. The problem is decomposed into different scenarios and is then scaled towards multiple vehicles along consecutive intersections. The constraints from the longitudinal position and feasible arrival moments of a vehicle with the presence of signals are mathematically described. Each scenario is explained with their transportation meaning and provided a solution of minimum travel time and trajectory.

As a solution for individual vehicles, the trajectory generates in an optimal control fashion. The state $x_{n,i}$ of a vehicle $n$ in intersection $i$ is defined as a combination of its longitudinal position $s_{n,i}$ within this block $i$ and longitude speed $v_{n,i}$:

$$x_{n,i} = (s_{n,i}, v_{n,i})^T \tag{1}$$

The system writes with a linear time-invariant system (LTI):

$$x_{n,i}(t) = Ax_{n,i}(t) + Bu_{n,i}(t), \tag{2}$$

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} B = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \tag{3}$$

where the control variable $u_{n,i}$ is the acceleration of the vehicle. The cost function to ensure optimal performances is defined as follows considering the comfort and terminal performances:

$$J_{n,i} = \min \int_{t=0}^{T_{n,i}} L(x_{n,i}(t), u_{n,i}(t))dt + \Phi(T_{n,i}, x_{n,i}(T_{n,i})), \tag{4}$$

where the ending time or the control horizon $T_{n,i}$ is a variable which is determined systematically. It is then discussed in Section 3.2, based on different scenarios. The running cost is set as an instantaneous cost showing the penalties concerning comfort. It is expressed as the quadratic term of acceleration:

$$L = \frac{1}{2}u_{n,i}{}^2. \tag{5}$$

The terminal cost gives penalties so that the final states can approach desired values (terminal speed and terminal distance):

$$\Phi = w_1\left(x_{n,i}{}^{(1)}(T_{n,i}) - l^*{}_{n,i}\right)^2 + w_2\left(x_{n,i}{}^{(2)}(T_{n,i}) - v^*{}_{n,i}\right)^2. \tag{6}$$

Again, $T_{k,i}$ will be determined systematically. Weighing factors $w_1$ and $w_2$ show the penalty for the state deviation from the terminal speed and the terminal distance at the end of the horizon. The desired speed is set to the terminal speed at each intersection for each vehicle: $v^*{}_{n,i} = v_0$. The block length between two intersections is used as terminal distance $l^*{}_{n,i} = l_i$. The problem then writes:

$$J_{n,i} = \sum_{K=1}^{T}\left(u_{n,it+k-1}{}^2\right) + w_1\left(x_{n,i}{}^{(2)}{}_T{}^2 - 2x_{n,i}{}^{(2)}{}_T v_{n,i}{}^* + v_{n,i}{}^{*2}\right) + w_2\left(x_{n,i}{}^{(1)}{}_T{}^2 - 2x_{n,i}{}^{(1)}{}_T l^* + l_{n,i}{}^{*2}\right), \tag{7}$$

s.t.

$$(x_{n,i}, u_{n,i}) \in \Omega \cap U, \tag{8}$$

where $\Omega$ represents the constraints from vehicle dynamics, including the limitation from maximal speed, maximal acceleration, distance, etc. $U$ represents the physical constraints from the preceding vehicle during the period when it follows preceding vehicle $f_{n,i}$.

$$\Omega = \left\{ x_{n,it+1} = A_d x_{n,it} + B_d u_{n,it}, u_{n,it} \in (u_{n,i,lb}, u_{n,i,ub}), x_{n,i}^{(1)} \in (0, l_i), x_{n,i}^{(2)} \in (v_{n,i,lb}, v_{n,i,ub}) \right\}, \tag{9}$$

$$U = \{ s_{n,i} \leq s_{n-1,i} + d_s + d_v, t \in (0, f_{n,i}) \}, \tag{10}$$

where $d_s$ is a safe distance that can ensure safety, and $d_v$ is the vehicle length; $f_{n,i}$ is the duration of following, determined differently in different scenarios in upper-level control.

### 3.2. Upper-Level Control: Determination of Travel Time

Having set the variable horizon optimal control, the horizon $T_{n,i}$ is to be determined systematically. Some prerequisites are provided.

#### 3.2.1. Following Behavior along Consecutive Signalized Intersections

With the availability of V2I techniques, CAV receives signal information including current state and future time phases such as $G_i$ and $\theta_i$. The arrival moments should be in a feasible region (the collection of green) and the physical constraints should always hold for safety concerns. To avoid stopping, for CAVs, the set of feasible arrival moments $M_{n,i}$ should be in the collection of green time $G$:

$$M_{n,i} \in [\theta_i + C * (k - 1), \theta_i + G_i + C * (k - 1)], \tag{11}$$

where $C$ is the cycle length. If no preceding vehicle exists, $k$ is the counter of the cycles after the current cycle in which the vehicle can pass. $k^*$ is the optimal $k$ that minimizes the travel time. If a vehicle is not able to pass within this cycle, it is natural that it passes at the next cycle, only if the preceding vehicle has passed. Generally, k could be 0 or 1 showing whether a vehicle is able to pass at this cycle or the next:

$$k^* = agrmin T_{n,i}. \tag{12}$$

Accumulative position $p_{n,i}(t)$ of a vehicle $n$ at time $t$ can be denoted as the addition of two parts: the accumulative position along previous blocks from 1 to $i - 1$, and the current position $p_{n,i}(t)$ in this block $i$ for vehicle $n$ is:

$$p_{n,i}(t) = \sum_1^{i-1} s_{n,i} + s_{n,i} \left( t - \sum_1^{i-1} T_{n,i} \right). \tag{13}$$

At time $t$, the vehicle has two state conditions which is either passed block $i$ or not. When the subject vehicle has a preceding vehicle in the same block, an inequality describes the situation:

$$\sum_{i-1} l_i < p_{n,i}(t) < p_{n-1,i}(t) < \sum_i l_i. \tag{14}$$

Similarly, when the preceding vehicle is not in the same block, an inequality writes:

$$\sum_{i-1} l_i < p_{n,i}(t) < \sum_i l_i < p_{n-1,i}(t). \tag{15}$$

If the subject vehicle has a preceding vehicle in the same block, its duration is constrained by the preceding vehicle. The moments that enter or leaves a block can be calculated from the values of accumulated travel time:

$$M_{n,i} = \sum_1^{i-1} T_{n,i}, \ M_{n,i+1} = \sum_1^i T_{n,i}. \tag{16}$$

When a vehicle has a preceding vehicle, $f_{n,i}$ stands for the time duration that the subject CAV following its preceding vehicle within this block. This duration is the subtraction of the moment the preceding vehicle leaves this block and the moment when the subject vehicle enters the block:

$$f_{n,i} = M_{n-1,i+1} - M_{n,i}. \tag{17}$$

To scale the problem to consecutive intersections, $G_{n,i}$ shows the duration of green before the vehicle passes the intersection at the moment $M_{n,i}$. This variable links the time of trajectories between two intersections.

### 3.2.2. Scenario Development

The continuation of position and speed are addressed by introducing variables such as the cycle length $C$, green time $G_i$, green elapse time $G_{n,i}$, and offset $\theta_i$. Each vehicle is planned only once in a block, the moment a vehicle passes the previous intersection becomes the starting moment the vehicle enters the next intersection; the information is indicated with the help of green elapse time. The final status of a vehicle becomes the initial status in the next.

For CAVs, the arrival moments at the stop line of each intersection are estimated ahead. For HVs, the arrival moments are estimated using travel time estimation methods. According to the categories of the estimated arrival moments and whether there is a preceding vehicle, four scenarios can be defined, and they are noted as scenario 0, scenario 1, scenario 2, and scenario 3, respectively:

$$0 < s_{n,i}(t) < l_i < s_{n-1,i}(t); M_{n,i} \in [\theta_i + C*(k-1), \theta_i + G_i + C*(k-1)], k \le 1, \quad (18)$$

$$0 \le s_{n,i}(t) < s_{n-1,i}(t) \le l_i; M_{n,i} \in [\theta_i + C*(k-1), \theta_i + G_i + C*(k-1)], k > 1, \quad (19)$$

$$0 \le s_{n,i}(t) < s_{n-1,i}(t) \le l_i; M_{n,i} \in [\theta_i + C*(k-1), \theta_i + G_i + C*(k-1)], k > 1, \quad (20)$$

$$0 < s_{n,i}(t) < l_i < s_{n-1,i}(t); M_{n,i} \in [\theta_i + C*(k-1), \theta_i + G_i + C*(k-1)], k \le 1. \quad (21)$$

When the subject CAV is the leading vehicle in the same block, the way to minimize travel time is to accelerate and maintain its desired speed to travel through the block to pass the intersection (setting the speed limit as the desired speed $v_0$). The minimal travel time is obtained when the subject CAV accelerates to the desired speed and maintains the speed until it passes the signal ahead:

$$T^*_{n,i} = \{ T_{n,i} | (u = u_0 | v \le v_0), (u = 0 | v = v_0) \}. \quad (22)$$

The value of $G_{n,i+1}$ in the next intersection $i + 1$ is calculated using travel time $T_{n,i}$ and the value of $G_{n,i}$, $\theta_i$ from the last intersection:

$$G_{n,i+1} = G_{n,i} + T_{n,i} - \theta_i. \quad (23)$$

For the subject CAV with no preceding vehicle in the same block, when it is not expected to pass the intersection within this cycle, it is planned to pass during the green in the next cycle, ($M_{n,i} \in [\theta_i + C*(k-1), \theta_i + G_i + C*(k-1)], k > 1$), via a smooth path without stopping. The corresponding $T^*_{n,i}$ for both scenario 1 is calculated by:

$$T^*_{n,i} = \theta_i + C*k^* - G_{n,i} + G_{n,i+1}. \quad (24)$$

$G_{n,i+1}$ varies the arrival moments, which is set as small as possible so that the startup time can be saved compared to human driving behavior.

For scenario 2, the calculation of $T^*_{n,i}$ and $G_{n,i+1}$ is the same as that of scenario 1. The difference is the subject vehicle has constraints from its preceding vehicle for the preceding vehicle is in the same block. $U$ is active as the physical constraints of the optimal control.

Scenario 3 shows when the subject CAV follows a preceding vehicle in this intersection, and it passes within the same green window as the preceding vehicle: $M_{n,i} \in [\theta_i + C*(k-1), \theta_i + G_i + C*(k-1)], k \le 1$. The corresponding $T^*_{n,i}$ is then calculated from:

$$T^*_{n,i} = max(T_{n-1,i} - f_{n,i} + t_{0,i}, \frac{l_i}{v_0}). \quad (25)$$

$$G_{n,i+1} = G_{n,i} + T_{n,i} - \theta_i - C*k^*. \quad (26)$$

Note that the minimal travel time cannot be smaller than the value when the vehicle is traveling with the desired speed (in that case, the travel time from scenario 3 is no smaller than that from scenario 0). $U$ is active as the physical constraints from the preceding vehicle.

Although an HV cannot respond to a CAV, a CAV can detect the position of its preceding HV. An estimation of the HV's travel time is conducted. The desired headway $t_{0,i}$ when a CAV following an HV is set to be larger than that an HV follows an HV to ensure safety. The travel time when a CAV follows an HV is calculated as:

$$T^*_{n,i} = max(T_{n-1,i} - f_{n,i} + t_{0,i(HV)}, \frac{l_i}{v_0}). \tag{27}$$

An HV is expected to slow down and stop if it cannot pass an intersection within the green duration. They will be modeled remaining at a standstill at the stop bars during the red phases. The subject CAV does not need to follow closely to an HV. Instead, it passes with a smooth trajectory without stopping. The calculations of $T^*_{n,i}$ and $G_{n,i+1}$ are the same as the case when it follows a CAV. In the schematic diagrams of Figure 2, the blue line shows the estimated trajectory of an HV, and a black line shows the preceding vehicle trajectory. A magenta line represents the trajectory of a CAV.
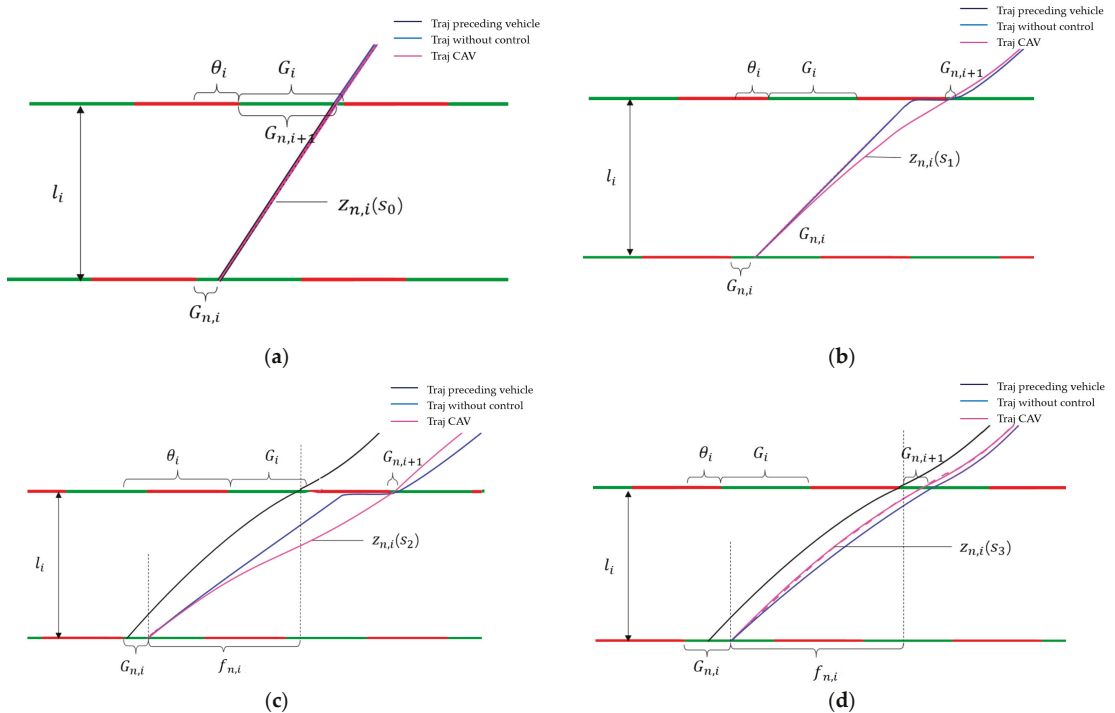


**Figure 2.** Schematic diagram of (**a**) scenario 0; (**b**) scenario 1; (**c**) scenario 2; (**d**) scenario 3.

*3.3. Synthesized Algorithm*

In lower-level control, the optimal control has been set up for each vehicle to calculate their optimal trajectories. In upper-level control, the scenarios are developed. In each scenario, the way to find the minimum travel time has been introduced. The problem in this paper is to minimize the total travel time for all vehicles therefore the hierarchical control is addressed systematically in a synthesized way.

According to the analysis of scenarios, scenario 0 is designed as the vehicle that can drive with its speed limit. Scenario 3 follows preceding vehicle successfully without being hampered by a red light. Both scenarios are with no time loss. Scenario 1 and 2 experienced time losses at red. It is obvious that, at the same intersection $i$, the travel time for each scenario has the following relations:

$$T^*_{n,i}(scnario\ 0) < T^*_{n,i}(scnario\ 3) \leq T^*_{n,i}(scnario\ 1) \leq T^*_{n,i}(scnario\ 2). \tag{28}$$

Apparently, the travel time reaches minimal when an ideal condition can occur in which all scenarios are scenario 0. Nevertheless, a vehicle may not be able to drive with scenario 0 along all the blocks. In this case, replacing one of the scenarios into another scenario with the least cost for vehicle $n$ achieves the minimal costs that are feasible. Therefore, a greedy heuristic is to try to plan scenario 0 or scenario 3 first, and then to plan scenario 1 or 2.

Define $z_{n,i} \triangleq \left[ u_{n,i}(0)^T, \ldots, u_{n,i}(t-1)^T \right]^T$ as the decision variable of vehicle $n$ from the time instant 0 to $t$ at each intersection $i$. Once a selection of scenarios is made, the minimal travel time $T_{n,i}^*$ is calculated. The decision variables of the preceding vehicle $\bar{z}_{n-1,i}$ and the constraints inputs into the next calculation. By assuming there are $N$ vehicles and $I$ intersections, the calculation process is listed as follows:

Start: start with intersection $i = 1, n = 1$

- Step 1: If $0 \leq s_{n,i}(t) < s_{n-1,i}(t) \leq l_i$ , go to Step 2a; otherwise, go to Step 3a.
- Step 2a: $M_{n,i} \in [\theta_i + C*(k-1), \theta_i + G_i + C*(k-1)], k = 1$, obtain the numerical solution as scenario 0; otherwise, go to Step 2b.
- Step 2b: $M_{n,i} \in [\theta_i + C*(k-1), \theta_i + G_i + C*(k-1)], k > 1$, obtain the numerical solution for optimal control problem as scenario 1.
- Step 3a: $M_{n,i} \in [\theta_i + C*(k-1), \theta_i + G_i + C*(k-1)], k = 1$, obtain the numerical solution for optimal control problem as scenario 3; otherwise, go to Step 3b.
- Step 3b: $M_{n,i} \in [\theta_i + C*(k-1), \theta_i + G_i + C*(k-1)], k > 1$, obtain the numerical solution for optimal control problem as scenario 2 (following a CAV) or scenario 3 (following an HV).
- Step 4: Find the solution $z_{n,i}^*$ for vehicle $n$ at intersection $i$ and broadcast all the outputs from current plan to all other CAVs. The known decision variables are $\bar{z}_{n-1,i}$ then.

End: End by $i = I, n = N$.

As described in the algorithm, the controller determines each CAV individually and broadcasts its information and solutions. Information is broadcasted to the follower if it is a CAV. This proceeds until all the vehicles have solutions for trajectory profiles. The process is demonstrated in Figure 3.
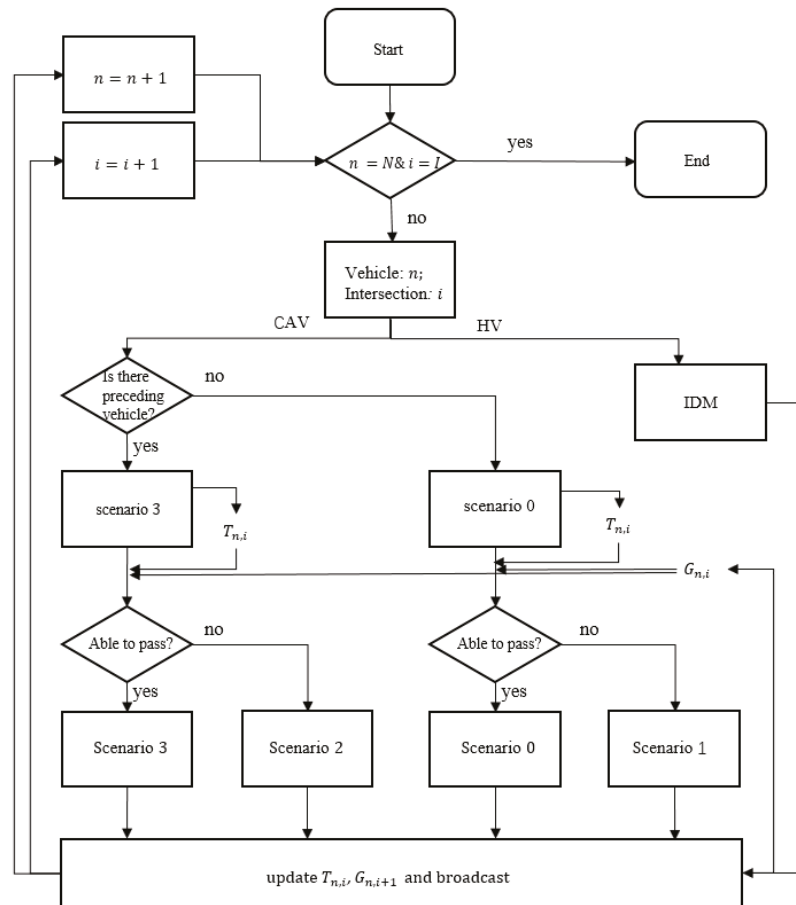
**Figure 3.** The flow chart of synthesized algorithm.

## 4. Numerical Simulations

The proposed method was implemented in MATLAB, and the numerical simulations are demonstrated below. To test conditions under light traffic does not have much value since no traffic backpropagation will happen, so only the cases with moderate demands were considered. Two cases were presented to validate the method. Case 1 compared the method with the situation when all vehicles were HVs. HVs were assumed to slow down and stop when approaching a signalized intersection if they expected to fail to pass and remain standstill at the stop bars during the red phases. HVs were assumed to follow preceding vehicles using the intelligent driver model (IDM) model [45]. Case 2 compared the proposed method with a benchmark when all CAVs drive smoothly to avoid stopping at intersections without the consideration of minimal travel time.

Both cases comprised two examples. In one example, the initial average headway input was set as 5 s. In the other example, the initial input headway was 3 s. The desired headway for a CAV and the IDM model was set as 3 s; the desired headway for a CAV following an HV was set at 4 s for safety concerns. Multiple runs with random seeds were applied in each case to calculate the average travel time savings under each penetration rate. The parameters used in the experiment are listed in Table 1.

**Table 1.** Values of parameters in the experiments.

| Parameters | Notation | Value |
|---|---|---|
| Block length (m) | $l_i$ | 1000 |
| Number of blocks | N | 3 |
| Number of vehicles | I | 60 |
| Cycle length (s) | C | 100 |
| Green duration (s) | $G_i$ | 60 |
| Offset (s) | $\theta_i$ | 30 |
| Weight for terminal distance | $w_1$ | 10 |
| Weight for terminal speed | $w_2$ | 10 |
| Number of vehicle inputs | N | 60 |
| Safe distance (meter) | $d_s$ | 5 |
| Vehicle length (meter) | $d_v$ | 5 |
| Desire headway (s) | $t_{0,i}$ | 3 |
| Desire headway a CAV following an HV (s) | $t_{0,i(HV)}$ | 4 |
| Maximum speed (m/s) | $v_{n,i,ub}$ | 20 |
| Minimal speed (m/s) | $v_{n,i,lb}$ | 5 |
| Maximum acceleration (m/s$^2$) | $u_{n,i,ub}$ | 2 |
| Minimal deacceleration (m/s$^2$) | $u_{n,i,lb}$ | $-2$ |

*4.1. Performance under Different Penetration of CAVs*

Case 1 compares the results when no CAVs and when some CAVs using the proposed are applied. The simulated results are presented in Figures 4 and 5.
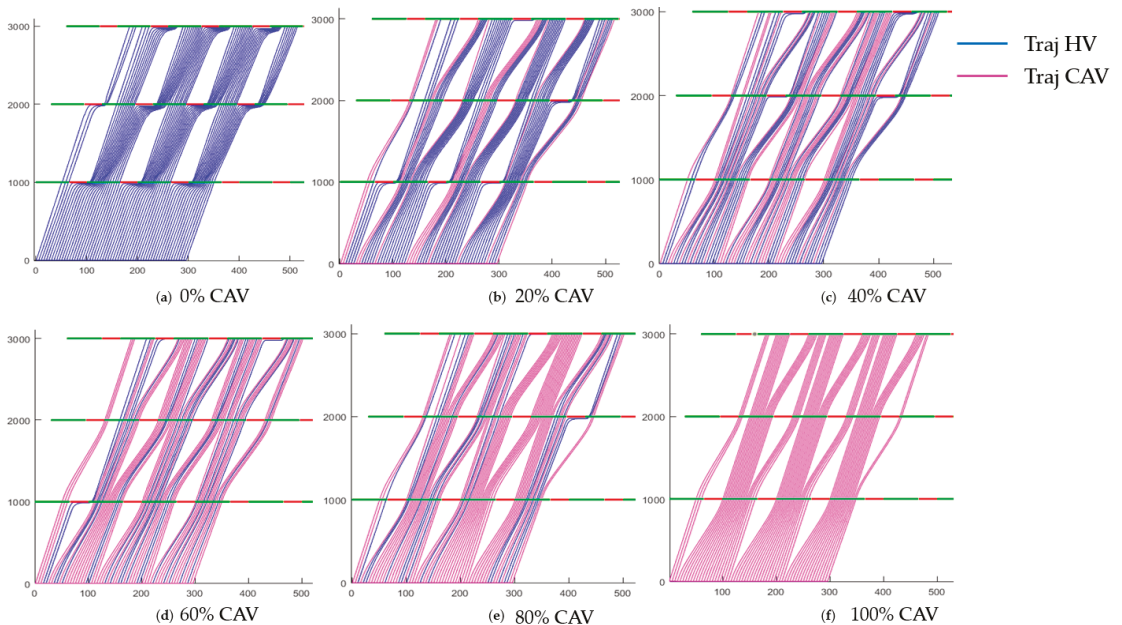


**Figure 4.** A comparison of trajectories between HVs (blue lines) and CAVs (magenta) under varying penetration rates of CAV when the initial headway for CAVs was 5 s (x-axis —time (s), y-axis—distance (m)).
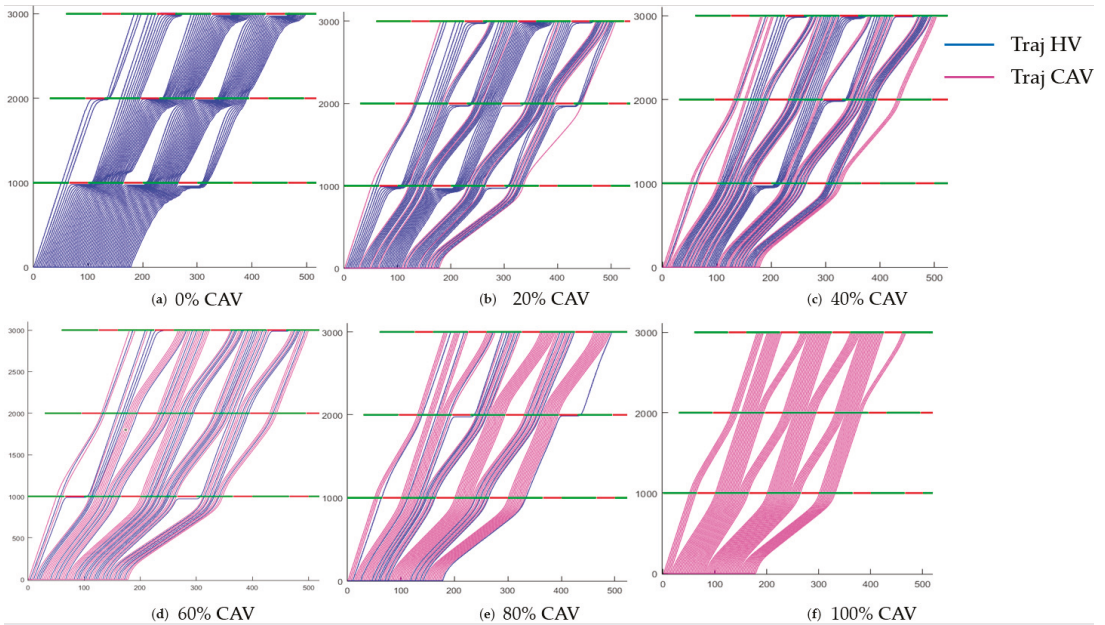
**Figure 5.** A comparison of trajectories between HVs (blue lines) and CAVs (magenta) under varying penetration rates of CAV, when the initial headway for CAVs was 3 s (x-axis—time (s), y-axis—distance (m)).

When the initial headway for CAVs was 5 s, the CAVs trajectories could lead the whole platoon to decompose and reconstruct reasonably. This led to a reduction in travel time in the first step. The results also showed that the proposed method can reduce the number of stops; as a result, the queues and backpropagation shockwaves were mitigated to reduce the startup time, which saved travel time in the second step. The method compressed the headways for CAVs when the initial headway was larger than the desired headway, which made the traffic stream compact, leading to a reduction of travel time in the third step. Compared to the situations when all vehicles are HVs (0%), the effects of mitigation of adverse phenomena became more significant with the increase of penetration rates. When the penetration rate was 100%, the stops were mostly eliminated, and no queue and backpropagation shockwave showed.

When traffic demand was higher, according to Figure 5, although the initial headways were so small that they cannot be compressed, travel time was saved from the first two steps: The whole platoon still decomposed and reconstructed in a certain manner to ensure vehicles could pass with the shortest time, and the queues and backpropagation shockwaves were also mitigated. The overall results after multiple runs are presented in Figure 6.

When the penetration rate of CAVs was as low as 20%, the methods could lead to a negative effect (−1.57% and −4.12 %). The reason was that a large desire headway (4 s) for a CAV following an HV was set to ensure safety, which was larger than the case when a CAV followed a CAV (3 s) or when an HV followed an HV (3 s). However, with the increasing penetration rates of CAVs, the travel time savings become effective. The travel time savings were significant when the penetration rate was larger than 60% for both cases. When a full penetration rate was assumed, the proposed method can provide travel time savings of 29.33 % and 26.85 % in two examples.
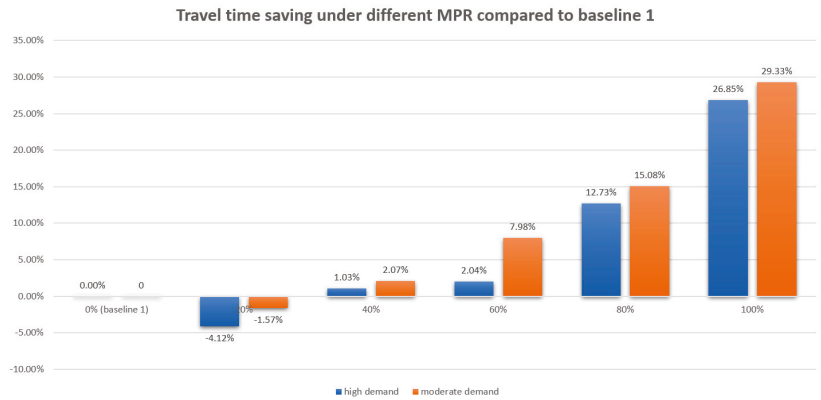
**Figure 6.** Travel time saving using the proposed method under different penetration rates of CAVs.

### 4.2. Compare with a Benchmark

A benchmark was configured with the following settings: (1) the trajectories of HVs were generated in the same way as in case 1; (2) the trajectories of CAVs were generated based on a benchmark. For case 2, only the optimal control was used to smooth the trajectories of the leading CAV at an intersection, and the others followed their leaders. Similarly, in these cases, different initial headways were demonstrated.

As seen in Figures 7 and 8, although smooth trajectories could reduce travel time by reducing time-consuming stop and startup driving behaviors at an intersection, they led to an increase in travel time if multiple intersections were involved and the local minimal travel time was not considered. This case showed the importance of the proposed method to calculate the minimal travel time locally under all possible scenarios.
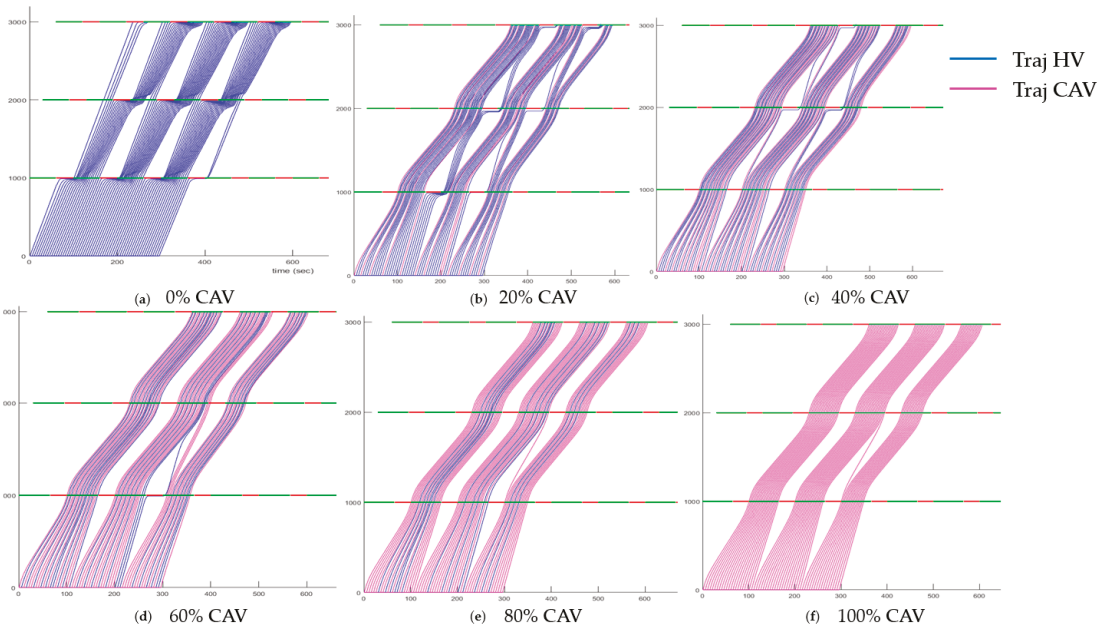


**Figure 7.** The trajectories between HVs (blue lines) and CAVs (magenta) under varying penetration rates of CAV when the initial headway was 5 s (x-axis—time (s), y-axis—distance (m)) controlling CAVs using benchmark.
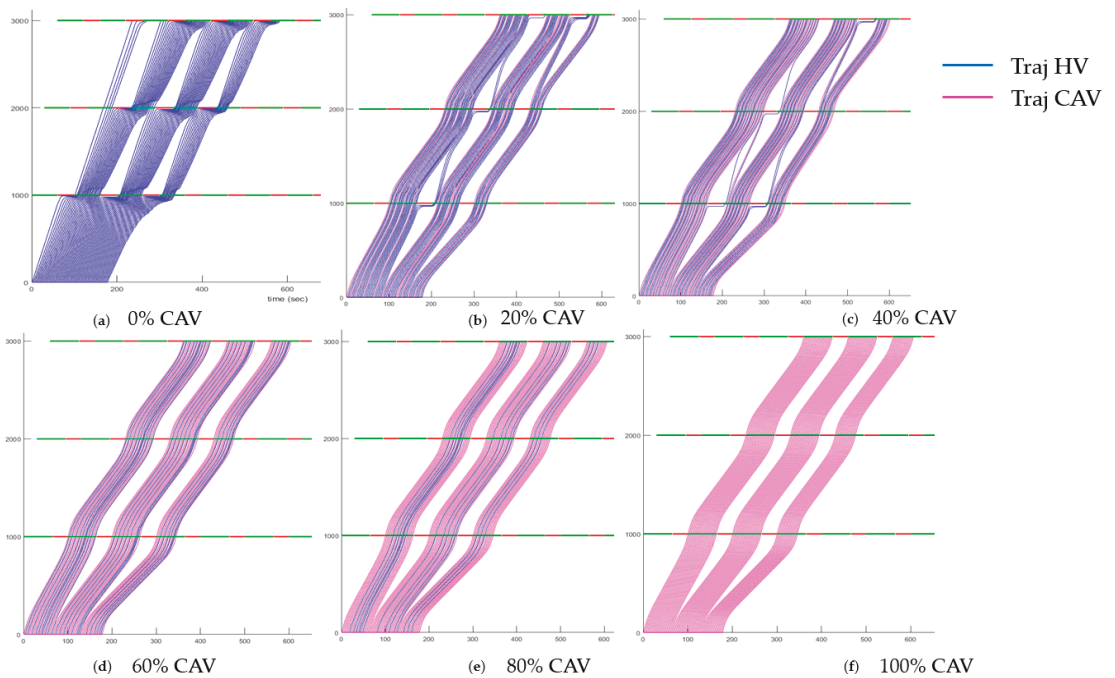
**Figure 8.** The trajectories between HVs (blue lines) and CAVs (magenta) under varying penetration rates of CAV when the initial headway was 3 s (x-axis—time (s), y-axis—distance (m)) controlling CAVs using benchmark.

The outputs from case 1 and case 2 showed a significant difference in Figure 9.
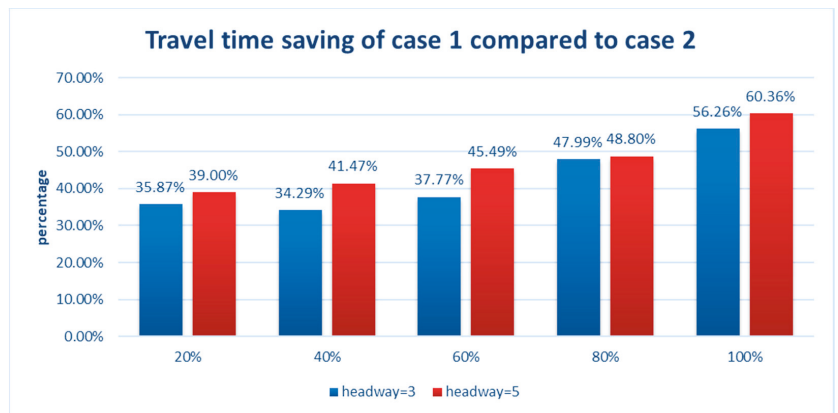


**Figure 9.** Travel time savings in case 1 compared to benchmark under different penetration rates of CAVs.

Comparing case 1 (using the proposed method to control CAVs) with case 2 (using a benchmark), 35.87% (shorter headway) and 39.00% (larger headway) travel time savings

were shown, even when the penetration rate was as low as 20%. The percentage increased to 56.26% and 60.36% when a full penetration rate was assumed.

**5. Conclusions**

Traffic oscillation and queue backpropagation caused by traffic signals can interrupt traffic streams periodically and increase the travel time for drivers. To ensure sustainable transport on a signalized urban street by improving mobility, a connected automated vehicle hierarchical longitudinal control for mixed traffic on consecutive signalized arterials was proposed to control multiple vehicles along multiple intersections, considering their varying control horizons. The main aim is to focus on vehicle mobility on signalized arterials to improve sustainable urban transportation.

In the lower-level layer, mathematical formulations were developed for the relations between vehicles and signals during the time vehicles were traveling along consecutive signalized intersections. In the upper-level layer, the conditions of vehicles are decomposed into four scenarios. In each scenario, a minimal travel time is calculated. A synthesized algorithm is used to connect lower-level and upper-level layers.

Two cases were developed to validate the proposed control strategy. Case 1 concerned a non-CAV setting and Case 2 assumed all CAVs with smooth trajectory without considering the travel time. The proposed method significantly reduced the number of stops. When it came to travel time savings, when the initial headway was larger, the travel time saving ranged from $-1.57\%$ to 29.33 %. When the initial headway was smaller, the travel time saving was also significant (ranging from $-4.12$ % to 26.85 %). Compared to case 2 using a benchmark, the proposed method can save travel time from 35.87% to 56.26% and 39.00% to 60.36%.

The limitation of this paper was that the status of the CAVs and HVs were assumed as deterministic, and only a single lane was considered in the problem. In the future, how these scenarios are stably switched in the real world will be considered. In addition, the method is to be generalized to multilane scenarios by considering lane changing and overtaking behaviors.

**Author Contributions:** Conceptualization, X.X., Y.Z., and X.B.W.; methodology, X.X. and Y.Z.; software, X.X.; validation, X.X.; formal analysis, X.X.; investigation, X.X., Y.Z., and T.C.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X., S.Y.; writing—review and editing, X.X., Y.Z., S.Y., T.C.; visualization, X.X.; supervision, Y.Z.; project administration, Y.Z. and X.B.W.; funding acquisition, Y.Z. and X.B.W. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

**References**

1. Zhao, X.; Ke, Y.; Zuo, J.; Xiong, W.; Wu, P. Evaluation of Sustainable Transport Research in 2000–2019. *J. Clean. Prod.* **2020**, *256*, 120404. [CrossRef]
2. Diao, M. Towards Sustainable Urban Transport in Singapore: Policy Instruments and Mobility Trends. *Transp. Policy* **2019**, *81*, 320–330. [CrossRef]
3. Gong, D.; Tang, M.; Liu, S.; Xue, G.; Wang, L. Achieving Sustainable Transport through Resource Scheduling: A Case Study for Electric Vehicle Charging Stations. *Adv. Prod. Eng. Manag.* **2019**, *14*, 65–79. [CrossRef]
4. Elvik, R. The Non-Linearity of Risk and the Promotion of Environmentally Sustainable Transport. *Accid. Anal. Prev.* **2009**, *41*, 849–855. [CrossRef] [PubMed]

5.    Moeinaddini, M.; Asadi-Shekari, Z.; Shah, M.Z. An Urban Mobility Index for Evaluating and Reducing Private Motorized Trips. *Measurement* **2015**, *63*, 30–40. [CrossRef]
6.    Treiber, M.; Kesting, A.; Helbing, D. Delays, Inaccuracies and Anticipation in Microscopic Traffic Models. *Phys. Stat. Mech. Its Appl.* **2006**, *360*, 71–88. [CrossRef]
7.    Denney, R.W., Jr. Traffic Platoon Dispersion Modeling. *J. Transp. Eng.* **1989**, *115*, 193–207. [CrossRef]
8.    Wang, Z.; Wu, G.; Barth, M.J. A Review on Cooperative Adaptive Cruise Control (CACC) Systems: Architectures, Controls, and Applications. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 2884–2891.
9.    Dey, K.C.; Yan, L.; Wang, X.; Wang, Y.; Shen, H.; Chowdhury, M.; Yu, L.; Qiu, C.; Soundararaj, V. A Review of Communication, Driver Characteristics, and Controls Aspects of Cooperative Adaptive Cruise Control (CACC). *IEEE Trans. Intell. Transp. Syst.* **2015**, *17*, 491–509. [CrossRef]
10.   Lee, J.; Park, B. Development and Evaluation of a Cooperative Vehicle Intersection Control Algorithm under the Connected Vehicles Environment. *IEEE Trans. Intell. Transp. Syst.* **2012**, *13*, 81–90. [CrossRef]
11.   González, D.; Pérez, J.; Milanés, V.; Nashashibi, F. A Review of Motion Planning Techniques for Automated Vehicles. *IEEE Trans. Intell. Transp. Syst.* **2015**, *17*, 1135–1145. [CrossRef]
12.   Rios-Torres, J.; Malikopoulos, A.A. A Survey on the Coordination of Connected and Automated Vehicles at Intersections and Merging at Highway On-Ramps. *IEEE Trans. Intell. Transp. Syst.* **2016**, *18*, 1066–1077. [CrossRef]
13.   Wang, M.; Daamen, W.; Hoogendoorn, S.P.; van Arem, B. Rolling Horizon Control Framework for Driver Assistance Systems. Part I: Mathematical Formulation and Non-Cooperative Systems. *Transp. Res. Part C Emerg. Technol.* **2014**, *40*, 271–289. [CrossRef]
14.   Wang, M.; Daamen, W.; Hoogendoorn, S.P.; van Arem, B. Rolling Horizon Control Framework for Driver Assistance Systems. Part II: Cooperative Sensing and Cooperative Control. *Transp. Res. Part C Emerg. Technol.* **2014**, *40*, 290–311. [CrossRef]
15.   Wang, Z.; Wu, G.; Barth, M.J. Developing a Distributed Consensus-Based Cooperative Adaptive Cruise Control System for Heterogeneous Vehicles with Predecessor Following Topology. *J. Adv. Transp.* **2017**, *2017*, 1–16. [CrossRef]
16.   Gong, S.; Zhou, A.; Peeta, S. Cooperative Adaptive Cruise Control for a Platoon of Connected and Autonomous Vehicles Considering Dynamic Information Flow Topology. *Transp. Res. Rec.* **2019**, *2673*, 185–198. [CrossRef]
17.   Wang, C.; Gong, S.; Zhou, A.; Li, T.; Peeta, S. Cooperative Adaptive Cruise Control for Connected Autonomous Vehicles by Factoring Communication-Related Constraints. *Transp. Res. Part C Emerg. Technol.* **2020**, *113*, 124–145. [CrossRef]
18.   Zhou, Y.; Wang, M.; Ahn, S. Distributed Model Predictive Control Approach for Cooperative Car-Following with Guaranteed Local and String Stability. *Transp. Res. Part B Methodol.* **2019**, *128*, 69–86. [CrossRef]
19.   Rakha, H.; Kamalanathsharma, R.K. Eco-Driving at Signalized Intersections Using V2I Communication. In Proceedings of the 2011 14th international IEEE conference on intelligent transportation systems (ITSC), Washington, DC, USA, 5–7 October 2011; pp. 341–346.
20.   Kamalanathsharma, R.K.; Rakha, H.A. Multi-Stage Dynamic Programming Algorithm for Eco-Speed Control at Traffic Signalized Intersections. In Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013); IEEE, Kurhaus, The Hague, The Netherlands, 6–9 October 2013; pp. 2094–2099.
21.   Chen, Z.; Zhang, Y.; Lv, J.; Zou, Y. Model for Optimization of Ecodriving at Signalized Intersections. *Transp. Res. Rec.* **2014**, *2427*, 54–62. [CrossRef]
22.   Jiang, H.; Hu, J.; An, S.; Wang, M.; Park, B.B. Eco Approaching at an Isolated Signalized Intersection under Partially Connected and Automated Vehicles Environment. *Transp. Res. Part C Emerg. Technol.* **2017**, *79*, 290–307. [CrossRef]
23.   Ma, J.; Li, X.; Zhou, F.; Hu, J.; Park, B.B. Parsimonious Shooting Heuristic for Trajectory Design of Connected Automated Traffic Part II: Computational Issues and Optimization. *Transp. Res. Part B Methodol.* **2017**, *95*, 421–441. [CrossRef]
24.   Zhou, F.; Li, X.; Ma, J. Parsimonious Shooting Heuristic for Trajectory Design of Connected Automated Traffic Part I: Theoretical Analysis with Generalized Time Geography. *Transp. Res. Part B Methodol.* **2017**, *95*, 394–420. [CrossRef]
25.   Yao, H.; Cui, J.; Li, X.; Wang, Y.; An, S. A Trajectory Smoothing Method at Signalized Intersection Based on Individualized Variable Speed Limits with Location Optimization. *Transp. Res. Part Transp. Env.* **2018**, *62*, 456–473. [CrossRef]
26.   Yu, C.; Feng, Y.; Liu, H.X.; Ma, W.; Yang, X. Integrated Optimization of Traffic Signals and Vehicle Trajectories at Isolated Urban Intersections. *Transp. Res. Part B Methodol.* **2018**, *112*, 89–112. [CrossRef]
27.   Asadi, B.; Vahidi, A. Predictive Cruise Control: Utilizing Upcoming Traffic Signal Information for Improving Fuel Economy and Reducing Trip Time. *IEEE Trans. Control Syst. Technol.* **2010**, *19*, 707–714. [CrossRef]
28.   Liu, M.; Wang, M.; Hoogendoorn, S. Optimal Platoon Trajectory Planning Approach at Arterials. *Transp. Res. Rec.* **2019**, *2673*, 214–226. [CrossRef]
29.   Mandava, S.; Boriboonsomsin, K.; Barth, M. Arterial Velocity Planning Based on Traffic Signal Information under Light Traffic Conditions. In Proceedings of the 2009 12th International IEEE Conference on Intelligent Transportation Systems, St. Louis, MO, USA, 4–7 October 2009; pp. 1–6.
30.   He, X.; Liu, H.X.; Liu, X. Optimal Vehicle Speed Trajectory on a Signalized Arterial with Consideration of Queue. *Transp. Res. Part C Emerg. Technol.* **2015**, *61*, 106–120. [CrossRef]
31.   Huang, X.; Peng, H. Speed Trajectory Planning at Signalized Intersections Using Sequential Convex Optimization. In Proceedings of the 2017 American Control Conference (ACC), Seattle, WA, USA, 24–26 May 2017; pp. 2992–2997.

32. Tang, T.-Q.; Yi, Z.-Y.; Zhang, J.; Wang, T.; Leng, J.-Q. A Speed Guidance Strategy for Multiple Signalized Intersections Based on Car-Following Model. *Phys. Stat. Mech. Its Appl.* **2018**, *496*, 399–409. [CrossRef]
33. Carlino, D.; Boyles, S.D.; Stone, P. Auction-Based Autonomous Intersection Management. In Proceedings of the 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), The Hague, The Netherlands, 6–9 October 2013; pp. 529–534.
34. Zhao, W.; Ngoduy, D.; Shepherd, S.; Liu, R.; Papageorgiou, M. A Platoon Based Cooperative Eco-Driving Model for Mixed Automated and Human-Driven Vehicles at a Signalised Intersection. *Transp. Res. Part C Emerg. Technol.* **2018**, *95*, 802–821. [CrossRef]
35. Wei, Y.; Avcı, C.; Liu, J.; Belezamo, B.; Aydın, N.; Li, P.T.; Zhou, X. Dynamic Programming-Based Multi-Vehicle Longitudinal Trajectory Optimization with Simplified Car Following Models. *Transp. Res. Part B Methodol.* **2017**, *106*, 102–129. [CrossRef]
36. Guo, X.; Peng, Y.; Ashraf, S.; Burris, M.W. Performance Analyses of Information-Based Managed Lane Choice Decisions in a Connected Vehicle Environment. *Transp. Res. Rec.* **2020**, *2674*, 120–133. [CrossRef]
37. He, S.; Guo, X.; Ding, F.; Qi, Y.; Chen, T. Freeway Traffic Speed Estimation of Mixed Traffic Using Data from Connected and Autonomous Vehicles with a Low Penetration Rate. *J. Adv. Transp.* **2020**, *2020*, 1–13. [CrossRef]
38. Du, Z.; HomChaudhuri, B.; Pisu, P. Hierarchical Distributed Coordination Strategy of Connected and Automated Vehicles at Multiple Intersections. *J. Intell. Transp. Syst.* **2018**, *22*, 144–158. [CrossRef]
39. Yao, H.; Li, X. Decentralized Control of Connected Automated Vehicle Trajectories in Mixed Traffic at an Isolated Signalized Intersection. *Transp. Res. Part C Emerg. Technol.* **2020**, *121*, 102846. [CrossRef]
40. Mahbub, A.I.; Malikopoulos, A.A.; Zhao, L. Decentralized Optimal Coordination of Connected and Automated Vehicles for Multiple Traffic Scenarios. *Automatica* **2020**, *117*, 108958. [CrossRef]
41. Jiao, J.; Trentelman, H.L.; Camlibel, M.K. Distributed Linear Quadratic Optimal Control: Compute Locally and Act Globally. *IEEE Control Syst. Lett.* **2019**, *4*, 67–72. [CrossRef]
42. Wu, J.; Ahn, S.; Zhou, Y.; Liu, P.; Qu, X. The Cooperative Sorting Strategy for Connected and Automated Vehicle Platoons. *Transp. Res. Part C Emerg. Technol.* **2021**, *123*, 102986. [CrossRef]
43. Cai, M.; Xu, Q.; Li, K.; Wang, J. Multi-Lane Formation Assignment and Control for Connected Vehicles. In Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019; pp. 1968–1973.
44. Cai, M.; Xu, Q.; Chen, C.; Wang, J.; Li, K.; Wang, J.; Zhu, Q. Formation Control for Connected and Automated Vehicles on Multi-Lane Roads: Relative Motion Planning and Conflict Resolution. Available online: https://arxiv.org/abs/2103.10287 (accessed on 7 August 2021).
45. Treiber, M.; Hennecke, A.; Helbing, D. Congested Traffic States in Empirical Observations and Microscopic Simulations. *Phys. Rev. E* **2000**, *62*, 1805. [CrossRef] [PubMed]

*Article*

# Built Environment Determinants of Pedestrian Activities and Their Consideration in Urban Street Design

**Regine Gerike [1,*], Caroline Koszowski [1], Bettina Schröter [1], Ralph Buehler [2], Paul Schepers [3], Johannes Weber [1], Rico Wittwer [1] and Peter Jones [4]**

1   "Friedrich List" Faculty of Transport and Traffic Sciences, Institute of Transport Planning and Road Traffic, Technische Universität Dresden, 01069 Dresden, Germany; caroline.koszowski@tu-dresden.de (C.K.); bettina.schroeter@tu-dresden.de (B.S.); johannes.weber1@tu-dresden.de (J.W.); rico.wittwer@tu-dresden.de (R.W.)
2   Virginia Tech Research Center (VTRC), Urban Affairs and Planning, Arlington, VA 22203, USA; ralphbu@vt.edu
3   Department of Human Geography and Spatial Planning, Utrecht University, 3584 Utrecht, The Netherlands; j.p.schepers@uu.nl
4   Centre for Transport Studies (CTS), University College London, London WC1E 6BT, UK; peter.jones@ucl.ac.uk
*   Correspondence: regine.gerike@tu-dresden.de

**Abstract:** Pedestrian facilities have been regarded in urban street design as "leftover spaces" for years, but, currently, there is a growing interest in walking and improving the quality of street environments. Designing pedestrian facilities presents the challenge of simultaneously accommodating (1) pedestrians who want to move safely and comfortably from point A to B (movement function); as well as (2) users who wish to rest, communicate, shop, eat, and enjoy life in a pleasant environment (place function). The aims of this study are to provide an overview of how the task of designing pedestrian facilities is addressed in international guidance material for urban street design, to compare this with scientific evidence on determinants of pedestrian activities, and to finally develop recommendations for advancing provisions for pedestrians. The results show that urban street design guidance is well advanced in measuring space requirements for known volumes of moving pedestrians, but less in planning pleasant street environments that encourage pedestrian movement and place activities. A stronger linkage to scientific evidence could improve guidance materials and better support urban street designers in their ambition to provide safe, comfortable and attractive street spaces that invite people to walk and to stay.

**Keywords:** walking; pedestrians; urban street design; pedestrian facilities; link and place functions; sidewalk; walkability

## 1. Introduction

For many years, spaces for pedestrians were treated as "leftover spaces" in urban street design. In regard to technical geometrical street design, motorised vehicle size was the main determinant for minimum lane widths. The provision of dedicated lanes for public transport depended on space availability and its level of prioritisation in local transport policy; defined target values for traffic quality for motorised vehicles, e.g., in terms of level of service for the forecasted traffic volumes, determined the number of lanes in street sections and at junctions. Additionally, the recent rise in the popularity of cycling has resulted in the increase in both the quality and quantity of cycling facilities. Yet the accommodation for pedestrian needs or place functions has fallen by the wayside, particularly in areas with limited street space availability. Furthermore, seen from an engineering perspective, with a width of about 0.75 m to 1.00 m, a "standard" pedestrian does not typically occupy much space, thus causing pedestrians to be perceived and treated as a more flexible user group compared to motorised vehicles and bicycles.

Two additional problems hamper the efforts of transport planners in providing for pedestrians: (1) Apart from the quality of the street environment, spatial structures and land use are also strong incentives for walking; thus, despite poor conditions, pedestrians will still walk if spatial structures and land use are supportive. (2) Planners rarely have reliable information about existing or expected pedestrian volumes. Even in the current era of digitalisation, pedestrians are still counted by hand in most cases, which is burdensome, time-consuming and rarely done.

Various combinations of the above-described issues have been the focus of many discussions concerning urban street design tasks, which has led to street layouts with overly narrow sidewalks. Those narrow sidewalks rarely accommodate pedestrians' movement functions and often do not encourage place activities such as resting, waiting, communicating, shopping, eating, and enjoying life in a pleasant environment.

At the same time, research interest in walking and in walkability has sharply increased, and new insights have surfaced about why people walk and about the various benefits of walking [1,2]. For example, the Health Economic Assessment Tool (HEAT-Tool, https://www.heatwalkingcycling.org/ (accessed on 14 August 2021), provided by the WHO/Europe, allows cities to compute in advance the monetised health effects of anticipated behavioural change as well as increased walking and cycling levels. It is consensus that walking is a key ingredient of liveable cities, and contributes to a healthier population as well as to more environmentally friendly travel behaviours.

Cities and stakeholders are increasingly aware of these positive effects. Thus, there is increasing interest around the world in walking and in improving the quality of street environments to be more walkable. Cities such as New York are redesigning parts of their street networks and urban spaces with a primary focus on an increased quality of space for pedestrian and dense urban areas. The City of Malmö places pedestrians at the highest level of their street-user hierarchy [3]. In London, the healthy street approach takes highest priority in the Mayor's Transport Strategy [4], and also at the national level, more and more pedestrian strategies are being put in place (see e.g., [5]). The current COVID-19 pandemic and related physical distancing requirements bring new challenges and opportunities for efforts to provide for pedestrians [6].

Seeing the scientific evidence on the positive effects of pedestrian activities and the increasing interest in encouraging walking and lively streets, it becomes clear that spaces for pedestrians must not be treated as "leftover spaces". They should be the focus of attention.

This study focuses on the design of streets and pedestrian facilities as one important determinant of pedestrian activities, as well as one main field in policy-making for promoting walking. This study compiles standards for pedestrian facilities, including both movement and place functions, from international guidelines on urban street design from five European cities and six nationwide guides from European countries and the USA (NACTO). It compares these with empirical evidence from the scientific literature on infrastructure-based determinants of pedestrian activity in urban streets.

Two goals are pursued with this approach: Our comparison of standards can be used separately by researchers who analyse covariates of pedestrian activities. Our overview of scientific evidence provides a concise summary of infrastructure-based determinants of pedestrian activities. Our comparison of scientific evidence and standards highlights how the transfer from research to practice works, and simultaneously allows us to derive recommendations for advancing the guidelines based on insights gained in research. These insights should help address the above-described tensions and challenges, and give urban street designers optimal guidance for reliably providing for pedestrian movement and place activities, while at the same time leaving flexibility for finding tailor-made solutions that fit to the local context and that overall contribute to the final goal of advancing provision for pedestrians.

The remainder of this paper is organised as follows: Section 2 presents scientific evidence on determinants of pedestrian activities related to street characteristics and the

built environment. It is followed by the summary of guidance material on pedestrian facilities in Section 3. Section 4 compares scientific evidence and guidance material in order to show how the transfer between research and practice works. Recommendations on providing for pedestrians in future guidelines on urban street design are developed in Section 5. The final Section 6 summarises main findings and gives an outlook to further research.

**2. Determinants of Walking and Place Activities**

Research on determinants of walking and place activities is as interdisciplinary as the research topic itself [1,7]. Public health researchers focus on minutes of walking as one part of overall physical activity, and particularly include person-related variables such as socio-psychological variables, body mass index, or physical activity at work and for leisure purposes into their analyses [8]. Transport planners try to understand, above all, the influence of network and street characteristics on pedestrian volumes [9–11]. Urban planning literature also considers the characteristics of street networks, but takes a much broader view, including variables describing land use and other neighbourhood and city characteristics [12–14]. Three main groups of determinants of pedestrian movement and place activities related to street characteristics and the built environment could be identified in the analysis of the scientific literature:

1. Urban design and land use are of utmost importance for achieving high levels of pedestrian activities in the streets.
2. Streetscape design also matters.
3. The successful provision for pedestrian movement and place activities requires far more than pedestrian-focused urban and transport design.

In what follows, the main findings from the literature are summarised for each of these three groups of determinants.

*2.1. Urban Design and Land Use*

The "5 Ds" (Density, Diversity, Design, Distance to public transport, Destination accessibility) are consistently significant and influential for pedestrian activities in the researched literature [8,11,15–19]. Ewing et al. [9,20] demonstrate that Density is particularly important, measured in their example as floor area ratio and population density within a quarter mile of the investigated commercial streets. Diversity is often captured by entropy measures describing the number and variety of different land use types in a given area [15,19,21]. Ewing et al. found it to be statistically significant in one study [20], but not another study [9]. Shorter Distances, particularly to rail-based public transport, consistently and significantly increase pedestrian volumes [9,22]. Design-variables describe the characteristics, and more specifically the connectivity, of the street network, measured, e.g., as intersection density or as proportion of four-way intersections [15,23]. Mixed findings exist for these Design variables, which are significant in some studies, and not in others [9]. Destination accessibility describes the level to which relevant activities can be reached [15,24]. Destinations are operationalised, e.g., by the number of nearby stores and amenities weighted by their distance; these are hardly significant in Ewing et al. [9,20] and show an overlap with Diversity.

Some authors work with Cs instead of the Ds described above in order to investigate the influence of the built environment on pedestrian volumes: Connectivity, Convenience, Comfort, Conviviality, Conspicuousness, Coexistence, Commitment [25–28]. These Cs are a mixture of variables on the neighbourhood and street level; they show a substantial overlap with the Ds, and findings on their impacts on pedestrian activities are consistent with the findings summarised above.

*2.2. Streetscape*

The Ds also apply to the streetscape itself. This holds particularly for Design, but also for the other Ds. Ewing et al. [9] show the significant influence of floor area ratios along

the streets themselves (computed as the total building floor area for parcels abutting the street, divided by the total area of tax lots) and of the proportion of retail frontage along the block face on pedestrian volumes. In their pioneering work on the Design variable on the street level, Ewing and Handy [11] measured more than 100 features of selected streetscapes. Based on expert rankings as the dependent variable, the following five urban design qualities were identified as the most important:

- Imageability—quality of a place that makes it distinct, recognisable and memorable, measured, e.g., by the proportion of historic buildings, buildings with non-rectangular silhouettes and with identifiers, such as major landscape features;
- Enclosure—degree to which the streets are visually defined by buildings, walls, trees and other vertical elements, measured, e.g., by the proportion of street walls, of sky visible across and ahead, and by long sight lines;
- Human scale—size, texture, and articulation of physical elements that match the size and proportions of humans and correspond to the speed at which humans walk, measured, e.g., by long sight lines, street furniture, proportion of first floor with windows, building height, small planters;
- Transparency—degree to which people can see or perceive what lies beyond the edge of a street, measured, e.g., by the proportion of buildings with windows at street level and of active uses of adjacent buildings;
- Complexity—visual richness of a place, measured, e.g., by the number of buildings, dominant building colours, accent colours, pieces of public art, people in the street, and by the presence of outdoor dining.

These criteria have been validated against counted pedestrian volumes in subsequent studies [11,20]. Controlling for the D variables as introduced above, on the street level, only transparency was found to significantly influence pedestrian volumes. This is consistent with findings from other studies [29,30]. The only exemption is imageability, which was identified in one study as a variable that significantly increases pedestrian volumes [29]. Ewing et al. [9] refined the above concepts and analysed the influence of around 20 variables measuring the physical features of streetscapes on pedestrian volumes separately, resulting in three significant variables: proportion of windows, street furniture, and active uses. Overall, the three streetscape design features added significantly to the explanatory power of the statistical models on pedestrian volumes, compared to models with only the D variables on the neighbourhood and street levels. Street furniture was defined as a variety of signs, benches, parking meters, trash cans, newspaper boxes, bollards, and street lights, and includes anything at the human scale that increases the complexity of the street. Public seating was found to be of special importance. The proportion of active uses was defined as shops, restaurants, public parks, and other uses that generate significant pedestrian traffic. Inactive uses include blank walls, driveways, parking lots, vacant lots, abandoned buildings, and offices with no apparent activity.

Kang [31] and Kim et al. [22] focus on the street layout itself. They find significant positive impacts of sidewalk widths, crosswalks and trees, and negative impacts of slopes, on pedestrian volumes. The number of traffic lanes is positively associated with pedestrian volume, but highly correlated with the distance to public transport. Lai and Kontokosta [19] computed a composite variable called "streetscape" as the combination of sidewalk coverage, pavement quality, and street amenity. This variable significantly increases pedestrian volumes on weekends but not on workdays.

While a large number of studies analyse pedestrian volume, only few research groups and studies focus on place activities [13,14,32–34]. These are operationalised either by the number of people in a place [33,34], or by the liveliness index, as the product of people undertaking place activities times the duration of these activities (15 s to <1 min, 1 min to <5 min, 5 min to <10 min, 10 min to <15 min, ≥15 min) [13,14,32,35]. Mehta and Bosson [14] distinguish various activity types and the following physical human postures for their studies: standing, sitting, lying, sleeping. The determinants of place activities show substantial similarities with those of pedestrian volumes, and add further

valuable insights to how to achieve lively streets, including both pedestrian movement and place activities. The existence of community places, such as stores, that are places to meet neighbours, friends, strangers, etc., are most important for the liveliness index, followed by the provision of seating, both commercial and public. Personalisation is also statistically significant and describes how the interface of businesses with the street (building façade, entrances, shop windows) is embellished with personal touches, such as displays, decorations, signs, banners, planters, flowerboxes, and other wares. The variables permeability and variety of businesses are only significant in one study each [14,32]. Sidewalk widths are only significant in a study by Metha [32], and seem to be more of a mediating variable that is less relevant on its own but allows for facilities, such as seating, on the sidewalk that foster place activities. No significant influences on the liveliness index have been identified for shade provided, the existence of street furniture besides seating, the articulation of façades, and the degree of independence of the adjacent stores.

In addition to these empirical analyses of the influence of streetscape, urban design, and land use on pedestrian activities, various schemes for assessing walkability exist, e.g., the Pedestrian Environment Review System (PERS) [36], the Microscale Audit of Pedestrian Streetscapes (MAPS) [37] or the Healthy Street Checks applied by Transport for London [38]. These studies mainly rely on expert knowledge. They formulate recommendations for how to check the friendliness and suitability of street network elements for walking and place activities, and for how to improve walkability. The street characteristics included in these walkability assessments correspond well with the significant variables identified in the literature as described above, but go beyond this empirical evidence based on expert knowledge. Various street characteristics are investigated in walkability assessments, and these can be grouped along (1) destinations and land use, (2) street scape, and (3) aesthetics and social aspects [37].

Gehl [12] distinguished twelve quality criteria for high-quality street spaces for pedestrians. The criteria are grouped into the following categories:

- Protection—objective and subjective (perceived) safety against traffic and traffic crashes, as well as security against crime, are prerequisites and motivating factors for walking and for place activities. In addition, "protection against unpleasant sensory experiences" is to be considered;
- Comfort—after taking safety and security issues into account, the provision of comfortable public spaces must be ensured in order to invite people into different movement and place activities. For pedestrians, sidewalks should offer sufficient space void of obstacles (e.g., a dedicated footway clear zone) and good surface quality. Providing space for different place activities invites place users to spend time in public spaces;
- Delight—to ensure the well-being of pedestrians and place users, the human scale (in regard to adequate street width and building height) must be taken into account. The delight of design with respect to details and materials, as well as green structures, promote walking and the enjoyment of public spaces by place users.

Gehl [12] does not provide any quantitative validation for these twelve criteria, such as a comparison with empirically measured volumes of pedestrian movement or place activities. However, he lists various examples for the successful application of these criteria in projects for redesigning streets and public spaces all over the world [39].

### 2.3. Governance and Stakeholder Engagement

Studies in urban design, and particularly the projects published by the groups around Mehta et al. [13,14,32,35,40] and Gehl et al. [12,39], clearly show that successful provision for pedestrians needs more than tailor-made and pedestrian-focused designs. Designing and managing liveable streets is an interdisciplinary task that can only be achieved if far more stakeholders collaborate than only urban and transport designers.

Cities have a prominent role in initiating and coordinating such collaboration and in developing policies that support the various community-based stakeholders to engage in improving and actively using the streets in their neighbourhood. Incentive schemes

might be set up that create or strengthen small independent businesses, especially those that are perceived as community places. Longer and more flexible opening hours for local businesses might be considered and encouraged, contributing to active street usage over the whole day, week, and year. Cities might transfer some level of control to businesses and users so that these local stakeholders are enabled and feel invited to claim street space, e.g., by providing movable street furniture or by allowing businesses to use parts of the street for their activities and facilities. Incentives might also be given for the organisation of events such as street closures, festivals, open classroom projects, or other activities that strengthen the community. Temporary changes in the use of parts of the streets, e.g., by allowing parklets in summer, by closing lanes or taking out parking lots, e.g., on selected weekends, might also encourage pedestrian activities and give a different perspective on the potential of streets and possible perspectives.

Local building codes might support permeable and articulated façades at the street level. Nooks, alcoves, small setbacks, steps and ledges serve multiple purposes, e.g., people might seek shelter, get out of pedestrian flow, or stop and rearrange their belongings.

Streets are ecosystems; their users and usages constantly evolve. Streetscapes that are perfect for today might not be suitable in the near future. In addition, successful, liveable streets are well maintained streets; therefore, street management should be treated as equally important as the design. Regular evaluations of users and usages are needed in order to modify the street accordingly if change happens. Regular street management includes the operation of removing trash, sweeping and keeping the sidewalk clean, repairing and replacing furniture, maintaining trees and plants, etc. Local stakeholders might engage in some of these activities, and they might be supported by small and flexible funding schemes provided, e.g., on the city level.

### 3. Recommendations of Facilities for Walking and Place Activities in Guidance Material on Urban Street Design

*3.1. Methodology for Collating and Synthesising Guidance Material*

Data on guidance material for facilities for walking and place activities were gathered based on the MORE project (Multimodal Optimisation of Roadspace in Europe, https://www.roadspace.eu/ (accessed on 14 August 2021), which brings together urban street designers from all over Europe. This project provides the unique opportunity to assemble guidance material on urban street design in local languages, to combine it into a standardized, approach as well as to gather background information about how this material is generated and used in daily planning practice. Guidelines and additional material in English—but also in various local languages—could therefore be synthesised for various European countries and, in particular detail, for the MORE city and corresponding country partners of Budapest, Constanta, Lisbon, London, and Malmö. Questionnaires with the following blocks of questions have been sent out to partners as the basis for collating relevant material: genesis and responsibilities for developing guidance, systems of road function classification, objectives and performance indicators for urban street design, specific recommendations for each street user group (pedestrians, cyclists, public transport, private motorised traffic, kerbside activities, etc.), and safety issues.

Partners from the MORE project filled in the questionnaires and provided relevant material. Intense discussions and feedback loops for translating materials and for compiling consistent information for all cities and countries followed and led to standardised comparisons for all street user groups. Further materials from other countries beyond the MORE partners have been included in order to get a broad picture of international practice in urban street design. Gerike et al. [6] have provided further information on this methodological approach.

The focus of this paper is on pedestrians and place activities. For these user groups, we analysed and summarised the following aspects in Table 1:

- Space requirements for moving pedestrians (movement function)—What width is assumed for "standard" pedestrians and for pedestrians with increased space require-

ments such as wheelchair users? Space requirements for two or more pedestrians are also provided in some references and included in Table 1. The reason for this is that sidewalks are never used in only one direction. Pedestrians are free to move in any direction on either side of the street and they extensively make use of this capability. This must be considered when designing pedestrian facilities;

- Space requirements for sidewalk equipment—What width is assumed in the guidelines for the various items that might be placed on sidewalks such as street furniture or greenery?
- Standard widths of sidewalks—How are the space requirements for movement and the place function translated into sidewalk widths? Which widths are recommended for sidewalks under differing conditions?
- Components/zones of sidewalks—Some references distinguish different zones of sidewalks;
- Recommendations on place functions—This part of the table summarises recommendations for supporting place functions of sidewalks;
- Crossing facilities—Besides the sidewalks, crossing facilities are very important for pedestrians as a vulnerable and highly detour-sensitive user group; recommendations on this topic are therefore also included in the table.

*3.2. Summary of Recommendations Provided from Guidance Material*

Table 1 combines the information taken from the researched guidance material on urban street design to provide an easily accessible comparative overview of the standards in the different countries and cities.

The combined research material shows that standards for space requirements of pedestrians are provided in most references and are comparable to one another. The width of a standard pedestrian varies between 0.55 m and 1.00 m. The main reason for this range seems to be the different definitions, as some references include (and others exclude) buffer space in the provided dimensions for standard pedestrians. Values for two pedestrians are given with few exceptions, and vary between 1.50 m and 2.00 m. Only the German guidelines on urban street design are clear and exacting, specifying that sidewalks should generally be scaled based on space requirements for two pedestrians. This specification is based on the fact that pedestrians walk in either direction on a sidewalk and that sidewalks should be generally designed in a way that allows two pedestrians walking in opposite directions to meet and pass each other.

Measurable differences were identified among buffer zones; these ranged from 0.00 m to 1.00 m. The criteria used for choosing buffer zone widths for each design task are consistent across locations. Buffers to the carriageway depend on speed and volume of motorised traffic. Buffers to the edge of the street depend on the type and size of adjacent buildings. However, the values themselves differ greatly.

The fairly similar space requirements for pedestrians summarised above translate within the researched guidance material into very different recommended sidewalk widths ranging from 1.00 m upwards. This wide range shows the difficulty of integrating adequate sidewalk widths into urban street layouts. A sidewalk of 1.00 m means that one standard pedestrian with an assumed width of 0.75 m can walk on this sidewalk with about 0.12 m buffer on both sides. One pedestrian needs to leave the sidewalk if two pedestrians walking in opposite directions meet each other. A wheelchair user with a width of 0.90 m can use this sidewalk with a 0.05 m buffer to each side. On the one hand, this is not very comfortable, and, on the other hand, it is also a safety issue when pedestrians use the carriageway when meeting each other. The authors of the guidance material are definitely aware of pedestrian space requirements and of the problems that might result from very narrow sidewalks. Nevertheless, they include these low values for sidewalk widths into their recommendations. The main reason for this is space scarcity. Particularly in historic city centres, it is rarely possible to accommodate all user requirements into the limited available street space. Low minimum values, e.g., for sidewalk widths, could help with

finding compromises for such challenging design tasks, and in the minds of the authors of the guidance material, these low values can be applied for pedestrians more easily than, e.g., for buses, which simply cannot pass a cross-section when lanes are too narrow.

Some references provide specific guidance for bottlenecks; these might help in such cases. For example, Transport for London [50] allows for a minimum width of the footway clear zone of 1.00 m, and for a maximum length of 6 m. Two pedestrians cannot meet each other here, but they might wait at a passing point until the bottleneck is cleared and can be passed. The Municipal Chamber of Lisbon [47] recommends coexistence streets (shared spaces) in situations of limited space availability; further references recommend taking out selected functions completely (such as parking), and thus allowing for regular widths for the remaining elements in the street [45].

The criteria for choosing sidewalk widths beyond minimum values are (1) the street type (Budapest/Hungary, Lisbon, London, Madrid, Malmö, Germany, Spain, The Netherlands), (2) speeds and volumes of motorised traffic (Austria, Germany, NACTO), (3) pedestrian volumes (Budapest/Hungary, Constanta, London, Switzerland), (4) the existence of parking or cycling facilities (Austria), (5) or proximity to specific destinations such as schools or retirement homes (Germany, Malmö).

The criteria for distinguishing street types for criterion (1) are based on road function classification, using mainly one-dimensional systems such as urban and district roads/local collector roads/local access roads in Madrid, or residential streets/major streets/commercial streets in Budapest. London's [70] approach to movement and place functions is a two-dimensional system for road function classification that disentangles user requirements in terms of pedestrian movement (walking) and place activities (staying). It is thus more detailed and better suitable for designing sidewalks that fit specific user needs in each of the two dimensions. Some references describe street types based on specific street characteristics, such as the location of the street section (e.g., inner versus outer city, proximity to specific destinations such as schools or retirement homes), characteristics and usage of adjacent buildings or traffic (e.g., volumes of motorised vehicles); these characteristics show an overlap with the more specific criteria (2) to (5).

The second criterion of speeds and volumes of motorised traffic focusses on safety and buffer zones. The third criterion (pedestrian volumes) seems to be very suitable for optimally matching sidewalk design and user needs. The disadvantage of this criterion is that it is based on the status quo and not on anticipated or desired pedestrian volumes. In addition, it is difficult to apply because of insufficient knowledge on pedestrian volumes. Discussions with city partners in the MORE project revealed that pedestrian volumes are hardly considered for sidewalk design, even when these are listed as criteria in the local or national guidance material, mainly because of a lack of data availability. Criterion (4) again focusses on safety and buffer zones, while criterion (5) is a suitable input for deciding on sidewalk width and is frequently applied.

More sophisticated references provide not only recommendations for the overall sidewalk width, but also give additional recommendations for different zones of the sidewalk [45,47,50,54,67]. This approach allows for a clear separation of movement and place functions. The footway clear zone (also called pedestrian through zone) is the part of the sidewalk that should be kept clear from any obstacles and that is dedicated to the movement function; it should allow pedestrians to move safely and comfortably. The recommended minimum width for footway clear zones is 1.20 m in Lisbon (on existing 4th or 5th level streets); 1.50 m in Budapest, Constanta (street category III), London (acceptable minimum) and the U.S.; 1.80 m in Germany, Madrid, Lisbon (for new streets), Spain and The Netherlands, and 2.00 m in Austria, London and Switzerland 2.00 m as the preferred minimum.

Table 1. Recommendations for Pedestrian Facilities in Guidance Material on Urban Street Design.

| | Austria | Budapest | Constanta | Germany | Lisbon | London | Madrid | Malmö | NACTO | Spain | Switzerland | The Netherlands | Summary |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Space Requirements (Width)** | | | | | | | | | | | | | |
| Standard Pedestrian | 1.00 m (including buffer on each side) | 0.55 m + 0.10 m buffer on each side = 0.75 m | No rec. | 0.80 m (value is given only in one figure where 2 pedestrians are shown) | 0.655 m | 0.75 m | 0.55 m | 0.70 m | No rec. | 0.65 m | 0.80 m/1.00 m | 0.80–1.00 m (person with pram narrowest width 0.80 m, person with suitcases widest width 1.00 m) | 0.55–1.00 m (with and without buffer space) |
| Two or More Pedestrians | 2.00 m, min. 1.50 m | Adult + child: 1.30 m + 0.10 m buffer on each side = 1.50 m Family (2 adults + 2 children): 2.80 m + 0.10 m buffer on each side = 3.00 m | No rec. | Two standard pedestrians: 1.80 m (each pedestrian 0.80 m + 0.20 m buffer in between) | Two standard pedestrians: 1.50 m | Two standard pedestrians: 1.50 m Adult + child: 1.20 m | ≥1.14 m (two persons holding each other tightly) | No rec. | No rec. | 1.50 m two standard pedestrians 1.60 m standard pedestrian + wheelchair 1.80 m three standard pedestrians | 2.00 m two standard pedestrians | 2.10 m (two pedestrians with suitcases) | Adult + child: 1.20–1.50 m Two pedestrians: 1.14–2.10 m Family: 3.00 m |
| **Increased Space Requirements (Width)** | | | | | | | | | | | | | |
| Blind Person with Assistance | No rec. | No rec. | No rec. | 1.20–1.30 m | No rec. | 1.20 m | No rec. | 1.20 m | No rec. | No rec. | 1.25 m | No rec. | 1.20–1.30 m |
| Person with Walking Cane | No rec. | 0.80 m + 0.10 m buffer on each side = 1.00 m | 0.95 m | 0.85–1.20 m | No rec. | 0.75 m | 1.21 m | No rec. | No rec. | 0.80 m | 1.25 m | No rec. | 0.75–1.25 m |
| Person with Crutches | 1.00 m | 0.80 m + 0.10 m buffer on each side = 1.00 m | 0.90 m | 1.00 m | No rec. | 0.90 m | 0.79 m | No rec. | No rec. | 0.80 m | 1.00 m | No rec. | 0.79–1.00 m |
| **References** | | | | | | | | | | | | | |
| | [41] | [42] | [43,44] | [45,46] | [47] | [48–50] | [51] | [52,53] | [54] | [55] | [56] | [57,58] | |

**Table 1.** *Cont.*

| | Austria | Budapest | Constanta | Germany | Lisbon | London | Madrid | Malmö | NACTO | Spain | Switzerland | The Netherlands | Summary |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Increased Space Requirements (Width)** | | | | | | | | | | | | | |
| Person in a Wheelchair | 0.90 m | 0.80 m + 0.10 m buffer on each side = 1.00 m | 0.80 m | 0.90 m | 0.90 m × 1.20 m, two persons with wheelchair ≥1.80 m | 0.90 m | 0.84 m | 0.80 m | No rec. | 0.90 m | 1.25 m | 0.85–0.90 m | 0.80–1.25 m |
| Person with Luggage | 1.00 m | 0.80 m + 0.10 m buffer on each side = 1.00 m | No rec. | No rec. | No rec. | No rec. | 1.24 m | No rec. | No rec. | 0.90 m | 1.25 m | 1.00 m | 0.90–1.25 m |
| Person with Pram | No rec. | 0.55 m + 0.10 m buffer on each side = 0.75 m | No rec. | 1.00 m | No rec. | Plus one adult beside: 1.50 m | 1.62 m (with pram and child) | 0.70 m | No rec. | 0.80 m | 1.00 m | 0.80 m | 0.75–1.50 m |
| **Space Requirements for Street Furniture (Width)** | | | | | | | | | | | | | |
| Benches | 1.00 m | No rec. | No rec. | ≥1.00 m | ≥1.20 m | ≥0.50 m (space to be kept clear in front of a bench) | ≥0.60 m (space to be kept clear in front of a bench) | 2.00 m | No rec. | ≥1.20 m | ≥1.50 m | ≥1.20 m | 1.00–2.00 m |
| Green Space without Trees | No rec. | No rec. | No rec. | ≥1.00 m | No rec. | No rec. | No rec. | No rec. | No rec. | No rec. | ≥1.50 m | No rec. | ≥1.00 m |
| Green Space with Trees | No rec. | No rec. | 0.75–1.00 m | 2.00–2.50 m | No rec. | No rec. | 1.00 m, 1.50 m for bigger trees | > 2.50 m | No rec. | 1.20 m | No rec. | Rec. to provide green at borough level | ≥0.75 m |
| Waiting Area at PT Stops | 1.50–4.25 m (depending on pass. volume) | ≥1.50 m | ≥2.00 m (≥2.80 m with cyclists) | ≥1.50 m (excluding the shielded area) | ≥2.60 m | Wide enough | ≥1.50 m | 2.30 m | 1,83–3.05 m | 1.50 m (including the shielded area) | ≥1.50 m | ≥1.80 m (bus stop, whole sidewalk width) | ≥1.50 m |
| **References** | | | | | | | | | | | | | |
| | [41] | [42] | [43,44] | [45,46] | [47] | [48–50] | [51] | [52,53] | [54] | [55] | [56] | [57,58] | |

28

**Table 1.** *Cont.*

|  | Austria | Budapest | Constanta | Germany | Lisbon | London | Madrid | Malmö | NACTO | Spain | Switzerland | The Netherlands | Summary |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Standard Width of Sidewalks** | | | | | | | | | | | | | |
| Width and Conditions | Standard widths: 2.00 m (residential streets or SL ≤ 40 km/h or next to parallel parking or independent sidewalk off the carriageway), 2.30 m (next to cycle lane), 2.50 m (SL=50km/h or next to perpendicular parking), 3.00 m (SL ≥ 60 km/h). Increased width for higher ped. volumes. Min. widths are provided in addition | Min. width 1.50 m, recommended width 3.00 m Useable width of sidewalk: 1.5 m + n × 0.75 m (n = number of pedestrians) Width of sidewalk depends on street type, available space and pedestrian volume. Recommended width per street type: Living/residential street: 1.50–3.00 m Major street: ≥3.00 m Commercial street: ≥4.50m PT stop area: ≥3.00 m | 1.00 m–4.00 m depending on pedestrian volumes | Standard width: 2.50 m (1.80 m for two persons + buffer to adjacent buildings and carriageway) Wider sidewalks for higher density, higher usage of adjacent buildings, high frequency PT Wider sidewalks also in the vicinity of specific destinations such as retirement homes, schools, shopping centres | Min. width: General (including trees, lighting, etc.): 3.00 m Usable width in new streets: 2.00 m Usable width in pre-existing streets: ≥1.20 m on 4th/5th level streets ≥1.50 m on 2nd/3rd level streets ≥1.50 m in every other situation Coexistence streets (shared space) in cases of space scarcity | Min. width: 2.00 m in lightly used streets (such as those with a purely residential function) The width of the sidewalk varies depending on pedestrian volumes | Urban and district roads 6 m (4 m min.) Local collector roads (2 lanes) 6 m (3 m min.) Local collector roads (4 lanes): 6 m (4 m min.) Local access roads 3 m (2.50 m min.) Min. for bottlenecks 2.00 m Rec. for pedestrian main routes 6 m | Min. width: 2.00 m In inner city environment next to higher buildings the sidewalk should not be less than 2.50 m | Desired minimum clear zone of 2.13 m and an absolute minimum of 1.52 m Where a sidewalk is directly adjacent to moving traffic, the desired minimum is 2.44 m, providing a minimum 0.61 m buffer for street furniture and utilities | Urban and district roads 6 m (4 m min.) Local collector roads (2 lanes) 6 m (3 m min.) Local collector roads (4 lanes): 6 m (4 m min.) Local access roads 3 m (2.50 m min.) Min. for bottlenecks 2.00 m Rec. for pedestrian main routes 6 m | Standard width: 2.50 m (min. 2.00 m) Rec. widths: clear zone (2.00–3.00 m, depending on ped. volumes and proportion of persons with increased space requirements) plus buffer space | Standard 1.80 m (sufficient for two wheelchair users or wheelchair and pram to pass) Min. 1.20 m (max. 10 m length) Min. 0.9 m (bottlenecks, e.g., to pass a light pole) Shopping streets: ≥1.5 m (<100 p/h); ≥2.4 m (>100 p/h) ≥4.0 m free space in main shopping streets | Width: ≥1.00 m Dependencies: Land use and height of buildings Street type Available space Pedestrian volume Frequency of PT Existing vs. new streets |
| **References** | | | | | | | | | | | | | |
|  | [41] | [42] | [43] | [45,46] | [47] | [48–50,59] | [51] | [52,53,60] | [54] | [61] | [56,62] | [57,58,63] |  |

**Table 1.** *Cont.*

| | Austria | Budapest | Constanta | Germany | Lisbon | London | Madrid | Malmö | NACTO | Spain | Switzerland | The Netherlands | Summary |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Components/Zones of Sidewalks** | | | | | | | | | | | | | |
| Clear Zone | 2.00 m, min. 1.50/1.20 m (max.1 m length) and 0.90 m at one point | ≥1.50 m | Street category I: 2.00 m Street category II: 1.50 m Street category III: 1.00–1.50m | 1.80 m | Min. 1.80 m in new streets Min. 1.50 m on 2nd/3rd level streets Min. 1.20 m on 4th/5th level streets | ≥2.00 m (preferred minimum, unobstructed width) | 1.80 m | Should be provided, but no information on width | 1.52–2.13 m in residential settings 2.44–3.66 m in downtown/commercial areas | Min. ≥ 1.50 mm Rec. 1.80 m | See sidewalk widths above | See sidewalk widths above | ≥1.00 m |
| Buffer to Adjacent Buildings | No rec. | 0.50 m | ≥1.00 m | 0.20 m (0.00 m in case of no buildings or low fences) | ≤0.60 m | 0.30 m | 0.45 m | Should be provided, but no information on width | No rec. | 0.50 m | ≥0.20 m | No rec. | 0.00–1.00 m |
| Buffer to Carriageway/Kerb Zone | No buffer (SL ≤ 40 km/h) 0.50 m (SL = 50 km/h) 1.00 m (SL ≥ 60 km/h) | 0–30 km/h: 0.00 m 31–50 km/h: 0.25 m 51–70 km/h: 0.50 m 71–100 km/h: 1.00 m | 0.25 m | 0.50 m in standard busy streets 0.30 m in case of low goods traffic and residential streets | 0.30 m | 0,45–0.60 m | 0.40 m | Should be provided, but no information on width | 0.61 m Otherwise enhancement zone with bike lanes, parklets or kerb extensions | 0.50 m | 0.20–0.50 m for SL ≥ 50 km/h ≥0.20 m to parallel parking ≥0.50 m to perpendicular parking | No horizontal space requirements, but required bicycle paths or lanes along distributor roads (SL ≥ 50 km/h) located between the carriageway and the sidewalk No buffer between bicycling path and pedestrian area | 0.00–1.00 m |
| **References** | | | | | | | | | | | | | |
| | [41] | [42] | [43,44] | [45,46] | [47] | [48–50] | [64] | [52,53,60, 65] | [54] | [61] | [56] | [58,63] | |

**Table 1.** *Cont.*

| | Austria | Budapest | Constanta | Germany | Lisbon | London | Madrid | Malmö | NACTO | Spain | Switzerland | The Netherlands | Summary |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Components/Zones of Sidewalks** | | | | | | | | | | | | | |
| Furniture Zone | No rec. | 1.00 m | No rec. | ≥1.00 m Reference values for width for specific place functions, higher widths for specific street types, next to specific POIs | Parklets: 2.00–2.50 m; Terrace/gastronomy: ≥2.00 m if terrace is provided, clear zone ≥2.00 m | 0.50–2.00 m with detailed information on space requirements of different types of street furniture | Kiosks 1.20 m | Should be provided, but no information on width | Yes, but no information on width For parklets: 1.68 m | Min 1.00 m, kiosks 2.50 m, light poles 0.70–1.00 m, trash cans 0.90 m, terrace 2.10–2.50 m, water fountains 1.50 m Min. 0.40 m between furniture and kerb to carriageway | See frontage zone | General recommendation for place advertisement, information panels, traffic sign, light poles and the like, outside of walking routes | 0.50–2.00 m |
| Frontage Zone | 1.00 m | 1.00–1.50 m | No rec. | ≥1.00 m | Shop displays and showcases: 1.00 m | Yes, but no information on width | 0.90 m | Should be provided, but no information on width | Yes, but no information on width | 1.50 m | ≥1.20 m for vitrines/sales booths, ≥0.50 m for street café, ≥0.20 m for advertisement/info panels/parking meter | 0.90 m clear from obstacles such as parked bicycles so that facades and fences can offer guidance to visually impaired pedestrians | ≥1.00 m |
| **References** | | | | | | | | | | | | | |
| | [41] | [42] | [43,44] | [45,46] | [47] | [48–50] | [64] | [52,53,60,65] | [54] | [61] | [56] | [58,63] | |

**Table 1.** *Cont.*

| Place Function | Austria | Budapest | Constanta | Germany | Lisbon | London | Madrid | Malmö | NACTO | Spain | Switzerland | The Netherlands | Summary |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Recommendations on Place Function** | | | | | | | | | | | | | |
| | No rec. | Staying, waiting, leaning against the wall: 0.70–1.00 m Two to three persons chatting/sitting: 1.50–2.00 m Places to stay (e.g., gastronomy, benches) or to play: 2.50–3.00 m Spacious places to stay or to play: ≥4.00 m | No rec. | Installation of benches at appropriate intervals Public spaces should primarily be created by widening a section of sidewalk in addition to the area provided for moving along. This also includes the creation of play spaces. | Installation of benches at appropriate intervals: 50–150 m | Seating on key pedestrian routes should be considered every 100 m to provide rest points and to encourage street activity, max. recommended spacing interval on high streets and city places 50 m Places to stay/chat: ≥2.50 m Places to play: ≥4.00 m | Benches: One every 30 m on sidewalks Playgrounds areas: 20 m² for every 100 m² of green spaces, parks or boulevards | Installation of benches every 25 m in pedestrian zones otherwise every 50 m Next to benches, garbage bins should be installed | Street space can be reused for different purposes, such as parklets, bike share, and traffic calming | Benches every 50 m in pedestrians routes | No rec. | Benches at intervals of about 100 m in city centres and near retirement homes, 200 m outside city centres, ≥1.20 m in front of a bench Further qualitative recommendations: Create car-free city centres, provide sufficient resting areas, provide green facilities at the level of boroughs, e.g., parks, dog walking areas, and playground Improve the attractiveness at street level with interesting architecture, green spaces such as planting, or rows of trees along a road, and water such as a pond or ditch | Benches: 25–150 m Places to stay/chat: 1.50–2.50 m Places to play: ≥4.00 m |
| | [41] | [42] | | [45,46] | [47] | [48–50] | [64] | [52,53,60] | [54] | [66] | | [58,63] | |

**Table 1.** *Cont.*

| | Austria | Budapest | Constanta | Germany | Lisbon | London | Madrid | Malmö | NACTO | Spain | Switzerland | The Netherlands | Summary |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Recommended Crossing Designs** | | | | | | | | | | | | | |
| Crossing Designs | Pedestrian crossing (Zebra Crossing) Traffic signal central median Under-/overpass (exemption) | Central median Physical without priority (plateau/raised block-paved area) Pedestrian crossing (Zebra Crossing) Pedestrian crossing with physical measures Traffic signal Under-/overpass | No rec. | Central median Physical without priority (plateau/raised block-paved area) Pedestrian crossing (Zebra Crossing) Pedestrian crossing with physical measures Traffic signal Under-/overpass | Central median Pedestrian crossing (Zebra Crossing) Traffic signal | Central median Physical without priority (plateau/raised block-paved area) Pedestrian crossing (Zebra Crossing) Pedestrian crossing with physical measures Traffic signal Under-/overpass | Central median Pedestrian crossing (Zebra Crossing) Traffic signal Safety islands in big roads Under-/overpass | Central median Physical without priority (plateau/raised block-paved area) Pedestrian crossing (Zebra Crossing) Pedestrian crossing with physical measures Traffic signal | Central median or central island Pedestrian crossing (e.g., Zebra Crossing) Traffic signal | Pedestrian crossing Central median Traffic signal | Pedestrian crossing (Zebra Crossing) Traffic Signal Central Median, refuge Under-/overpass (exemption) Pedestrian zone, shared space | Pedestrian crossing; Zebra Crossing Signal controlled crossing Under-/overpass The first two types of crossings may be combined with a median island Speeds should be reduced to 30 km/h at zebra crossings | Central median Physical without priority (plateau/raised block-paved area) Pedestrian crossing (Zebra Crossing) Pedestrian crossing with physical measures Traffic signal Under-/overpass |
| **Criteria for Selecting Types of Crossing Facilities** | | | | | | | | | | | | | |
| Criteria | Criteria: ped. and veh. volumes No need for crossing facility: veh. volumes < 100 v/h Crossing facility recommended: 100–300 v/h (depending on ped. volumes) Crossing facility mandatory: higher veh. volumes | Crossing facilities are necessary if there is a distinct crossing need; Traffic volume > 1000 v/h, speed limit 50 km/h; or Traffic volume > 500 v/h and speed limit > 50 km/h. | No rec. | Crossing facilities are necessary if there is a distinct crossing need; Traffic volume > 1000 mot. veh. per hour, SL = 50 km/h; or Traffic volume > 500 v/h, SLt > 50 km/h. | Zebra Crossings should be used whenever no traffic lights could be provided, To reduce speed and to avoid accidents | Zebra Crossings are only recommended for low-speed environments, 35 mph or less Underpass only under exceptional circumstances with high pedestrian demand | 3 main criteria: Veh./ped. volumes, road hierarchy, land use Pedestrian crossings when >1000 v/h and >100 ped/h Local roads: Zebra Crossings Collector roads zebra crossings or traffic signal Urban and district roads, traffic signal crossings or underpass | Crossing design depends on: Traffic safety Whether a carriageway or a cycle path should be crossed | On streets with higher volume higher speeds (> 20 mph), or more lanes (2+), crosswalks should be provided At places with high ped. demand marked crossings may be beneficial regardless of traffic conditions | No rec. | Criteria: Local or linear crossing needs, widths of carriageway, vehicle volumes, speed limit, pedestrian volumes, proportion of pedestrians with special requirements (proximity to destinations such as schools) | Road category (distributor or access road) Location at crossing or road section (and local situation such as nearby a school) | Typical criteria: Crossing need for pedestrians High traffic volume High speed High pedestrian demand |
| | [41] | [67] | | [45,46] | [47] | [50] | [64] | [68] | [54] | [66] | [62,69] | [63] | |

No rec. = no recommendation; PT = public transport; Veh = vehicles; Min = minimum; AADT = annual average daily traffic; rec. = recommendation; SL = speed limit; p/h = persons per hour; Ped. = pedestrians.

The frontage zones, furniture zones, and kerb zones are spaces that are dedicated to place functions or that serve as buffer zones, as described above. Recommendations for place functions are very technical in the researched guidance material, and include mainly space requirements for street furniture such as benches, parklets, terraces, gastronomy tables/seating, waiting areas at public transport stops, or parking facilities for bicycles. Malmö is the most advanced in providing space requirements for greenery. Transport for London [50,59] lists possible place activities for different widths of the furniture zone. Some references work with pictograms to visualise possible sidewalk usages for specific sidewalk widths; for example, they provide a pictogram showing a group of pedestrians who chat and give the necessary sidewalk width for this scenario [6]. Provision for place functions is additionally included in the increased sidewalk width for specific street types, as described above.

Overall, the focus of the researched guidance material is clearly on the movement function for pedestrians; rarely is any information given about how to design pleasant spaces for place users that fit to the human dimension and that encourage users to stay, sit, chat, etc.

### 4. Comparison of Empirical Evidence and Guidance Material in Urban Street Design

Empirical evidence in the researched literature consistently shows the dominance of the D variables for pedestrian volumes, including pedestrian movement and place activities. Density, Diversity of land uses and Distance to public transport are significant determinants of walking and, with less comprehensive empirical evidence, also for place activities in all the studies identified in the literature research. Streetscape also matters, but with less importance compared to the D variables at the neigbourhood level. Floor area ratios, the proportion of retail frontage or other active uses of the adjacent buildings, as well as façade design, are the most important variables at the street level. Transparency at the ground floor level is of particular relevance; people like to see what happens inside the buildings next to the street. These street characteristics, as well as the D variables on the neighbourhood level, are shaped by urban planning rather than by transport engineering.

Sidewalk width, street furniture and amenities are the relevant variables related to actual street design. Sidewalk width shows ambiguous causality: wider sidewalks are implemented in locations with observed or anticipated high pedestrian volumes, and they allow the placing of (more) street furniture and amenities, thus inviting pedestrian activities. Empirical evidence clearly shows that street furniture and particularly seating increase pedestrian volumes, and the relationship between sidewalk width (other things being equal) and pedestrian volumes is thus clear.

The comparison of this empirical evidence in the scientific literature with the compiled guidance material shows that they are not well linked. Guidance material for pedestrian facilities focusses on space requirements for specific furniture and usages of sidewalks. Recommendations on which sidewalk design to choose in a specific location are based on criteria that focus on safety and buffer zones (e.g., existence of parking), pedestrian volume (a criterion that is hardly measured and only represents the current situation), or street types, without good support from scientific evidence. The street type approach as such, in combination with the proximity to relevant destinations, seems to be the most suitable criterion for deciding on sidewalk width and design. However, it should make use of the determinants for pedestrian movement and place activities, as these have been identified in the literature. These are the D variables, particularly Density, Diversity and Distance to public transport. In terms of classification, the characteristics of adjacent buildings, particularly at street level, should be considered as one criterion for defining the street type. Based on street type classification, recommendations should be given for sidewalk widths, design and equipment. These should cover both the movement function for pedestrians (walking) and place activities.

## 5. Recommendations for Advancing Guidance on Urban Street Design

Based on the findings so far, this section develops recommendations for pedestrian facilities in future guidelines on urban street design.

Movement Function:

- In a supply-oriented approach, an adequate standard width for sidewalks should be provided in the guidance material as a basic standard value, independent of expected pedestrian volumes. These should include a footway clear zone that allows two pedestrians to meet/pass each other and buffer zones to adjacent usages;
- For the footway clear zone that should be kept free of any obstacles, a minimum value of 1.80 m seems to be suitable. This is the width that allows two standard pedestrians to pass each other (0.80 m + 0.20 m + 0.80 m). This value would be 1.90 m if the goal were to allow one standard pedestrian and one wheelchair user to pass each other (0.80 m + 0.20 m + 0.90 m). The chosen standard width for a pedestrian of 0.80 m is in the upper range of values identified in the guidance material, but seems to be suitable given the ageing population in many countries all over the world, which is related to a growing number of pedestrians with increased space needs;
- Buffer zones to buildings and the carriageway should be scaled depending on the height of the buildings and the usage of the carriageway. For residential streets with low traffic volumes and speed, small buffer values are sufficient. For busy streets with higher speeds and volumes of motorised vehicles, bigger buffer zones between the pedestrians and the moving motorised traffic are necessary ($\geq$0.30 m). Guidelines should also provide recommendations for the adequate separation of pedestrians from cyclists, scooters, and other micromobility vehicles. Guidelines might not only provide guidance on the dimensions of buffer spaces, but also on their design, with possible reference to the design-for-all principles, water treatment, and the provision for place functions;
- For street sections with higher observed or expected pedestrian volumes, greater widths for the footway clear zones should be recommended, following again a supply-oriented approach. These street sections can be identified based on the street type approach, as described above. Alternatively, pedestrian volumes can be counted. Automated counting facilities for pedestrians are increasingly available, and allow for counting at more locations and for longer time periods. Future expected or envisaged changes in pedestrian volumes need to be considered in this case;
- Guidelines should also provide recommendations for types and locations of crossing facilities. These are paramount for achieving high levels of subjectively perceived and objective safety.

Place Function:

- The street type approach also seems to be a suitable basis for providing recommendations for place functions. It allows for implicitly considering differences in place functions resulting from different types and usages of the buildings next to the street and in the neighbourhood (the D variables), as well as resulting from the vicinity to public transport stops or further specific destinations;
- Recommendations should be given for the amount of space to be provided for place functions (quantity), and also for how to design and equip this space (quality). The qualitative descriptions of requirements for benches, characteristics of attractive spaces, etc., in [57,58] might be a suitable starting point for this;
- Guidance on seating should be provided, as this variable was found to increase pedestrian activities significantly in all the researched scientific references. Seating should be preferably located near activity-supporting businesses or facilities, and it should also allow groups of people to sit together and to engage in any kind of social activity;

- Shade and shelter, street furniture, and greenery are further significant determinants of pedestrian activities and should also be included in future guidelines on urban street design in terms of location, quality, and quantity;
- Road function classification is also of great importance for pedestrians. A strategic concept for pedestrian networks, including a hierarchy of main and secondary pedestrian facilities, is the basis for deciding on extra space beyond standard values and on the equipment of sidewalks (e.g., benches or public toilets). The concept of main pedestrian arteries in Madrid is a good example of such strategic development of pedestrian networks and facilities [51].

Bottlenecks:

Bottlenecks are a major problem in planning for pedestrians. Guidance should be provided about how to deal with bottlenecks. Examples of such guidance are given above in Section 3. For example, selected functions such as parking might be taken out completely in narrow parts of a street in order to gain space for pedestrians. Shared space concepts might be a solution, as proposed for Lisbon. Low speeds and volumes of motorised traffic are necessary for successfully implementing such concepts. Gehl [12] concludes from his practical work and research that these shared space concepts only work if, firstly, priority is legally given to pedestrians. Narrow sidewalks for limited and clearly defined distances, as suggested in London and in the Netherlands, are another opportunity for dealing with bottlenecks. Narrow values such as 1.00 m should be limited in their application, as otherwise, there is the risk that these become the standard values commonly used. These standard values for sidewalk width should instead be values that allow pedestrians to at least move safely and comfortably in both directions and to meet each other.

Streets as ecosystems:

Streets are vital parts of urban ecosystems. They are places where man-made infrastructure interferes with natural systems. Street design is a significant determinant for various aspects of environmental quality at the street level itself, as well as beyond. It influences the micro-climate, as well as the exposure of street users and residents in the adjacent properties to noise and air pollution, and it is one core component of water management at the city level. Designing for streets as ecosystems is an interdisciplinary task that requires collaboration between urban, transport and environmental planning, including, e.g., public works and water departments. These aspects regarding how to provide for ecosystem services and how to maximise synergies between all the different street functions are hardly covered at all in the researched guidance material on urban street design. They should be included in future guidelines with the final goal of designing streets and cities that are resilient, efficient in moving people and goods, sustainable, and enjoyable.

The NACTO guides can be seen as a best practice example for including environmental aspects into guidance on urban street design. The Urban Street Design Guide [54] stresses the importance of planning for streets as ecosystems, and gives brief guidance on important design elements, such as stormwater management, bioswales or flow-through planters. The Urban Street Stormwater Guide [71] details these aspects with a particular focus on the important aspect of stormwater management.

## 6. Conclusions, Summary and Outlook for Further Research

Planning for pedestrians is an interdisciplinary task that requires contributions from (1) transport planning, (2) urban planning, and (3) environmental planning, as well as (4) commitment from the city, local businesses and communities, and from other local stakeholders. Our review of scientific literature has shown that all four of these aspects are important, and that no clear priorities can be identified. Some level of trade-off seems to be possible between the four criteria. For example, one weak element, e.g., in transport planning/street design, might be compensated by strong urban design and stakeholder

engagement. However, none of these four aspects can fail entirely when the goal of lively streets must be achieved.

The review of guidance material on urban street design shows that urban street designers are well advanced in measuring space requirements for pedestrians and for pedestrian facilities, but less in planning pleasant urban environments that fit the human dimension and invite pedestrian movement and place activities.

It will be neither possible nor meaningful to integrate all relevant aspects of successfully providing for pedestrian activities as identified in the scientific literature into guidelines on urban street design. However, a better linkage with scientific evidence can greatly improve the guidance material. The recommendations given in guidelines on urban street design could be far more focused on the significant aspects as identified in scientific literature, with two types of possible positive effects: In a supply-oriented approach, sidewalk width and design match with pedestrian needs and activities at each specific location. In a demand-oriented approach, wider and more attractive sidewalks including space for pedestrian movement and place activities can be provided at the most suitable locations based on scientific evidence, thus inviting people to come and stay in the streets and to support lively cities and streets, with various positive side effects.

The suggestions of more targeted recommendations for pedestrian facilities, and particularly for place functions, in future guidelines on urban street design hopefully contribute beneficially to the discussion on how to promote walking and lively streets. This could contribute to various positive side effects in overall travel behaviour, the economy and the environment. Planning for walking and place activities will only be successful if this is done in the context of all street functions and user needs. The challenge is to find the right balance between movement and place functions for all the different user groups anew for each design task.

The current COVID-19 pandemic brings new challenges, but also opportunities. Walking is one essential aspect of resilient transport systems, and has substantially increased in importance in the last few months. Insights into behavioural changes due to COVID-19 restrictions, and also into the effects of policy measures implemented in various cities all over the world for supporting social distancing and for generally promoting walking and place activities (see e.g., [6]), should feed into future guidelines.

Sufficient evidence exists in the literature that can reliably be translated into recommendations for planners and urban street designers. Further research on the determinants of walking and pedestrian place activities would help to additionally validate the findings from the studies published so far, and to elaborate on issues that have not been addressed in detail in the existing studies.

# References

1. Koszowski, C.; Gerike, R.; Hubrich, S.; Götschi, T.; Pohle, M.; Wittwer, R. Active Mobility: Bringing Together Transport Planning, Urban Planning, and Public Health. In *Towards User-Centric Transport in Europe: Challenges, Solutions and Collaborations*; Müller, B., Meyer, G., Eds.; Lecture Notes in Mobility; Springer International Publishing: Cham, Switzerland, 2019; pp. 149–171. [CrossRef]
2. Litman, T.A. Economic Value of Walkability. *Transp. Res. Rec.* **2003**, *1828*, 3–11. [CrossRef]
3. City of Malmö. Sustainable Urban Mobility Plan: Creating a more Accessible Malmö. 2016. Available online: https://ec.europa.eu/transport/sites/default/files/cycling-guidance/malmo_sump.pdf (accessed on 14 August 2021).
4. Mayor of London. Mayor's Transport Strategy. 2018. Available online: https://tfl.gov.uk/corporate/about-tfl/the-mayors-transport-strategy (accessed on 14 August 2021).
5. Austrian Ministry for Transport, Innovation and Technology (BMVIT). Masterplan Walking: Strategy for Supporting Walking in Austria. 2015. Available online: https://www.bmk.gv.at/themen/mobilitaet/fuss_radverkehr/publikationen/masterplangehen.html (accessed on 14 August 2021).
6. Gerike, R.; Dean, M.; Koszowski, C.; Schröter, B.; Wittwer, R.; Weber, J.; Jones, P. Urban Corridor Road Design: Guides, Objectives and Performance Indicators. Deliverable D1.2, Multimodal Optimisation of Roadspace in Europe. 2019. Available online: https://www.roadspace.eu/ (accessed on 14 August 2021).
7. Gerike, R.; Hubrich, S.; Koszowski, C.; Schröter, B.; Wittwer, R. Active Transport: Heterogeneous Street Users Serving Movement and Place Functions. In *International Encyclopedia of Transportation*; Vickerman, R., Ed.; Elsevier: Amsterdam, The Netherlands, 2021; pp. 140–146.
8. Gascon, M.; Götschi, T.; de Nazelle, A.; Gracia, E.; Ambròs, A.; Márquez, S.; Marquet, O.; Avila-Palencia, I.; Brand, C.; Iacorossi, F.; et al. Correlates of Walking for Travel in Seven European Cities: The PASTA Project. *Environ. Health Perspect.* **2019**, *127*, 97003. [CrossRef] [PubMed]
9. Ewing, R.; Hajrasouliha, A.; Neckerman, K.M.; Purciel-Hill, M.; Greene, W. Streetscape Features Related to Pedestrian Activity. *J. Plan. Educ. Res.* **2016**, *36*, 5–15. [CrossRef]
10. Ewing, R.H.; Clemente, O. *Measuring Urban Design: Metrics for Livable Places*; Metropolitan Planning + Design; Island Press: Washington, DC, USA, 2013.
11. Ewing, R.; Handy, S. Measuring the Unmeasurable: Urban Design Qualities Related to Walkability. *J. Urban Des.* **2009**, *14*, 65–84. [CrossRef]
12. Gehl, J. *Cities for People*; Island Press: Washington, DC, USA, 2010.
13. Mehta, V. *The Street: A Quintessential Social Public Space*; Routledge: London, UK, 2013.
14. Mehta, V.; Bosson, J.K. Revisiting Lively Streets: Social Interactions in Public Space. *J. Plan. Educ. Res.* **2018**, *14*, 0739456X1878145. [CrossRef]
15. Ewing, R.; Cervero, R. Travel and the Built Environment. *J. Am. Plan. Assoc.* **2010**, *76*, 265–294. [CrossRef]
16. Cervero, R.; Duncan, M. Walking, Bicycling, and Urban Landscapes: Evidence from the San Francisco Bay Area. *Am. J. Public Health* **2003**, *93*, 1478–1483. [CrossRef]
17. Götschi, T.; de Nazelle, A.; Brand, C.; Gerike, R. Towards a Comprehensive Conceptual Framework of Active Travel Behavior: A Review and Synthesis of Published Frameworks. *Curr. Environ. Health Rep.* **2017**, *4*, 286–295. [CrossRef]
18. Stead, D.; Marshall, S. The Relationships between Urban Form and Travel Patterns. An International Review and Evaluation. *Eur. J. Transp. Infrastruct. Res.* **2001**, *1*, 113–141.
19. Lai, Y.; Kontokosta, C.E. Quantifying place: Analyzing the drivers of pedestrian activity in dense urban environments. *Landsc. Urban Plan.* **2018**, *180*, 166–178. [CrossRef]
20. Ewing, R.; Connors, M.B.; Goates, J.P.; Hajrasouliha, A.; Neckerman, K.; Nelson, A.C.; Greene, W. Validating Urban Design Measures. In Proceedings of the TRB 92nd Annual Meeting Compendium of Papers, Washington, DC, USA, 13–17 January 2013.
21. Cervero, R.; Kockelman, K. Travel demand and the 3Ds: Density, diversity, and design. *Transp. Res. Part D Transp. Environ.* **1997**, *2*, 199–219. [CrossRef]
22. Kim, S.; Park, S.; Jang, K. Spatially-varying effects of built environment determinants on walking. *Transp. Res. Part A Policy Pract.* **2019**, *123*, 188–199. [CrossRef]
23. Berrigan, D.; Pickle, L.W.; Dill, J. Associations between street connectivity and active transportation. *Int. J. Health Geogr.* **2010**, *9*, 20. [CrossRef] [PubMed]
24. Vale, D.S.; Saraiva, M.; Pereira, M. Active accessibility: A review of operational measures of walking and cycling accessibility. *JTLU* **2013**, *9*, 1. [CrossRef]
25. Garau, C.; Annunziata, A.; Coni, M. A Methodological Framework for Assessing Practicability of the Urban Space: The Survey on Conditions of Practicable Environments (SCOPE) Procedure Applied in the Case Study of Cagliari (Italy). *Sustainability* **2018**, *10*, 4189. [CrossRef]
26. Moura, F.; Cambra, P.; Gonçalves, A.B. Measuring walkability for distinct pedestrian groups with a participatory assessment method: A case study in Lisbon. *Landsc. Urban Plan.* **2017**, *157*, 282–296. [CrossRef]
27. Cambra, P. Pedestrian Accessibility and Attractiveness. Indicators for Walkability Assessment. Ph.D. Thesis, Technico Lisbon, Lisbon, Portugal, 2012.

28. Cambra, P.; Moura, F.; Gonçalves, A.B. On the correlation of pedestrian flows to urban environment measures: A space syntax and walkability analysis comparison case. In Proceedings of the 11th International Space Syntax Symposium, Lisbon, Portugal, 3–7 July 2017.
29. Ameli, S.H.; Hamidi, S.; Garfinkel-Castro, A.; Ewing, R. Do Better Urban Design Qualities Lead to More Walking in Salt Lake City, Utah? *J. Urban Des.* **2015**, *20*, 393–410. [CrossRef]
30. Hamidi, S.; Moazzeni, S. Examining the Relationship between Urban Design Qualities and Walking Behavior: Empirical Evidence from Dallas, TX. *Sustainability* **2019**, *11*, 2720. [CrossRef]
31. Kang, C.-D. The effects of spatial accessibility and centrality to land use on walking in Seoul, Korea. *Cities* **2015**, *46*, 94–103. [CrossRef]
32. Mehta, V. Lively Streets. *J. Plan. Educ. Res.* **2007**, *27*, 165–187. [CrossRef]
33. Njeru, A.M.; Kinoshita, I. Stay Activities on Carfree Neighborhood Shopping Streets. *E-BPJ* **2018**, *3*, 14. [CrossRef]
34. Hall, C.M.; Ram, Y. Measuring the relationship between tourism and walkability? Walk Score and English tourist attractions. *J. Sustain. Tour.* **2019**, *27*, 223–240. [CrossRef]
35. Mehta, V. Look Closely and You Will See, Listen Carefully and You Will Hear: Urban Design and Social Interaction on Streets. *J. Urban Des.* **2009**, *14*, 29–64. [CrossRef]
36. Mindell, J.S. Street Mobility Project: How to Do a Survey: A Guide for Local Authorities, Voluntary Organisations and Community Groups. 2017. Available online: http://discovery.ucl.ac.uk/1540725/1/Mindell_How%20to%20do%20a%20survey.pdf (accessed on 14 August 2021).
37. Cain, K.L.; Millstein, R.A.; Geremia, C.M. Microscale Audit of Pedestrian Microscale Audit of Pedestrian Streetscapes (MAPS): Data Collection & Scoring Manual. 2012. Available online: https://activelivingresearch.org/microscale-audit-pedestrian-streetscapes (accessed on 14 August 2021).
38. Transport for London. Healthy Street Checks for Designers. 2019. Available online: https://tfl.gov.uk/corporate/about-tfl/how-we-work/planning-for-the-future/healthy-streets (accessed on 14 August 2021).
39. Gehl, J.; Svarre, B. *How to Study Public Life*; Island Press: Washington, DC, USA, 2013.
40. Mehta, V. Streets and social life in cities: A taxonomy of sociability. *Urban Des. Int.* **2019**, *24*, 16–37. [CrossRef]
41. Austrian Research Association Road-Rail-Traffic (FSV). *Designing Pedestrian Facilities: RVS 03.02.12*; FSV: Vienna, Austria, 2015.
42. MAUT. Traffic Facilities in Roads for Disabled Persons. e-UT 03.05.12. 2009. Available online: http://ume.kozut.hu/dokumentum/49 (accessed on 26 April 2019). (In Hungarian).
43. Institutul Roman de Standardizare. *Streets: Sidewalks, Footways and Bicycle Tracks, Design Specifications: STAS 10144/2-91*; Institutul Roman de Standardizare: Bucarest, Romania, 2010. (In Romanian).
44. Dumitrescu, D. *City of Constanta: Lupascu, George. Transport Infrastructure in Constanta*; Telco: Paris, France, 2019.
45. FGSV. *Directives for the Design of Urban Roads: Translation 2012*; FGSV: Cologne, Germany, 2006. Available online: https://www.fgsv-verlag.de/pub/media/pdf/200_E_PDF.v.pdf (accessed on 14 August 2021). (In German)
46. FGSV. *Recommendations for Pedestrian Traffic Facilities*; FGSV: Cologne, Germany, 2002. Available online: https://www.fgsv-verlag.de/efa (accessed on 14 August 2021). (In Romanian)
47. Municipal Chamber of Lisbon. Lisbon Street Design. *Manual of Public Space*. 2018. Available online: http://www.cm-lisboa.pt/viver/urbanismo/espaco-publico (accessed on 26 June 2021). (In Portuguese).
48. Department for Transport. Inclusive Mobility. 2005. Available online: https://www.gov.uk/government/publications/inclusive-mobility (accessed on 14 August 2021).
49. Department for Transport. *Manual for Streets*; Department for Transport: London, UK, 2007. Available online: https://www.gov.uk/government/publications/manual-for-streets (accessed on 14 August 2021).
50. Transport for London. Streetscape Guidance. 1st Revision. 2019. Available online: https://tfl.gov.uk/corporate/publications-and-reports/streets-toolkit (accessed on 14 August 2021).
51. Municipality of Madrid. Accessibility Manual for Urbanised Public Spaces of the Madrid City Council. Available online: https://diario.madrid.es/wp-content/uploads/2017/02/Manual-accesibilidad-para-espacios-públicos-urbanizados-2016-1.pdf (accessed on 14 August 2021). (In Spanish)
52. City of Malmö—Streets and Parks Department. Street Sections: Recommendations and Examples for Street Environment Design. 2006. Available online: http://www.projektering.nu/files/Gatusektioner.pdf (accessed on 14 August 2021). (In Swedish)
53. Trafikverket; Swedish Association of Local Authorities and Regions. *Road and Street Design: Terms and Basic Values*; Trafikverket: Stockholm, Sweden, 2015. (In Swedish)
54. National Association of City Transportation Officials (NACTO). Urban Street Design Guide. 2013. Available online: https://nacto.org/publication/urban-street-design-guide/ (accessed on 14 August 2021).
55. Sanz Alduán, A. *Calm the Traffic. Steps for a New Culture of Urban Mobility*, 3rd ed.; Centro de Publicaciones: Madrid, Spain, 2008. (In Spanish)
56. Swiss Association of Road and Transportation Experts (VSS). *Provision for Pedestrians: Base Norm*; VSS: Zurich, Switzerland, 2009.
57. CROW. Walking Pays off; The Pedestrian in Policy, Design and Management. 2014. Available online: https://www.crow.nl/publicaties/lopen-loont (accessed on 14 August 2021). (In Dutch)
58. CROW. Accessibility Guideline for the Design of Walking Routes, Bus Stops, Car Parks and Travel and Route Information. 2014. Available online: https://www.crow.nl/publicaties/richtlijn-toegankelijkheid (accessed on 14 August 2021). (In Dutch)

59. Transport for London. Pedestrian Comfort Guidance for London. 2019. Available online: http://content.tfl.gov.uk/pedestrian-comfort-guidance-technical-guide.pdf (accessed on 14 August 2021).
60. City of Malmö—Streets and Parks Department. Technical Design Manual. 2019. Available online: http://projektering.nu/ (accessed on 24 June 2021). (In Swedish).
61. Sanz Alduán, A. *Pedestrian Mobility Guidelines: Walking in the City*; Colección Seinor; Garceta; Colegio de Ingenieros de Caminos, Canales y Puertos: Madrid, Spain, 2016; Volume 56. (In Spanish)
62. Swiss Association of Road and Transportation Experts (VSS). *Design of Urban Main Streets*; VSS: Zurich, Switzerland, 2017.
63. CROW. *Guidelines for Urban Roads and Streets (ASVV)*, 6th ed.; CROW: Ede, The Netherlands, 2012.
64. Municipality of Madrid. Guidelines for Public Roads. 2000. Available online: http://www.carreteros.org/normativa/travesias/pdfs/ccaa_pdf/ivp_ay_madrid.pdf (accessed on 24 June 2021).
65. Nordlund, J. *City of Malmö: Brodde Makri, Maria. Transport Infrastructure in Malmö*; Telco: Malmö, Sweden, 2019.
66. Spanish Ministry of Housing. Order VIV/561/2010, of 1 February: Technical Document on Basic Conditions of Accessibility and Non-Discrimination for Access and Use of Urbanised Public Spaces. 2010. Available online: https://www.boe.es/boe/dias/2010/03/11/pdfs/BOE-A-2010-4057.pdf (accessed on 14 August 2021).
67. MAUT. Design of Public Road Facilities for Pedestrian Traffic. e-UT 03.07.23. 2009. Available online: http://ume.kozut.hu/dokumentum/56 (accessed on 26 April 2019). (In Hungarian).
68. City of Malmö—Streets and Parks Department. Accessibility Programme for Malmö: Requirements and Recommendations for Good Accessibility for People with Reduced Mobility in Terms of Construction and Conversion on Public Space. 2008. Available online: http://www.projektering.nu/files/Tillganglighetsprogr.pdf (accessed on 14 August 2021). (In Swedish).
69. Swiss Association of Road and Transportation Experts (VSS). *Crossings for Pedestrian and Cycling Traffic: Foundations*; VSS: Zurich, Switzerland, 2003.
70. Jones, P.; Boujenko, N. 'Link' and 'Place': A New Approach to Street Planning and Design. 2009. Available online: https://www.australasiantransportresearchforum.org.au/sites/default/files/2009_Jones_Boujenko.pdf (accessed on 14 August 2021).
71. National Association of City Transportation Officials (NACTO). *Urban Street Stormwater Guide*; Island Press: Washington, DC, USA, 2017.

*Article*

# Guidance and Practice in Planning Cycling Facilities in Europe—An Overview

**Bettina Schröter [1,\*], Sebastian Hantschel [1], Caroline Koszowski [1], Ralph Buehler [2], Paul Schepers [3], Johannes Weber [1], Rico Wittwer [1] and Regine Gerike [1]**

[1] Institute of Transport Planning and Road Traffic, Technische Universität Dresden, 01069 Dresden, Germany; sebastian.hantschel@tu-dresden.de (S.H.); caroline.koszowski@tu-dresden.de (C.K.); johannes.weber1@tu-dresden.de (J.W.); rico.wittwer@tu-dresden.de (R.W.); regine.gerike@tu-dresden.de (R.G.)

[2] Virginia Tech Research Center (VTRC), Urban Affairs and Planning, Arlington, VA 22203, USA; ralphbu@vt.edu

[3] Department of Human Geography and Spatial Planning, Utrecht University, 3508 Utrecht, The Netherlands; j.p.schepers@uu.nl

\* Correspondence: bettina.schroeter@tu-dresden.de

**Abstract:** The provision of convenient, safe and seamless facilities for cyclists is one core success factor in promoting cycling as a mode of transport. Cycling infrastructures and planning philosophies differ greatly between countries, but there is no systematic overview or comparison of similarities and dissimilarities. The aim of this study is to provide an in-depth international overview of guidance material for cycling facilities in European countries and to develop recommendations for advancing provisions for cyclists. International guidance materials for cycling facilities along street sections are collated, systemised and compared. For researchers, the findings provide background information to better understand cycling behaviour and safety. For planners, the findings support their efforts to support cycling and to improve guidance materials. The results show that, in general, countries that are just beginning to promote cycling tend to offer a greater variety of cycling infrastructures in their guidance materials than more mature cycling countries. Countries differ in whether they prefer to put cyclists on the street level or on the sidewalk and whether they mix cyclists with other user groups in the same space. There was even greater variability among countries in the criteria for selecting types of cycling facilities than in the design characteristics (width, buffer zones, etc.).

**Keywords:** cycling; urban street design; cycling facilities; bike lanes

## 1. Introduction

Cycling is trending in research and in practice. The dynamically growing literature on cycling demonstrates how integral the establishment of safe and convenient cycling facilities is for increasing cycling levels [1], besides socio-demographic/-economic/-psychological variables, land-use and external factors such as climate and topography [2,3]. Cycling infrastructures need to be seamless and perceived as safe as well as provide appropriate levels of objective safety, e.g., in terms of crashes or conflicts. Literature also consistently shows that cycling is associated with various positive effects on the efficiency and environmental performance of transport systems as well as on the health and well-being of individuals [3].

Cyclist volumes are increasing in many cities and countries all over the world [4–6]. Many stakeholders agree that cycling, along with other active modes such as walking, should be regarded as a vital feature of transport systems to create attractive, comfortable, safe and healthy communities. They are working hard to promote cycling as a mode of transport and to improve cycling conditions; ambitious goals are being established in strategic urban and transport planning—for example the Sustainable Urban Mobility

Plans (SUMPs)—which target cycling either as a sole means of transport or in combination with walking and public transport. Examples for the latter are the cities of London and Vienna which aim for modal split proportions of 80 to 20 percent (walking/cycling/public transport vs. car) [7,8]. Lobby groups, such as national cycling associations or the European Cyclists' Federation (ECF) have increased their activities and influence substantially in the last decades and are today stronger in terms of membership and political influence for cycling than ever before. In summary, there is a pressure on planners to pay particular attention to cycling, both from the demand side (as a result of increasing cycling volumes) and from the policy side (resulting from the positive image of cycling).

These developing and multifaceted incentives toward an increase in the use and awareness of cycling have led to a variety in cycling facilities between countries and cities, and also to a dynamic collection of guidance material for cycling provision [9]. The aim of this study is to provide an in-depth international overview of this guidance material for cycling facilities and to develop on this basis recommendations for advancing provisions for cyclists. To our best knowledge, such a systematic overview is missing so far, only few and often non-scientific collations could be identified [10–13]. These are not very detailed and fragmented, and thus do not allow for systematic comparisons of standards for cycling facilities.

International guidance materials are therefore collated in this study, systemised and compared to each other and also with findings on infrastructure-based determinants of cycling safety from the scientific literature. The findings provide background information for researchers to better understand cycling behaviour and safety; they should support policy makers and planners in their efforts to support cycling and to advance and apply guidance material in a way that actually improves cycling conditions in terms of comfort, perceived and actual safety.

This paper focusses on the design of cycle facilities on sections in urban areas. It first describes the methodology used for researching the various materials including the development of a scheme for classifying cycling facilities in Section 2. Results are presented in Section 3 for the widths of cycling facilities and in Section 4 for the criteria used for selecting specific types of cycling facilities in the different countries. The summarised information in Section 5 and the comparison with the literature on determinants of cycling behaviour and safety in Section 6 lead to recommendations in providing for cycling in future guidelines on urban street design in Section 7. The paper ends with a summary in Section 8.

## 2. Methodology and Classification of Cycling Infrastructure

The basis of this study is a comprehensive research of guidance material on urban street design in European countries with a focus on the partner countries and cities in the MORE project (Multi-Modal Optimisation of Road-Space in Europe, https://www.roadspace.eu/ accessed on 13 July 2021). A questionnaire was sent to the MORE partner cities Budapest, Lisbon, London and Malmö and further technical partners (ECF, International Federation of Pedestrians (IFP), International Road Union (IRU), POLIS, PTV Group, International Association of Public Transport (UITP)) asking for material and information with relevance for urban street design in their city or from their specific perspective (technical partners). Partners were highly engaged in providing insights and references including their partial translation if necessary. Various feedback loops with discussions in teleconferences and personal meetings followed, gave background information and helped to better understand the material provided in local languages. As a result, the scope of this paper focuses on the MORE partner cities and countries. Material on other countries, for which there were no local partners in place, was added if information could be identified via desk research only. The guides published by the National Association of City Transportation Officials (NACTO) have been included as it is widely used [14–16].

In summary, recommendations for cycling infrastructure from Budapest, Lisbon, London, Malmö, Germany and the Netherlands are included in this study. Recommenda-

tions are valid on national level for Budapest (Hungary), Germany and the Netherlands. Information given in Lisbon, London and Malmö is valid on the municipal level.

Researched guidance material on cycling is, in most cases, more recently updated than for other street user groups (e.g., for motorised traffic or pedestrians) and it is more often in the active process of being updated (e.g., in Budapest, Germany, London, Malmö). This shows the high dynamics in cycling provision that is currently ongoing in all the researched countries across Europe. In addition, heterogeneity in types and range of application of cycling infrastructures are much greater compared to other user groups. One possible reason for this might be the relatively recent developments and changes in this area as described above. Another reason might be that cycling is (besides the other micro modes such as scooters) the only transport mode that can share the same space with other street users in the carriageway or on the sidewalks or that can be accommodated in dedicated cycling facilities, again either in the carriageway or on the sidewalks.

Terminology and also types of cycling facilities differ greatly between the researched references. To compare standards and their range of application, a consistent classification of cycling infrastructures is developed for this study as shown in Table 1.

**Table 1.** Types of cycling provision as identified in the guidance material.

| Type | Horizontal Location | Vertical Location | Dedicated to Cyclists | Segregation from Motorised Traffic | Segregation from Pedestrians | Exemplary Figure |
|---|---|---|---|---|---|---|
| **Mixed traffic** | On carriageway | Carriageway | No (shared with general traffic) | None | Physical |  |
| **Advisory cycle lane** Also called: Advisory bike lane | On carriageway | Carriageway | No (shared with general traffic) | Striped line or coloured surface | Physical |  |
| **Shared bus and cycle lane** | On carriageway | Carriageway | No (shared with buses) | Solid line | Physical |  ([16], ch. 4, p. 43) |
| **Mandatory cycle lane** Also called: Conventional bike lane Full bike lane | On carriageway | Carriageway | Yes | Solid/striped line (optionally with horizontal segregation) | Physical |  |
| **Cycle track** Also called: Protected bike lane | Off carriageway (adjacent to the carriageway) | Half sidewalk or sidewalk | Yes | Physical | Marking or physical |  |
| **Cycle path** Also called: Shared use path | Off carriageway (adjacent to the carriageway) | Sidewalk | No (shared with pedestrians) | Physical | None |  |
| **Cycle way** Also called: Bike path Bike trail Stand-alone path | Off carriageway (alignment independent from the carriageway) | - | Yes/No (optionally shared with pedestrians) | - | Optional |  |

Advisory cycle lanes are defined in the above Table 1 as one type of cycling infrastructure but, technically, they are a sub-type of mixed traffic because the advisory cycle lane is not exclusively dedicated to cyclists. It might also be used by general traffic. In contrast, mandatory cycle lanes are on-carriageway facilities and exclusively dedicated to cyclists. They can be separated by a striped or solid line or even have light segregation to motorised traffic.

The greatest variety in design is found for cycle lanes and cycle tracks. The standard design option for cycle lanes is a dedicated lane for cyclists on carriageway level. Cycle

tracks are usually on sidewalk level. Additionally, separated cycle lanes as well as stepped cycle tracks are recommended in the researched guidance material [17,18,20]. Those cycling facilities provide higher comfort and safety for cyclists compared to mandatory cycle lanes. Transport for London [20] recommends cycle lanes with either light or full segregation from motorised traffic. Light separation is designed with discontinuous pre-formed separators such as planters or flexible posts along the cycle lane and has buffer markings in some cases. Fully separated cycle lanes have a raised curb, separating strips, islands, grass verges or lines of planting which all create a continuous physical barrier between motorised traffic and cyclists. Stepped cycle tracks are located on an intermediate level between the carriageway and the sidewalk.

Cycle paths are always on sidewalk level as they are shared with pedestrians. Cycle ways are away from motorised traffic, e.g., in parks and may be dedicated to cyclists or shared with pedestrians.

### 3. Width of Cycling Infrastructure

The different types of cycling facilities are classified along their horizontal and vertical location relative to the carriageway and to the sidewalk: The horizontal location describes whether the cycling facility is on or off the carriageway, whereas the vertical location describes whether or not there is a difference in height between the carriageway/sidewalk and the cycling facility. In addition, information is given about whether or not the cycling facility can be used by other street users and whether or not (and how) it is separated from motorised traffic and pedestrians.

The degree of separation from pedestrians on off-carriageway cycling facilities on sidewalk level can differ. There might be a marking or a physical barrier (e.g., change in pavement or greenspace) separating cyclists and pedestrians or both might use the same space.

Mixed traffic does not require any provision for cyclists, but can be complemented by sharrows. Sharrows (also called pictograms) are defined in this context as non-contiguous lane markings and aim to make clear that cyclists are allowed and welcome in the carriageway. They also give direction about where to cycle in the carriageway, support cyclists in maintaining safe distances from parked cars and discourage overtaking by cars in narrow sections. Sharrows are mainly used where space is too narrow to provide a dedicated cycling facility [17,18] and should only be used if all conditions for mixing cyclists and motorised traffic on the carriageway are guaranteed (see Section 4).

Bicycle streets are a cycle-friendly design option for mixed traffic and are frequently used in the Netherlands [19]. These are residential streets with low link function for motorised traffic but with high link function for bicycle traffic. Bicycle traffic should be dominant in bicycle streets and should have higher volumes than car traffic. Bicycle streets might be also planned if current volumes of bicycles are lower than volumes of motorised traffic but an increase is expected or should be supported by providing a high-quality facility for cyclists. Service roads are small lower speed streets parallel to main streets with high speed or volumes of motorised traffic. Cyclists and local motorised traffic share the space in the service road. Cycling is often prohibited in the main street in these cases. Service roads often come in combination with sharrows or the dedication as a bicycle street.

Shared bus and cycle lanes where buses and cyclists are allowed to use the same lane and service roads are other design options in the category of mixed traffic. Recommendations on the width of specific types of cycling infrastructure are based on assumptions or measurements for the space needed by individual cyclists, the number of cyclists that is supposed to use the infrastructure, the allowed movements (passing, meeting) and the adjacent infrastructures. The basis for determining the space for cycling facilities is in most cases the definition of the space requirements of a standard cyclist in combination with buffer zones. All these aspects were therefore included in the analysis of the guidance material. Table 2 presents the space requirement of a standard cyclist and the recommended

widths of buffer zones in the researched guidance material. Table 3 gives an overview of the recommended widths of cycle facilities.

The space requirement of a standard cyclist is either 0.75 m (Lisbon, Malmö, the Netherlands) or 1.00 m (Budapest, London, Germany). The 1.00 m-value appears to already include a certain buffer zone, while the 0.75 m does not. For example, in Germany and Budapest Table 3 shows that there is no buffer zone needed between two cyclists in contrast to most other countries.

In addition to the various possible types and locations of cycling infrastructures as introduced in Section 2, there is also a wide variety of possible adjacent users and usages. Providing sufficient space between cyclists and these adjacent usages is of highest relevance for both the objective and perceived safety of cyclists. These buffer zones between two cyclists, or cyclists and other users describe the required space for safe overtaking or passing events. The recommended widths of the different buffer zones in the researched guidance material are presented in Table 2.

Buffer zones between two cyclists range from 0.00 m to 0.50 m and are, together with the cyclist' space requirements, highest in London with 2.50 m for two cyclists and a buffer zone in between.

Buffer zones for the general traffic are given as approximate values which are to be applied in all cases or are dependent on speed. These buffer zones vary between 0.00 m for on-carriageway cycling facilities in Germany and 2.50 m for streets with speed limits above 50 km/h in Lisbon. Having no buffer zones particularly between cycling facilities on the carriageway and motorised traffic might lead to low distances between cyclists and the car overtaking the cyclist with negative impacts on objective and perceived safety.

Buffer zones to static obstacles describe the space required to manoeuvre along high kerbstones or other objects and are recommended in most researched guidance material; their size differs with the type and height of these obstacles. The minimum as well as the maximum value is given in Lisbon with 0.20 m to obstacles of low height and up to 1.20 m to built elements. Medium buffer zones to static elements seem to be 0.25 m to 0.50 m.

Buffer zones to parking/loading facilities are recommended in order to avoid dooring crashes with cars opening their doors while being passed by a bicycle. These vary between 0.25 m and 1.00 m with medium values of around 0.75 m which are applied most frequently.

Space requirements for the different street users taken together with the buffer zones result in the recommendations for the width of cycling facilities. In general, dedicated cycling facilities need to fit to the space requirements of minimum one cyclist and buffer zones to adjacent traffic or objects and must ensure sufficient space to allow passing events (one-way) or meeting events (two-way).

**Table 2.** Space requirement standard cyclist, recommended widths of buffer zones for cyclists in selected countries and cities.

| Adjacent User/Usage | Budapest | Lisbon | London | Malmö | The Netherlands | Germany | NACTO |
|---|---|---|---|---|---|---|---|
| Cyclists | 0.00 m | 0.50 m | ≥0.50 m | No recommendation | 0.25 m/0.50 m | 0.00 m | 0.76 m |
| General Traffic | No recommendation | - 2.50 m to roads with speed limits > 50 km/h <br> - 0.70 m to roads with speed limits ≤ 50 km/h | ≥0.50 m | 0.50–1.00 m | - 1.00 m to roads with speed limits of 50 km/h <br> - 0.80 m to roads with speed limits < 50 km/h | - 0.00 m for on-carriageway cycle facilities <br> - 0.50–0.75 m for cycle facilities alongside the carriageway | No recommendation |
| Pedestrians | - | - | - | 0.30–0.40 m | - | - 0.00 m for on-carriageway cycle facilities <br> - 0.25 m for cycle facilities alongside the carriageway | No recommendation |
| Obstacles | - 0.25 m to kerbs, <br> - 0.35 m to obstacles on bridges and in exceptions <br> - 0.50 m to other obstacles | - 0.20 m to obstacles <0.15 m <br> - 0.30 m to obstacles from 0.15 m to 0.90 m <br> - 0.30–0.60 m to obstacles > 0.90 m to 0.60 m to obstacles > 0.90 m <br> - 0.90–1.20 m to built elements <br> - 0.70 m to watersides | ≥0.50 m | No recommendation | - 0.25 m to obstacles < 0.05 m, e.g., kerbs <br> - 0.50 m to obstacles > 0.05 m, e.g., kerbs <br> - 0.70 m to fixed object, e.g., railings, lamp posts, traffic signs, trees <br> - 1.00 m to build elements, e.g., walls, facades <br> (Measures exclude space requirements of cyclists) | 0.25 m to, e.g., walls, trees, traffic signs | No recommendation |
| Parking/Loading | 0.80 m | 0.70 m | ≥0.50 m | 0.80–1.00 m (where stopping is allowed) | 0.50 m | - 0.25–0.75 m to longitudinal parking for on-carriageway cycle facilities <br> - 0.75 m in any other case | No recommendation |
| References | [17] | [18] | [20] | [21] | [19] | [22,23] | [15] |

**Table 3.** Recommended widths of cycle facilities in selected countries and cities.

| | Budapest | Lisbon | London | Malmö | The Netherlands | Germany | NACTO |
|---|---|---|---|---|---|---|---|
| **Mixed Traffic** | | | | | | | |
| Width: Narrow profile | 3.50 m lane width | 3.80 m lane width | ≤3.20 m lane width | 5.50 m carriageway width | 4.80 m carriageway width | ≤6.00 m carriageway width | - |
| Width: Wide profile | 4.50 m lane width | 4.50 m lane width | ≥4.00 m lane width | | 5.80 m carriageway width | ≥7.00 m carriageway width | - |
| Operational criteria for narrow profile | Only general criteria mixed traffic | Height of adjacent buildings: <5 floors | Volume of motorised traffic: ≤500/h and 85th percentile speed ≤40 km/h and proportion of HGVs ≤5% | Standard solution in residential streets | Volume of cyclists ≤100/h or Volume of cyclists ≤400/h and Volume of motorised traffic ≤200/h | Volume of motorised traffic ≤500/h | Not recommended |
| **Advisory Cycle Lane** | | | | | | | |
| Width | 1.25 m | - | ≥2.00 m | - | - | 1.50 m | - |
| Remaining carriageway | Min. 3.5 m | - | - | | | Min. 4.5 m | - |
| **Bus/Cycle Lane** | | | | | | | |
| Narrow profile | 4.25 m | 3.20–3.25 m | 3.00–3.20 m | - | - | 3.00–3.50 m | - |
| Wide profile | | | ≥4.50 m | | | ≥4.75 m | - |
| Operational criteria for narrow profile | Only general criteria mixed traffic | - | Volume of buses ≤20/h or Volume of buses and taxis ≤100/h | Not recommended | | Volume of cyclists ≤200/h | Not recommended |
| **Mandatory Cycle Lane** | | | | | | | |
| Width | 1.25 m | ≥1.50 m | ≥2.00 m | - | 2.00–2.25 m | ≥1.85 m | 1.83 m |
| **Cycle Track** | | | | | | | |
| Width: One-Way | 2.00 m | ≥1.50 m | ≥1.50 m | ≥1.50 m | 2.00–4.00 m | 2.00 m | 1.99 m |
| Width: Two-Way | 2.50 m | ≥2.60 m | ≥2.0 m | 2.50–3.50 m | 2.50–4.50 m | 2.50 m | 3.66 m |
| **Cycle path (shared with pedestrians)** | | | | | | | |
| Width | 3.50–4.25 m | 2.70–3.00 m | 2.20–4.50 m | - | - | >2.50 m | - |
| **Cycle way (dedicated to cyclists)** | | | | | | | |
| Width | - | ≥2.60 m | 1.20–3.50 m | - | - | 2.50–4.00 m | - |
| **Reference** | [17] | [18] | [16,20] | [21] | [19] | [22,23] | [15] |

Carriageway widths for bicycles in mixed traffic should be kept either low so as to cause cars to remain behind a bicycle when faced with oncoming traffic or kept wide so that cars can safely overtake cyclists even in the face of oncoming traffic. Intermediate carriageway widths that might lead to situations of doubt for car drivers on whether or not to overtake a bicycle should be avoided. This principle of either narrow or wide lanes (technically called profiles) for mixed traffic is recommended in references from London, the Netherlands and Germany. Transport for London [20] recommends avoiding carriageway width of 6.40 m to 8.00 m (doubled lane width) and FGSV in Germany [21–23] does not recommend values of around 6.00 m to 7.00 m. The differences in width result from different space requirements of standard cyclists, motorised vehicles and buffer zones. For example, and as mentioned above, in London the buffer zone between cyclists and general traffic is 0.50 m whilst there is no buffer zone in Germany (both countries consider 1.00 m as space requirement for one cyclist). CROW [19] recommends a narrow profile of 4.80 m and a wide profile of 5.80 m, which both are narrower than the recommendations in Germany and London; Budapest only recommends the wide profile. Lisbon recommends different carriageway widths for cycling in mixed traffic depending on the height of the adjacent buildings. Narrow profiles only work with low volumes of motorised traffic; higher volumes cause irritation and might eventually result in risky overtaking manoeuvres.

A similar approach is used for shared bus and cycle lanes in London and Germany as described in Table 3. London recommends the profile dependent on the number of buses or buses plus taxis per hour. In Germany, the width of bus/cycle lanes depends on the volume of cyclists. Budapest and Lisbon recommend general width of shared bus and cycle lanes.

The variety of recommended widths for dedicated cycling facilities is quite low in the researched guidance material. Widths range from 1.25 m to 2.25 m for one-way cycling facilities and are ≥2.00 m (London) or ≥2.50 m for two-way-facilities (Budapest, Lisbon, Malmö, The Netherlands, Germany).

Advisory cycle lanes are usually narrower than mandatory cycle lanes or cycle tracks/paths, because they are used if space is too narrow to provide a dedicated cycling facility, and in addition, cyclists are allowed to leave the advisory lane and to cycle in the carriageway, e.g., when overtaking other cyclists. Recommendations are given for the remaining carriageway width between two advisory cycle lanes because enough space has to be provided for motorised traffic to pass vehicles in meeting events.

Shared paths for cyclists and pedestrians are wider than dedicated cycling facilities because they have to accommodate the two user groups with substantial differences in their velocity.

## 4. Operational Criteria for Selecting Suitable Types of Cycling Infrastructure

Similarities were identified in recommended types and widths of cycling infrastructure in Sections 2 and 3. However, the criteria for their application differ greatly; specific criteria and thresholds are provided to select the type of cycling infrastructure for each application with substantial differences particularly in the used thresholds. In what follows, the approaches of each city/country are presented individually. At the end of this Section, Table 4 gives and overview of the selection criteria for all the researched guidance material.

Malmö only verbally explains the operational criteria for selecting suitable types of cycling infrastructure and gives the general recommendation for main streets to provide separated cycling infrastructure (cycle tracks/paths; usually two-way). Outside the main street network, cyclists cycle in mixed traffic. This distinction of main/lower level streets mainly refers to volumes of motorised traffic (max. 3000 vehicles/24 h), the speed of motorised traffic is not considered because speed limits in Malmö are generally low (max. 40 km/h).

**Table 4.** Speed and volume criteria for selecting cycle facilities in selected countries and cities.

| Cycling Infrastructure | | Budapest | Lisbon | London | Malmö | The Netherlands | | Germany |
|---|---|---|---|---|---|---|---|---|
| **Criteria** | Speed | Volume of motorised traffic (veh/24 h) | Volume of motorised traffic (veh/24 h) | Volume of motorised traffic (veh/24 h) | Volume of motorised traffic (veh/24 h) | Volume of motorised traffic (veh/24 h) | Volume of cyclists (cyc/24 h) | Volume of motorised traffic (veh/24 h) |
| **Mixed Traffic** | 20 | <5000 | - | | | <5000 | <2500 | <8000 |
| | 30 | <12,000 * | <3000 | <5000 | <3000 | | | |
| | 40 | | | | | | | <6000 |
| | 50 | <3000 | | | | | | <4000 |
| | 60 | - | - | - | - | - | - | <2000 |
| | 70 | | | | | | | <1000 |
| **Advisory Cycle Lane** | 30 | 5000–12,000 | | | | | | <18,000 |
| | 40 | 2000–4000 | | | | | | <16,000 |
| | 50 | | - | - | - | - | - | <11,000 |
| | 60 | - | | | | | | <5000 |
| | 70 | | | | | | | <3000 |
| **Mandatory Cycle Lane** | 30 | - | | | | 2000–5000  >4000 | >2000  in any case | >18,000 |
| | 40 | 5000–19,000 | - | | | | | >16,000 |
| | 50 | 5000–15,000 | 3000–8000 | - | - | | | >11,000 |
| | 60 | 2000–8000 | - | | | - | - | >5000 |
| | 70 | - | | | | | | >3000 |
| **Segregated Lane/Stepped track** | 40 | >19,000 | - | | | | | |
| | 50 | >15,000 | 3000–10,000 | | - | - | - | - |
| | 60 | 8000–15,000 | - | Where mixed traffic cannot be applied | | | | |
| | 70 | <12,000 | | | | | | |
| **Cycle Track/Path** | 30 | | - | | | 2000–5000  >4000 | >2000  in any case | >18,000 |
| | 40 | - | | | - | - | - | >16,000 |
| | 50 | | 3000–8000 | | | | | >11,000 |
| | 60 | >15,000 | >10,000 | | in any case | Always required with speed limits ≥50 km/h | | >5000 |
| | 70 | in any case | >10,000 | | | | | >3000 |

* may be supplemented by bike lane or advisory bike lane.

Volume and speed of motorised traffic are used as operational criteria for selecting a suitable provision for cyclists in all cities/countries using operational criteria. For example, Hungary and Germany use these two criteria and give recommendations for mixed traffic, advisory cycle lanes, mandatory cycle lanes and cycle lanes/tracks depending on volumes and speed of motorised traffic as shown below in Figure 1; Figure 2 show the approach for selecting cycle facilities in Hungary and Germany. The green box in Figure 1 describes situations with low speeds (≤30 km/h) and low volumes (≤10,000 vehicles/day) of motorised traffic where cyclists are guided in mixed traffic. The white area (described as joint traffic zone) is a sub-type of mixed traffic and is recommended up to speeds of 50 km/h.

The blue section describes provisions for cyclists on the carriageway with speed limits between 40 km/h and 60 km/h. Mandatory cycle lanes are recommended with volumes

<15,000 vehicles/day and speed limits of 50 km/h. With higher volumes, a vertical or horizontal segregation is recommended (protected or raised cycle lane).
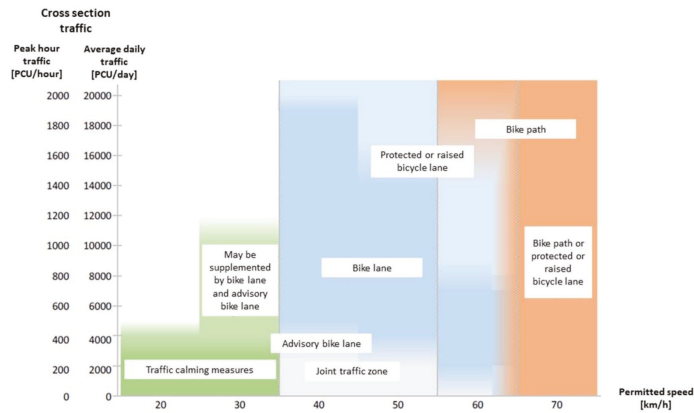


**Figure 1.** Selection plan for cycle facilities in Budapest ([17], p. 14).
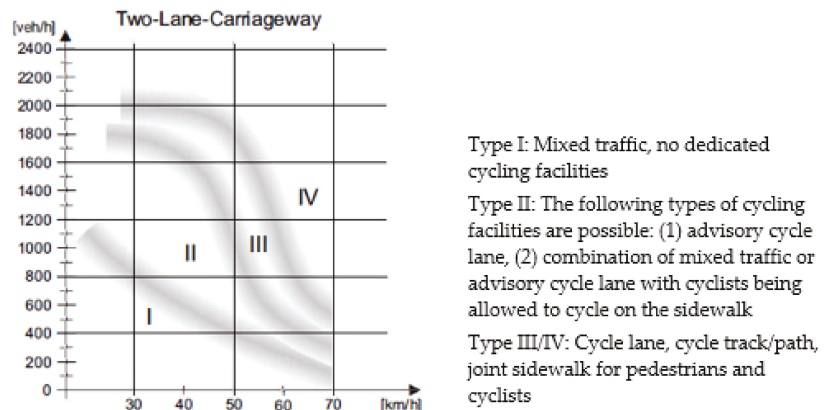


**Figure 2.** Selection plan for bicycle facilities in Germany ([23], p. 1.9).

Separation (orange section) is clearly required with speed limits of 70 km/h. The standard solution for separation is bike paths—protected or raised cycle lanes are possible with low or medium volumes and high speeds.

The German guidelines use a similar approach (see Figure 2). Mixed traffic (denoted as Area I) is possible with a speed limit of 30 km/h and with maximum 8000 vehicles/24 h or with a speed limit of 50 km/h and with maximum 4000 vehicles/24 h without any additional measures.

Area II denotes the combinations of traffic volumes and speed that are suitable for advisory lanes or solutions where cyclists are allowed to cycle either on the carriageway or the sidewalk. These still shared cycling facilities are recommended up to 18,000 vehicles/24 h combined with speed limits of 30 km/h or 8000 vehicles/24 h combined with speed limits of 50 km/h. The allowed volumes of motorised traffic at specific speed limits are almost twice as high as the allowed volumes in Hungary.

A clear recommendation to physically separate cyclists from motorised traffic does not exist in Germany as both, cycle lanes and cycle tracks/paths, are recommended in Area III and Area IV.

The Lisbon guidelines also consider volumes and speed limits of motorised traffic but add the street category as third criterion as shown in Figure 3. Lisbon gives the strictest recommendation on guiding cyclists in mixed traffic with maximum speed limits of 30 km/h and volumes up to 3000 vehicles/24 h on local streets. Alternatively, contra-flow cycle lanes in one-way streets may be implemented. Mandatory cycle lanes are recommended on local streets with 50 km/h speed limit. If the volumes exceed 5000 vehicles/24 h, cycle tracks are required.

| Street Hierarchy | Motorised traffic volume [veh/24 h] | Speed limit [km/h] | Cycle facility |
| --- | --- | --- | --- |
| **Local street** | <1500 | ≤30 km/h | Mixed traffic |
| | 1500 ≤ v <3000 | ≤30 km/h | Contra-flow cycle lane<br>Mixed traffic |
| | 3000 ≤ v <8000 | 50 km/h | Mandatory cycle lane<br>Cycle track (v > 5000) (unidirectional or bidirectional) |
| **Distributional street** | 3000 ≤ v <8000 | 50 km/h | Mandatory cycle lane (elevated) |
| | 8000 ≤ v <10,000 | 50 km/h | Mandatory cycle lane (elevated or intermediate level) |
| | > 10,000 | 50 km/h | Cycle track (unidirectional or bidirectional) |
| **Higher level Street** | > 10,000 | ≥50 km/h | Cycle track (unidirectional or bidirectional) |

**Figure 3.** Selection plan for bicycle facilities in Lisbon ([18], p. 12.5).

On distributional streets, elevated lanes (optionally on intermediate level between carriageway and sidewalk) are recommended with speed limits of 50 km/h and volumes up to 10,000 vehicles/24 h. Higher volumes require the implementation of cycle tracks.

At higher level streets (with speed limits ≥50 km/h), cycle tracks are generally recommended.

CROW [19] recommends cycling facilities according to the volume and speed of motorised traffic, the road category and the cycle network category which represents the volumes of cyclists (see Figure 4). Recommendations for mixed traffic in the Netherlands are with maximum 30 km/h and 5000 vehicles/24 h almost as strict as in Lisbon. With high volumes of cyclists (>2000 cyclists/24 h) and low volumes of motorised traffic, a cycle street is preferred over standard mixed traffic solutions to emphasise the dominance of cyclists. Cycle paths as separated facilities are recommended for speed limits of 50 km/h onwards (independent of volumes of motorised traffic) or with lower speed limits and high volumes of motorised traffic or high volumes of cyclists.



**Figure 4.** Selection plan for bicycle facilities in built-up areas in the Netherlands ([19], p. 102).

Transport for London (TfL) developed recently a new approach to decide whether or not cyclists can be mixed with motorised traffic [24]. TfL defines target green levels and minimal requirements for six criteria and a scheme to decide which level has to be met for each of these criteria in different combinations, these are summarised in 4 scenarios.

Target green levels are defined as:

1. Volume of motorised traffic: 5000 veh/24 h
2. 85th percentile speed of motorised traffic: 25 mph (40 km/h)
3. Width of traffic lanes: $\leq$3.20 m with 85th percentile speeds $\leq$ 40 km/h and the proportion of Heavy Goods Vehicles (HGV) $\leq$ 5%; $\geq$4.50 m otherwise
4. Minimisation of collision risk at junctions: infrastructure measures at all priority junctions' volume of motorised traffic $\geq$ 2000 veh/24 h
5. Minimisation of the impact of kerbside activities and
6. Proportion of HGV ($\leq$5% with volume of motorised traffic $\leq$ 5000 veh/24 h; $\leq$10% with volume of motorised traffic $\leq$ 2000 veh/24 h)

The scheme in Figure 5 illustrates scenarios that are suitable for guiding cyclists in mixed traffic. Scenario 1 is the preferred one when all target green levels are met. Higher volumes of motorised traffic or higher speed can both be compensated by defined combinations of criteria for which the target green levels have to be met as the minimum. In Scenario 2, too high volumes of motorised traffic are compensated by sufficiently low speed and proportions of HGV combined with at least two out of three of the remaining criteria meeting target green levels. Scenario 3 describes how too high speed can be compensated. When volumes and speed of motorised traffic meet target green levels, two out of four of the other criteria have to meet the target green level. Safety at junction has to be ensured in in all cases (turning risk): Measures for mitigating turning risks are required if safety issues exist.

| Scenarios considered acceptable for people cycling to mix with general traffic | Criteria 1 Flows | Criteria 2 Speed | Criteria 3 Width | Criteria 4 Turning risk | Criteria 5 Kerbside activity | Criteria 6 HGVs |
|---|---|---|---|---|---|---|
| Scenario 1 | All target green levels met | | | | | |
| Scenario 2 | Falls below the target green level | Target green level met | At least 2 out of 3 criteria achieve the target green level of provision, with turning risk mitigation measures at junctions required where there is a known safety issue | | | Proportion of HGVs* is less than 5%**(except where width requirements are met) |
| Scenario 3 | Target green level met | Falls below the target green level | At least 3 out of 4 criteria achieve the target green level of provision, with turning risk mitigation measures at junctions required where there is a known safety issue | | | |
| Scenario 4 | Target green level met | Target green level met | At least 2 out of 4 criteria achieve the target green level of provision, with turning risk mitigation measures at junctions required where there is a known safety issue | | | |

\* Heavy Goods Vehicle (HGV) – defined as lorries and trucks over 3.5 tonnes

\*\* Based on the peak hour HGV % as a proportion of the corresponding motor vehicle traffic flow, 7am to 7pm

**Figure 5.** Scenarios for guiding cyclists in mixed traffic in London ([16], p. 6).

Table 4 summarises the identified operational criteria and target values for deciding on suitable types of cycling facilities in the form of a table the crosses the two criteria that are consistently applied in all references, this is volumes and speed of motorised traffic. Volume of cyclists is also included in the table as this is a speed limit concerned criterion (even though is just applies in the Netherlands). Further criteria are only used in some of the references and are not included in the table; these are explained in the descriptions and figures above. NACTO [15] does only give one operational criterion on the suitability of mixed traffic solutions, this is maximum 3000 vehicles/24 h and 30 km/h and is therefore not included in the table.

## 5. Summary of Practices in Providing for Cyclists

The researched material shows a high variety in cycling facilities and criteria for their operation. Infrastructure for cyclists ranges from integrated solutions with cyclists in mixed traffic or on cycle streets to fully separated cycling facilities off the carriageway, e.g., on

cycle paths. Overall, more than 10 variations of infrastructure for cyclists were identified within the seven countries/cities.

As mixed traffic and cycle tracks/paths are defined in all cities and mandatory cycle lanes are recommended in all guidance material besides Malmö, these three provisions can be summarised as standard cycling infrastructure types within the researched countries and cities. Within these types the application of mixed traffic is typically used for streets with low volumes and speed of motorised traffic, cycle lanes and tracks/paths are more likely on streets with higher volumes and speed of motorised traffic.

Recommendations for accommodating cyclists in mixed traffic together with motorised vehicles in the carriageway are the most stringent in the Netherlands and Lisbon who only allow mixed traffic with speed limits below 30 km/h [18,19]. In Budapest and Germany, mixed traffic (including advisory cycle lanes) is possible up to speed limits of 50 km/h [17] or 70 km/h [23].

Bicycle streets are a special type of mixed traffic and recommended in the Netherlands for streets with high cycling volumes to emphasise the dominance of cyclists [19]. Carriageways with cyclists riding in mixed traffic should either be narrow so as to force cars to remain behind a bicycle when faced with oncoming traffic, or kept wide so that cars can safely overtake cyclists even in the face of oncoming traffic.

Dedicated cycling facilities are recommended for high volumes of motorised traffic and high speed limits. Cycle tracks/paths and cycle lanes bring vertical and/or horizontal separation. Budapest and Lisbon recommend mandatory cycle lanes with intermediate volumes/speed of motorised traffic and cycle tracks/paths with higher volumes/speed of motorised traffic [17,18]. The Netherlands and Malmö generally recommend off-carriageway cycle tracks/paths for main streets [19,21]. The German recommendations equally recommend cycle lanes and cycle tracks/paths for streets with high volumes of motorised traffic and speed limits [23].

The recommended widths for dedicated cycling facilities range from 1.25 m to 2.25 m for one-way cycling facilities and are $\geq 2$ m for two-way-facilities. It tends to be higher in countries with a well-established cycling culture such as the Netherlands and Sweden compared to starter countries such as Lisbon/Portugal.

## 6. Comparison of Guidance on Cycling Facilities with Literature on Determinants of Cyclist Safety and Comfort

The assessment of this diversity in recommendations for types of cycling infrastructures and criteria for their selection against the relevant criteria of (objective and perceived) safety and comfort proves to be difficult for at least two reasons: (1) The cause-effect chain from the characteristics of cycling facilities in individual street sections to travel behaviour and cycling choices is complex. Decisions about travel behaviour are shaped by various influences and the characteristics of single street sections is only one of them (see, e.g., [24] for the relevance of seamless cycling networks and safe intersections for successfully providing for cyclists). (2) The literature on the influence of the type of cycling infrastructure at street sections on safety is fragmented and hardly allows to draw general conclusions.

Mueller et al. [1] demonstrate the relationship between the length of cycling facilities in a city and the modal share of cycling without any consideration of the type of cycling facilities. Le et al. [25] report similar findings and Buehler and Pucher [26] find comparably high influences of cycle lanes and cycle paths on bicycle commuting in American cities. The TEMS tool (http://tems.epomm.eu/ accessed on 13 July 2021) shows that cycling shares are highest in Dutch and Swedish cities—two countries that favour separating cyclists from motorised traffic. At the same time, the TEMS tool shows substantial differences in modal shares of cycling within countries even when these have guidelines that are valid at national level such as Germany or Hungary.

Literature on infrastructural determinants of cycling safety consistently shows that higher speed (allowed or driven) leads to higher severity of crashes and in some studies also to higher crash numbers [27–30] and presence of parking increases crash risk [30–32].

Crash numbers are higher for streets with cycling facilities [33]. The main reason for this is higher car and cycle traffic volumes and speed in such streets. Differences in crash risks per car volumes are less consistent in the researched references. Lusk et al. ([34] see also [35]) find higher risks for on-carriageway cycling facilities compared to cycle tracks. Teschke et al. [31] find higher risks for sections without any cycling facility compared to sections with cycle lanes. Harris et al. [29] find lower risks for sections with cycle tracks compared to advisory lanes and mixed traffic. Canadian studies find higher risks for cycle tracks compared to cycle lanes [36,37].

The literature review on perceived safety supplements the findings of the risk analysis. User surveys show that perceived safety is very low when cyclists are guided in mixed traffic [38–41]. The safety perception is even lower with high speed limits [38,41], the presence of parking [38–40], high volumes of motorists or a high proportion of heavy vehicles or presence of buses [38,41].

On streets with cycling infrastructure, users generally prefer facilities off-carriageway over on carriageway facilities [38,39,41]. However, well-designed and protected cycle lanes (with coloured surface, buffer elements and sufficient width) can achieve a similarly high level of perceived safety as cycle tracks [38]. In a mental mapping study in Ireland, separation of cyclists was found to have the greatest impact on perceived safety (compared to motorists' volumes, width of infrastructure, number of junctions and parking) [42].

Some general conclusions can be drawn besides the fragmented character of the literature. Slow speed increases safety and also the willingness of people to cycle in the streets. Study design, specific location and infrastructure design matter when comparing the safety of on- and off-carriageway facilities. Cycling facilities are safer and perceived safer than no cycling facilities and there is the tendency of better safety for separated cycling facilities compared to unprotected on-carriageway facilities such as cycle lanes. The number of cars and also cyclists consistently matters [43]. Higher car volumes increase crash risk for cyclists, the safety-in-numbers effect leads to relatively lower risks for cyclists with increasing cycling volumes.

## 7. Recommendations on Providing for Cyclists in Future Guidelines on Urban Street Design

Based on the insights gained from summarising the various guidance materials on cycling provision, the following recommendations were developed:

Keep it simple: "Starter countries" in terms of cycling tend to offer many more types of cycling facilities in their guidance materials than countries with a longer history in cycling provision. A variety of solutions might be necessary in starter countries because the optimal solutions might not have enough political support (e.g., would require taking too much space from cars). This is a critical point because (potential) cyclists are not familiar with participating in traffic as cyclists nor are car drivers and other street users used to cycling infrastructure or expect cyclists in the streets. With this in mind, the first recommendation is to keep cycling provision simple, wherever possible. The three basic options for accommodating cyclists in the streets are a solid basis and, in most cases, sufficient; these are (1) mixed traffic, (2) on-carriageway mandatory cycle lanes and (3) off-carriageway cycle tracks/paths. Too many types of cycling infrastructure might cause confusion for users. Even though there are many different types of cycling infrastructure available, this disadvantage might outweigh the advantage of having the opportunity to provide tailor-made solutions for each design task.

Mixed traffic or dedicated cycling facilities: The decision between accommodating cyclists in mixed traffic with motorised vehicles on the one hand and dedicated cycling facilities on the other is of special importance. Speed of motorised cars of maximum 30 km/h and low volumes of motorised vehicles appear to be the two key deciding factors. Dedicated cycling facilities should be provided if either of these two is exceeded. Bicycle volumes should also be considered if these reach relevant levels. Lane widths (profiles) for cycling in mixed traffic should be either narrow or wide in order to clearly indicate whether or not the overtaking of bicycles is safely possible for cars. Narrow lane widths

seem to be more suitable as these support low speed for all street users. Bicycles should be prioritised over motorised traffic, particularly if their current or expected number exceeds car volumes, e.g., by providing bicycle streets.

Dedicated cycling facilities on or off the carriageway: Once the decision for a dedicated cycling facility has been made, these might be placed on the carriageway as cycle lanes or off the carriageway as cycle tracks/paths. Both of these options have pros and cons which can be evaluated on a case-by-case basis or addressed in a general manner as is carried out in Malmö and the Netherlands for off-carriageway cycle tracks/paths. Both options are good choices for objectively safe and convenient cycling networks if these are sufficiently wide and well designed. Cyclists feel safer on off-carriageway or at least protected facilities and although the scientific literature does not indicate a clear risk reduction for cycle tracks, elevated or protected design solutions are recommended when traffic volumes or speed limits are high. Solutions for street sections always need to fit with the solutions at the adjacent junctions as these are very important for cyclists' safety and comfort.

Mixing pedestrians and cyclists: This is a popular solution for limited space and high volumes (and speed) of motorised traffic but might lead to conflicts between pedestrian movement and place activities and cyclists. Dedicated and separated facilities for cyclists and pedestrians should therefore be implemented whenever possible, even if that requires taking space from motorised traffic.

Width of cycle lanes and tracks/paths: With high cycle volumes, it is desirable to offer a width of minimum 2.00 m in one direction to allow passing events without leaving the cycle lane/track. Narrower facilities should only be provided where a low number of cyclists is expected (e.g., due to alternative attractive routes in the network). Wide facilities might demand physical separation to discourage other road users from driving or parking within the cycle infrastructure. Buffer zones to adjacent usages particularly for parking are paramount for safe cycling.

Future needs: In general, cycling infrastructure should cover current and future needs. Due to an increasing number of cargo bicycles (higher space requirements) and electric bicycles (higher speeds) and the fast developments in Personal Light Electric Vehicles (PLEV), cycling infrastructure should provide enough space for non-standard and standard users. One such example would be the provision of lane widths which make is easy for faster cyclists to pass slower cyclists even though the slower bike has extended dimensions. Cycle Highways are another example as a new trend in cycling practice in planning [44].

## 8. Summary and Conclusions

This study provides a comprehensive international overview of guidance material for cycling facilities. It shows similarities and differences in the practice of cycling provision. Some general trends could be identified. "Cycling countries" such as the Netherlands and Malmö/Sweden use fewer types of cycling facilities, are strict in mixing cyclists and motorised traffic only in streets with low speed and volumes of motorised traffic and recommend greater widths for dedicated cycling facilities. At the same time, also substantial differences emerged. For example, Germany treats mandatory cycle lanes in the carriageway equally in terms of operational criteria as separated cycle tracks/paths. Malmö uses mainly two-way cycle tracks/paths on their main roads, these cycling facilities should be avoided following the recommendations in German guidelines for cycling facilities. More empirical evidence on the effects of the different types of cycling facilities would help to advance guidance material towards safe and comfortable solutions in each specific case study. These investigations should include both street sections and junctions as the latter are even more decisive for cyclists' safety compared to street sections.

# References

1. Mueller, N.; Rojas-Rueda, D.; Salmon, M.; Martinez, D.; Ambros, A.; Brand, C.; de Nazelle, A.; Dons, E.; Gaupp-Berghausen, M.; Gerike, R.; et al. Health impact assessment of cycling network expansions in European cities. *Prev. Med.* **2018**, *109*, 62–70. [CrossRef]
2. Gerike, R.; de Nazelle, A.; Wittwer, R.; Parkin, J. Special Issue "Walking and Cycling for better Transport, Health and the Environment". *Transp. Res. Part A* **2019**, *123*, 1–6. [CrossRef]
3. Gerike, R.; Parkin, J. *Cycling Futures: From Research into Practice*; Routledge: London, UK, 2016.
4. Schepers, P.; Helbich, M.; Hagenzieker, M.; de Geus, B.; Dozza, M.; Agerholm, N.; Niska, A.; Airaksinen, N.; Papon, F.; Gerike, R.; et al. The development of cycling in European countries since 1990. *Eur. J. Transp. Infrastruct. Res.* **2021**, *21*. [CrossRef]
5. Buehler, R.; Pucher, J. (Eds.) *Cycling for Sustainable Cities*; The MIT-Press: Cambridge, MA, USA, 2021.
6. Buehler, R.; Pucher, J. COVID-19 Impacts on Cycling, 2019–2020. *Transp. Rev.* **2021**, *41*, 393–400. [CrossRef]
7. Mayor of London. Mayor's Transport Strategy. 2018. Available online: https://tfl.gov.uk/corporate/about-tfl/the-mayors-transport-strategy (accessed on 24 June 2021).
8. Vienna Municipality. Urban Mobility Plan Vienna. 2015. Available online: https://www.wien.gv.at/stadtentwicklung/studien/pdf/b008379d.pdf (accessed on 24 June 2021).
9. Gerike, R.; Dean, M.; Koszowski, C.; Schröter, B.; Wittwer, R.; Weber, J.; Jones, P. Urban Corridor Road Design: Guides, Objectives and Performance Indicators. Deliverable D1.2, Multimodal Optimisation of Roadspace in Europe. 2019. Available online: https://www.roadspace.eu/ (accessed on 24 June 2021).
10. Report for Transport for London. International Cycling Infrastructure Best Practice Study. 2014. Available online: http://content.tfl.gov.uk/international-cycling-infrastructure-best-practice-study.pdf (accessed on 24 June 2019).
11. FHWA. Bicycle Network Planning & Facility Design Approaches in The Netherlands and the United States: FHWA-PL-16-019. 2016. Available online: https://www.fhwa.dot.gov/environment/bicycle_pedestrian/publications/network_planning_design/network_planning_design.pdf (accessed on 24 June 2021).
12. Furth, P.G. Bicycle Route Facilities: An International Comparison. In *City Cycling*; Pucher, J.R., Buehler, R., Eds.; Urban and Industrial Environments; MIT Press: Cambridge, MA, USA, 2012; Chapter 6.
13. Furth, P.G. Bicycling infrastructure for all. In *Cycling for Sustainable Cities*; Buehler, R., Pucher, J., Eds.; The MIT-Press: Cambridge, MA, USA, 2021; Chapter 5.
14. National Association of City Transportation Officials (NACTO). Urban Street Design Guide. 2013. Available online: https://nacto.org/publication/urban-street-design-guide/ (accessed on 24 June 2021).
15. National Association of City Transportation Officials (NACTO). Urban Bikeway Design Guide. 2014. Available online: https://nacto.org/publication/urban-bikeway-design-guide/ (accessed on 24 June 2021).
16. Transport for London. New Cycle Route Quality Criteria: Accompanying Technical Note to the Quality Criteria Spreadsheet Tool v1. 2019. Available online: http://content.tfl.gov.uk/cycle-route-quality-criteria-technical-note-v1.pdf (accessed on 24 June 2021).
17. MAUT. Design of Bikeable Public Roads. e-UT_03_04_13. 2019. Available online: https://ume.kozut.hu/dokumentum/479 (accessed on 24 June 2021).

18. Municipal Chamber of Lisbon. Lisbon Street Design. Manual of Public Space. 2018. Available online: http://www.cm-lisboa.pt/viver/urbanismo/espaco-publico (accessed on 24 June 2021).
19. CROW. *Design Manual for Bicycle Traffic*; CROW-Record; VROW: Ede, The Netherlands, 2016; Volume 28.
20. Transport for London. London Cycling Design Standards. 2016. Available online: https://tfl.gov.uk/corporate/publications-and-reports/streets-toolkit (accessed on 24 June 2021).
21. City of Malmö: Brodde Makri, Maria; Nordlund, Jesper. *Transport Infrastructure in Malmö*; Telco: Malmö, Sweden, 2019.
22. FGSV. *Directives for the Design of Urban Roads: Translation 2012*; FGSV-Verlag: Cologne, Germany, 2006.
23. FGSV. *Recommendations for Cycle Facility Design*; FGSV-Verlag: Cologne, Germany, 2010.
24. Hull, A.; O'Holleran, C. Bicycle infrastructure: Can good design encourage cycling? *Urban Plan. Transp. Res.* **2014**, *2*, 369–406. [CrossRef]
25. Le, H.T.K.; Buehler, R.; Hankey, S. Correlates of the Built Environment and Active Travel: Evidence from 20 US Metropolitan Areas. *Environ. Health Perspect.* **2018**, *126*, 77011. [CrossRef] [PubMed]
26. Buehler, R.; Pucher, J. Cycling to work in 90 large American cities: New evidence on the role of bike paths and lanes. *Transportation* **2012**, *39*, 409–432. [CrossRef]
27. Chen, P.; Shen, Q. Built environment effects on cyclist injury severity in automobile-involved bicycle crashes. *Accid. Anal. Prev.* **2016**, *86*, 239–246. [CrossRef]
28. Collins, D.J.; Graham, D.J. Use of Open Data to Assess Cyclist Safety in London. *Transp. Res. Rec.* **2019**, *2673*, 27–35. [CrossRef]
29. Harris, M.A.; Reynolds, C.C.O.; Winters, M.; Cripton, P.A.; Shen, H.; Chipman, M.L.; Cusimano, M.D.; Babul, S.; Brubacher, J.R.; Friedman, S.M.; et al. Comparing the effects of infrastructure on bicycling injury at intersections and non-intersections using a case-crossover design. *Injury Prev.* **2013**, *19*, 303–310. [CrossRef]
30. Eluru, N.; Bhat, C.R.; Hensher, D.A. A mixed generalized ordered response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accid. Anal. Prev.* **2008**, *40*, 1033–1054. [CrossRef] [PubMed]
31. Teschke, K.; Harris, M.A.; Reynolds, C.C.O.; Winters, M.; Babul, S.; Chipman, M.; Cusimano, M.D.; Brubacher, J.R.; Hunte, G.; Friedman, S.M.; et al. Route Infrastructure and the Risk of Injuries to Bicyclists: A Case-Crossover Study. *Am. J. Public Health* **2012**, *102*, 2336–2343. [CrossRef]
32. Schüller, H.; Plesker, M.; Hantschel, S.; Kühn, B.; Gerike, R. *Akzeptanz und Verkehrssicherheit des Radverkehrs im Mischverkehr Auf Hauptverkehrsstraßen: Entwurf Schlussbericht*; 2019; unpublished.
33. Alrutz, D.; Bohle, W.; Maier, R.; Enke, M.; Pohle, M.; Zimmermann, F.; Ortlepp, J.; Schreiber, M. *Einfluss von Radverkehrsaufkommen und Radverkehrsinfrastruktur auf das Unfallgeschehen*; Forschungsbericht/Gesamtverband der Deutschen Versicherungswirtschaft e.V.; Nr. 29; Gesamtverband der Deutschen Versicherungswirtschaft e. V.; Unfallforschung der Versicherer: Berlin, Germany, 2015.
34. Lusk, A.C.; Furth, P.G.; Morency, P.; Miranda-Moreno, L.F.; Willett, W.C.; Dennerlein, J.T. Risk of injury for bicycling on cycle tracks versus in the street. *Inj. Prev.* **2011**, *17*, 131–135. [CrossRef]
35. FHWA. Separated Bike Lane Planning and Design Guide. Appendix. 2013. Available online: http://www.bv.transports.gouv.qc.ca/mono/1169015/02_Appendix.pdf (accessed on 24 June 2021).
36. Aultman-Hall, L.; Hall, F.L. Ottawa-Carleton commuter cyclist on-and off-road incident rates. *Accid. Anal. Prev.* **1998**, *30*, 29–43. [CrossRef]
37. Aultman-Hall, L.; Kaltenecker, M.G. Toronto bicycle commuter safety rates. *Accid. Anal. Prev.* **1999**, *31*, 675–686. [CrossRef]
38. FixMyBerlin. Study on Subjective Safety When Travelling by Bicycle: Results and Data from a Survey of 21,000 Participants. 2020. Available online: https://fixmyberlin.de/research/subjektive-sicherheit (accessed on 27 July 2021).
39. Winters, M.; Babul, S.; Becker, H.J.E.H.; Brubacher, J.R.; Chipman, M.; Cripton, P.; Cusimano, M.D.; Friedman, S.M.; Harris, M.A.; Hunte, G.; et al. Safe Cycling: How Do Risk Perceptions Compare with Observed Risk? *Can J. Public Health* **2012**, *103*, S42–S47. [CrossRef]
40. Chataway, E.S.; Kaplan, S.; Nielsen, T.A.S.; Prato, C.G. Safety perceptions and reported behavior related to cycling in mixed traffic: A comparison between Brisbane and Copenhagen. *Transp. Res. Part F* **2014**, *23*, 32–43. [CrossRef]
41. Rossetti, T.; Guevara, C.A.; Galilea, P.; Hurtubia, R. Modeling safety as a perceptual latent variable to assess cycling infrastructure. *Transp. Res. Part A* **2018**, *111*, 252–265. [CrossRef]
42. Manton, R.; Rau, H.; Fahy, F.; Sheahan, J.; Clifford, E. Using mental mapping to unpack perceived cycling risk. *Accid. Anal. Prev.* **2016**, *88*, 138–149. [CrossRef] [PubMed]
43. Elvik, R. Exploring factors influencing the strength of the safety-in-numbers effect. *Accid. Anal. Prev.* **2017**, *100*, 75–84. [CrossRef] [PubMed]
44. Cabral Dias, G.J.; Gomes Ribeiro, P.J. Cycle Highways: A new concept of infrastructure. *Eur. Plan. Stud.* **2021**, *29*, 1003–1020. [CrossRef]

*Article*

# Measurement Quality Appraisal Instrument for Evaluation of Walkability Assessment Tools Based on Walking Needs

**Sanaz Tabatabaee [1], Mahdi Aghaabbasi [2,\*], Amir Mahdiyar [3], Rosilawati Zainol [4] and Syuhaida Ismail [1,\*]**

[1]   Green Cities and Construction Research Group, Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur 54100, Malaysia; tsanaz@utm.my

[2]   Centre for Sustainable Urban Planning and Real Estate (SUPRE), Department of Urban and Regional Planning, Faculty of Built Environment, University of Malaya, Kuala Lumpur 50603, Malaysia

[3]   School of Housing, Building and Planning, Universiti Sains Malaysia, Penang 11800, Malaysia; amirmahdiyar@usm.my

[4]   Centre for Civilisational Dialogue, University of Malaya, Kuala Lumpur 50603, Malaysia; rosilawatizai@um.edu.my

**\***   Correspondence: mahdi@um.edu.my (M.A.); syuhaida.kl@utm.my (S.I.)

**Abstract:** Walking is a sustainable commute mode, and walkability is considered an essential sign of sustainable mobility. To date, many walkability assessment tools have been developed to assess the walkability conditions across the world. However, there is a paucity of comprehensive methods to assess current walkability tools based on walking needs and ensure all walking requirements are included. Thus, researchers and experts are unable to select the most comprehensive tool systematically. The present study attempts to develop a system to evaluate the quality of the existing tools. The instrument focuses on factors related to walking needs frequently observed in all types of walkability assessment tools. Hence, a pilot measurement quality appraisal instrument (MQAI) is developed and tested by a research team with planning and public health backgrounds. The final MQAI is tested by suitable reliability, criterion, and content validity tests. Most appraisal scales display moderate to high reliability for both audits and questionnaires. The MQAI appears as ready for use in several applications, including meta-analyses and systematic reviews. Additionally, the MQAI can be used by practitioners and planners to identify the most comprehensive and efficient assessment tools based on their needs.

**Keywords:** sustainable commute mode; walkability assessment tool; measurement quality appraisal; walking environment; walking needs

## 1. Introduction

Walking is the simplest class of physical movement that benefits individual health. In addition, walking is regarded as a sustainable transport mode that benefits an individual, society, and environment [1–3]. Several studies focused on identifying pedestrian needs [4,5]. These studies identified a range of factors that affect pedestrian behavior and decisions. These factors can be summarized into four main groups that include accessibility [6–10], safety [11–16], comfort [4,5,17–19], and pleasurability [4,5,20]. A few studies also focused on a single dimension of walking needs. For example, Tiwari [21] explored the safety concerns of an individual while accessing metro stations, and Zakaria and Ujang [22] determined pedestrian comfort based on walking experience.

Several studies used these walking needs to develop assessment tools, including pedestrian level of service (PLOS) methods and walkability assessment tools [23–30]. Factors used in the aforementioned studies include accessibility, traffic factors, safety (from crime and traffic), geometry/environmental/footpath factors, pedestrian movement factors, aesthetics, comfort, attractiveness, functionality, destinations, environmental appearance, activity potential, shade, convenience, walking facilities, usability, and exploration.

The existing physical activity tools and walkability assessment tools aim to assess the walking environment and improve recreational spaces for health advancement in societies [31]. Walkability assessment tools use audits [13,32–34] and questionnaires [35,36] to collect the required data. In order to perform an audit, the streets are split into segments, and each part is examined by one or more evaluators. In audits, a set of qualitative judgments or quantitative measurements is designated for each assessment item. Like the audits, the questionnaires are effective instruments to assess pedestrian environments. The questionnaires are utilized to evaluate the perceptions of neighborhood residents towards walking and cycling facilities in their area.

According to Litman [37], walkability is considered an essential indicator of sustainable mobility. Typically, researchers and practitioners from various domains, including urban planning, transport planning, urban design, and public health have an interest in the topic of walkability. In addition, they are the main users of the walkability assessment tools. There are many walkability assessment tools, and it is a challenging task to select the best one. Furthermore, there is no guideline or a systematic manner to help these users to select the most appropriate walkability assessment tool. They need to ensure that the tool that they select to work with is comprehensive and sufficiently detailed. This is because the future investments in infrastructures may depend on this assessment. Thus, if an inappropriate tool is used, undesirable consequences will be brought about. Each type of walkability assessment tool uses certain indicators to assess the walking environment and urban design-related factors. The walking needs are extremely diverse, and thus it is important to ensure that the assessment tools consider a wide range of urban design-related factors to the maximum possible extent for assessment purposes. Consequently, there is a need to develop an instrument to appraise the strength of assessment tools to evaluate walking needs. Currently, there is a paucity of research dedicated to the measurement quality examination of walkability assessment tools [11]. The present study aims to develop a measurement quality appraisal instrument (MQAI) to evaluate walkability assessment tools based on walking needs. This paper presents the development process of the MQAI. To exhibit this process, the MQAI was applied to some walkability assessment tools and indicated the reliability, validity, and applicability of these tools. The successful development of MQAI ensures planners and researchers can efficiently employ this tool for choosing the most appropriate walkability assessment tool among the candidate tools.

## 2. Walking Needs

Various walking needs and their contributory urban design variables affect people's decision to walk. Accessibility is among the most cited walking needs that must be met to motivate people to walk. Accessibility simply refers to the ability (easiness) of obtaining desired services and activities [4,6–10]. Several urban design factors affect the accessibility needs of walking, including, but not limited to, availability/completeness of sidewalk network, number of destinations, proximity to transit points, presence/number of barriers, and public spaces.

Safety is another important walking need that is frequently found in the literature. Safety of walking refers to whether an individual feels safe from the danger of falling due to wet conditions, the hazard of conflicts with vehicles, and the threat of crime [2,4,11–14,38]. Urban design factors that may affect safety from crime include lighting, landscape and trees, and vacant buildings. Design factors that may contribute to safety from traffic include signage, signals, and pedestrian crossings. Safety from falling also can be affected by surface, materials, and lighting.

A considerable amount of literature has been published on comfort as an important need for walking. Comfort refers to a person's level of satisfaction, ease, and pleasure [4,5,17]. The design factors that may affect the comfort needs of walking include landscape and trees, the presence of traffic calming features, canopies, and drinking fountains. Pleasurability is also an important need for walking. Pleasurability simply refers to whether an individual experiences an enjoyable and interesting area for walking [4,5,20].

The presence of a varied streetscape, architectural elements, and outdoor dining areas can affect the pleasurability level of pedestrians. Table 1 illustrates the walking needs and the urban design factors that affect these needs.

**Table 1.** Walking requirements and their design considerations.

| Walking Needs | Design Factors That Affect the Walking Needs | Reference |
|---|---|---|
| Accessibility | Presence/completeness of sidewalk network | [4,12,22] |
| | Presence/number of barriers | [4,39] |
| | Distance to destinations | [4,39,40] |
| | Number of destinations | [4] |
| | Pavement width | [24] |
| | Pattern of street network | [39] |
| | Variety and proximity of activities | [4,39] |
| | Connectivity between uses | [39] |
| | Mix land use | [12,39] |
| | Alternative routes | [22] |
| | Proximity to transit points | [22] |
| | Directness of walkways/routes | [22] |
| | Accessibility of transit stations | [22] |
| | Public space | [12] |
| Safety | Landscape and trees (buffering and safety from crime) | [26,41–44] |
| | Signage | [32,35,45,46] |
| | Bollards | [44,47–49] |
| | Lighting | [12,14,32,50–52] |
| | Signals | [32,34,53,54] |
| | Surface and material | [35,46,55] |
| | Windows facing the street | [12] |
| | Abandoned buildings | [12] |
| | Rundown buildings | [12] |
| | Vacant buildings | [12] |
| | Graffiti | [12] |
| | Undesirable land uses | [12] |
| | Driveways | [56] |
| | Pedestrian crossings | [57–59] |
| | Landscape and trees | [24,60] |
| | Sidewalk width | [49,52,61,62] |
| | Surface/Material | [46,52] |

**Table 1.** *Cont.*

| Walking Needs | Design Factors That Affect the Walking Needs | Reference |
|---|---|---|
| | Lighting | [63] |
| | Cleanliness | [6] |
| | Shading | [6] |
| | Presence of traffic calming features | [4,12] |
| | i. Roundabouts/Traffic circle | |
| | ii. Medians | |
| | iii. curb bulb-outs | |
| Comfort | Length of blocks | [12,39] |
| | Buffer | [12,39] |
| | Arcades | [4] |
| | Canopies | [4] |
| | Number of lanes | [12] |
| | Street width | [12] |
| | Pavement treatment | [12] |
| | Drinking fountains | [64] |
| | Bench and sitting areas | [35,65] |
| | Slope | [52,66] |
| | Landscape and trees | [12,67] |
| | Presence of a varied streetscape | [4] |
| | Mixed uses | [4] |
| Pleasurability | Architectural elements | [4] |
| | Historic or unique architecture | [4] |
| | Color | [4] |
| | Presence of public space | [4] |
| | Outdoor dining areas | [4] |

The existing walkability assessment tools have various factor classifications. In walkability assessment tools, the major groups of assessment items are street facilities, sidewalk characteristics, land use, and road attributes. Street facilities include signage, signals, drinking fountains, surveillance, and items related to the disabled [33,34,68]. Sidewalk characteristics include items such as sidewalk completeness, the width of the sidewalk, presence/number of barriers (obstacles), and surface/material of the sidewalk [69–71]. Land use is another frequently used grouping that contains a mixture of land use, undesirable land uses, and destinations [72,73]. The walkability assessment tools also use items related to road attributes, including traffic calming features, street width, cleanliness, lighting, and directness of walkways/routes [71,74]. Table 2 presents walking needs-related factors based on the major factor classifications in the existing walkability assessment tools. The walking needs information obtained from the literature and summarized in Tables 1 and 2 were used to develop a comprehensive instrument to assess current tools based on walking needs. This instrument can assess the quality of the existing walkability assessment tools and determine their capability for assessing pedestrian environments. Such an instrument also can act as a decision-making system for selecting the most appropriate assessment tool for evaluating the walking environments.

**Table 2.** Intersections between the walking needs and the major assessment categories.

| | Sidewalk | | Land Uses and Destinations | | | Street Facilities | | | | | | | | | | | | Road Attributes | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Sidewalk Network | Path Condition | Mixture of Land Use | Undesirable Land Uses | Destinations | Signage | Signals | Drinking Fountains | Bollards | Landscape/Trees | Buffers | Benches/Sitting Areas | Surveillance | Items Related to Disabled | Streetscape Characters | Driveways | Transit Points | Traffic Calming | Features | Road Attributes | Qualitative Characteristics | Network Design |
| Accessibility | 1.2 32 | | 2.26 | | 1.7 26 | 5.10 16.28 | 5.15 27.30 | | | | | | | 3 | | | 32 | | | | | 26.32 |
| Safety | | | | 2 | | | | | 3.22 25.31 | 8.14 22.24 | 21.22 | | 2 | 3 | | 38 | | | | | 11 | 34.35 37 |
| Comfort | 4.17 18.25 | 4.28 36 | | | | | | 39 | | 20.29 | 2.26 | 16.33 | | 3 | | | | | 1.2 | 2 | 19.12 | 2.26 |
| Pleasurability | | | 1 | | | | | | | 2.23 | | | | 3 | 1 | | | | | | 1 | |

1 = Alfonzo [4]; 2 = Alfonzo, Boarnet, Day, McMillan and Anderson [12]; 3 = Asadi-Shekari, Moeinaddini and Zaly Shah [48]; 4 = Azemati, Bagheri, Hosseini and Maleki [52]; 5 = Clifton, Livi Smith and Rodriguez [32]; 6 = Crews and Zavotka [50]; 7 = Cubukcu [40]; 8 = Cui, Allan, Taylor and Lin [42]; 9 = Foster and Giles-Corti [51]; 10 = Funk [45]; 11 = Haans and de Kort [14]; 12 = Handy and Clifton [6]; 13 = Harkey and Zegeer [55]; 14 = Hernandez [41]; 15 = Karim and Azmi [53]; 16 = Kihl, Brennan, Gabhawala, List and Mittal [35]; 17 = Kim, Choi and Kim [62]; 18 = Krambeck [61]; 19 = Landis, Vattikuti, Ottenberg, McLeod and Guttenplan [63]; 20 = Lee, Jang, Wang and Namgung [60]; 21 = MacNeil [43]; 22 = Matan and Newman [44]; 23 = Rahimiashtiani and Ujang [67]; 24 = Samarasekara, Fukahori and Kubota [26]; 25 = Samarasekara, Fukahori and Kubota [49]; 26 = Sapawi and Said [39]; 27 = Slater, Nicholson, Chriqui, Barker, Chaloupka and Johnston [54]; 28 = Southworth [46]; 29 = Talavera-Garcia and Soria-Lara [24]; 30 = Tropped, Cromley, Fragala, Melly, Hasbrouck, Gortmaker and Brownson [34]; 31 = Van Cauwenberg, Van Holle, Simons, Deridder, Clarys, Goubert, Nasar, Salmon, De Bourdeaudhuij and Deforche [47]; 32 = Zakaria and Ujang [22]; 33 = Galanis and Eliou [65]; 34 = Keat, Yaacob and Hashim [57]; 35 = Kolbe-Alexander, Pacheco, Tomaz, Karpul and Lambert [59]; 36 = Monteiro and Campos [66]; 37 = O'Connor, Borscheid and Reid [58]; 38 = Otak [56]; 39 = Sarkar [64].

## 3. Methods

As previously mentioned, this paper shows the development process of the MQAI. This process included two main parts: (1) pilot version development and (2) final version development. Each part involved a series of assessments and techniques. The development process of MQAI is indicated in Figure 1.
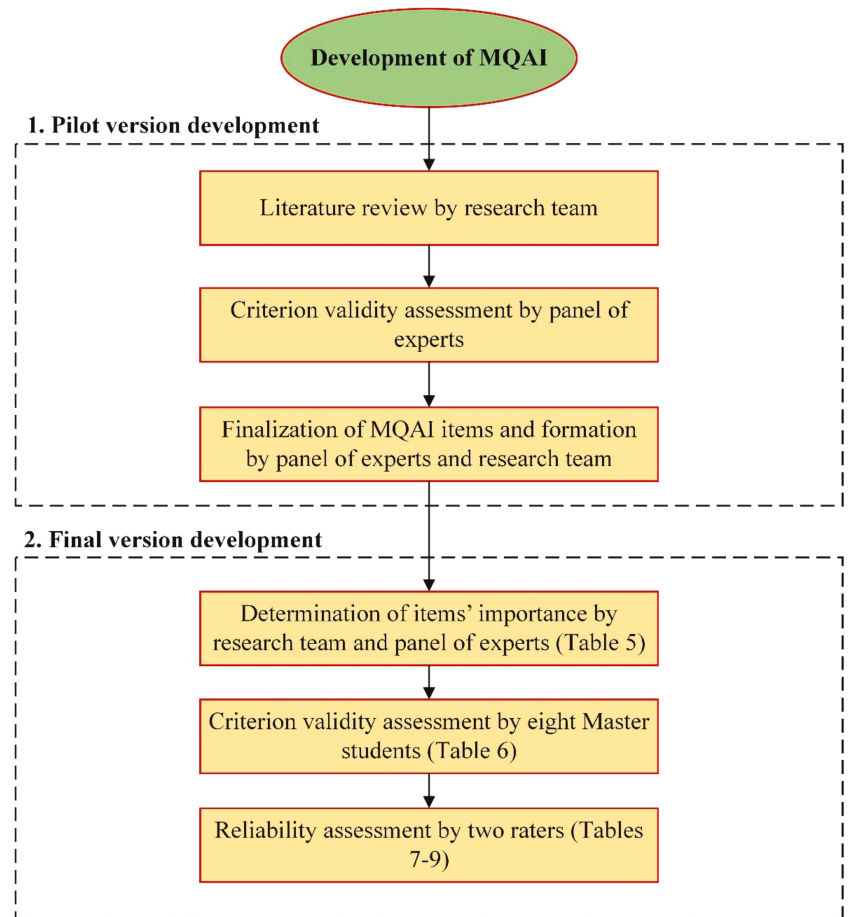


**Figure 1.** MQAI development process.

### 3.1. Identifying Walking Needs and Developing the Pilot MQAI

A literature review has been conducted to identify the walking needs and their widest range of contributory urban design factors. The walking needs information extracted from this literature (refer to Tables 1 and 2) were used to develop a pilot MQAI (refer to Appendix A) to assess the current tools based on walking needs. Table 3 lists the key characteristics of the MQAI. This tool is based on a pointing system in which each point corresponds to a specific condition. In this system, the worst and best conditions receive the lowest and greatest points, respectively. This method facilitates a systematic comparison among the walkability assessment tools and allows for determining the tools' capability for evaluating the walkability. To assess each item, the evaluator must select 'no assessment' (determines that the tool does not assess the indicator); 'simple assessment' (determines that the tool simply assesses the availability of an indicator and does not assess the quality

of indicator); 'partial assessment' (determines that the tool assesses the availability in addition to the quality but does not provide a complete assessment for the quality); and 'complete assessment' (determines that the tool presents a complete assessment (availability and quality) for the indicator). The 'no assessment', 'simple assessment', 'partial assessment', and 'complete assessment' conditions receive points of zero, one, two, and three, respectively. These four levels of responses allow for simultaneously assessing both availabilities of design factors and their assessment quality in the tools. The score of each measurement scale is computed by the sum of the marks assigned to the different items. Appendix A shows the scoring pattern and related explanations.

**Table 3.** Characteristics of MQAI.

| Type of Appraisal | Scale | Question No. |
|---|---|---|
| Sidewalk | Sidewalk network and path condition | 1–2 |
| Land uses and destinations | A mixture of land uses, undesirable land uses, and destinations | 3–5 |
| Street facilities | Signage, signals, drinking fountains, landscape and trees, bollards, buffers, benches and sitting areas, surveillance, items related to disability, streetscape characters, driveways, and transit points | 6–17 |
| Road attributes | Traffic calming features, road attributes, network design, and qualitative characteristics | 18–21 |

To investigate the content validity of the proposed MQAI, some meetings were held with a panel of experts which included two experts in urban transport planning and public health. The outcomes of these meetings were minimal changes to the content of some scales and/or the explanation attached. The pilot version of MQAI was made through the results of this step.

A criterion validity test was conducted in this step. Two pedestrian environment assessment tools, including one audit and one questionnaire, were assessed utilizing the pilot version of MQAI by the research team (authors). Each member of the research team was benchmarked relative to the team leader (first author). The average level of agreement was 41.5%.

Once the assessment of criterion validity of the MQAI pilot version was completed, the outcomes of this evaluation were discussed in a series of meetings in which both the research team and experts were involved. These meetings engaged the experts in discussion and the developing of a refined list of suitable MQAI appraisal items. During the meetings, the research team and experts confirmed the purpose and scope of the MQAI. They also ensured that the widest range of appraisal items was included in the proposed instrument. Thus, a few changes were implemented, such as adding more explanations to the description of the responses to clarify the differences between answer categories in a better way (refer to Table 4). Additional improvements included adding an instruction to respond to the appraisal items. Step-by-step instructions were provided to aid users in selecting a suitable answer concerning 'No, simple, partial, and complete' (refer to Appendix A). Additionally, a graphical scale was provided to help the users recognize the right response (refer to Figure 2).

**Table 4.** An example of an added explanation to a given question in *MQAI*.

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the condition of the path | 0 |
| Simple assessment | The tool simply assesses the path condition * in the study area | 1 |
| Partial assessment | The tool assesses the path condition, and one of the following path condition issues: (1) material ** used; and (2) slope *** | 2 |
| Complete assessment | The tool assesses the path condition and all the following path condition issues: (1) material used; and (2) slope | 3 |

* Poor (several weeds, breaks, and holes), moderate (a few weeds, breaks, and holes), good (very few weeds, breaks, and holes), under repair. ** Flat segmented concrete slabs, paving stones, Portuguese mosaic, rustic natural stones, slippery material (smooth ceramic tiles), rough material (hydraulic tiles, interlocked blocks, flattened concrete), regular, firm, antiskid, and ant vibration material (high strength paving). *** Flat or gentle, moderate slope, steep slope.
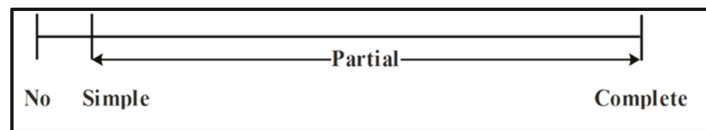


**Figure 2.** Scoring graphical scale.

*3.2. Final Version Development*

The research team and panel of experts assessed the significance of each tool item. They rated the importance of the items by utilizing a five-point scale varying between 'not important' and 'very important'. The median score for each item was calculated to determine the weight of the items. In order to gain the consensus of the research team, the team computed the agreement level for the importance of every factor. Then, the weight for each item was adjusted based on the number of items in each category.

The formula that was utilized is [weight − expected weight]. The expected weight is the score that is assigned if the items equally contributed to a category. For instance, if it is required to weigh two items, the expected weight is 2.50 for each; and if it is needed to measure four items, then the expected weight is 1.25. The inter-quartile range (IQR) is calculated for these modified weights to assess the degree of consensus among the evaluators on the scored importance of items. Items with an IQR < 1 correspond to a high level of consensus among the evaluators.

The final version of MQAI was tested for criterion validity and reliability. The reference degree of correlation and agreement for individuals with a background and familiarity with urban planning and urban design was investigated to assess the criterion validity of the MQAI. For each rater, the agreement level was calculated with respect to the leader of the research team. A total of eight students who registered for a Master of Science (advanced urban planning course) participated in this step. Two tools were selected by the team leader and were classified based on the MQAI% interpretation section (Appendix A) as poor (20 ≤ MQAI% < 40) and regular (40 ≤ MQAI% < 60). A tool was given to each student, and they were asked to complete the assignment in four days.

In order to test the reliability, two raters were asked to evaluate six walkability assessment tools (three audits and three questionnaires). The users of walking assessment tools are mainly from the domains of urban planning, transportation planning, and public health. Thus, two raters were selected, namely an urban and transport planner and a public health expert. The main goals of this step were: (1) to verify the inter-rater degree of agreement for each of the four levels of answers employed in the MQAI; and (2) to assess the inter-rater degree of agreement for each of the six tools. The inter-rater reliability was tested by using Kappa, which is a statistical measure of inter-rater reliability.

## 4. Results

Based on the IQR definition, an IQR of less than one indicates a high level of agreement, and an IQR of more than one indicates a low level of agreement. Thus, sixteen factors exhibited high levels while five factors exhibited moderate levels of consensus (Table 5). All items, including 'sidewalk', 'land use and destinations', and 'road attributes', exhibited high levels of agreement while those items with moderate levels belonged to the 'street facilities' category.

**Table 5.** Relative weightings and level of consensus on the items.

| | Median | Final Weight = [Weight − Expected Weight *] | Relative Weight (Normalized Weight) | IQR |
|---|---|---|---|---|
| **Sidewalk** | | | | |
| Sidewalk network | 5 | 2.50 | 0.50 | 0 |
| Path condition | 5 | 2.50 | 0.50 | 0.5 |
| **Land use and destinations** | | | | |
| Mixture of land use | 3 | 1.33 | 0.19 | 0.5 |
| Undesirable land uses | 4 | 2.33 | 0.33 | 0.5 |
| Destinations | 5 | 3.33 | 0.48 | 0 |
| **Street Facilities** | | | | |
| Signage | 3 | 2.58 | 0.07 | 1 |
| Signals | 4 | 3.58 | 0.09 | 1 |
| Drinking fountains | 3 | 2.58 | 0.07 | 0.5 |
| Bollards | 4 | 3.58 | 0.09 | 0 |
| Landscape and trees | 4 | 3.58 | 0.09 | 1 |
| Buffers | 4 | 3.58 | 0.09 | 0.5 |
| Benches and sitting areas | 4 | 3.58 | 0.09 | 0.5 |
| Surveillance | 4 | 3.58 | 0.09 | 0.5 |
| Items related to disabled | 4 | 3.58 | 0.09 | 0.5 |
| Streetscape characters | 4 | 3.58 | 0.09 | 1 |
| Driveways | 3 | 2.58 | 0.07 | 0.5 |
| Transit points | 3 | 2.58 | 0.07 | 1 |
| **Road Attributes** | | | | |
| Calming features | 4 | 2.75 | 0.23 | 0.5 |
| Road attributes | 4 | 2.75 | 0.23 | 0 |
| Qualitative characteristics | 4 | 2.75 | 0.23 | 0 |
| Network design | 5 | 3.75 | 0.31 | 0.5 |

High level of agreement (IQR < 1); moderate level of agreement (IQR = 1); * low level of agreement (IQR > 1). * Expected weight: sidewalk factors = 2.5; land use and destination factors = 1.67; street facilities factors = 0.42; road attribute factors = 1.25.

The final version of MQAI was tested for criterion validity and reliability. As shown in Table 6, the total baseline of agreement level between the evaluators and the team head was 82%. The lowest agreement belonged to the sidewalk scale (75%). Agreement values for the other three scales were 79% for land use and destinations, 83% for street facilities, and 88% for road attributes. The Spearman correlations were 0.78 for the regular tool and 0.92 for the poor tool. The average MQAI% for the tools was 38% for the poor tool and 43% for the regular tool. The difference in MQAI% between the 'poor' and 'regular' tools were statistically non-significant at the 5% level. Additionally, there was no statistically significant difference in MQAI% between the tools assessed by the research team leader and the tools assessed by the individuals ($p$-value = 0.3 for the poor tool; $p$-value = 0.1 for the regular tool).

**Table 6.** Baseline degree of agreement with respect to the team leader (criterion validity testing).

| | Questions | Testing of MQAI Final Version | | |
| --- | --- | --- | --- | --- |
| | | Poor Tool | Regular Tool | Overall |
| Sidewalk | 1–2 | 75% | 75% | 75% |
| Land use and destinations | 3–5 | 83% | 75% | 79% |
| Street facilities | 6–17 | 92% | 73% | 83% |
| Road attributes | 18–21 | 94% | 81% | 88% |
| Overall | 1–21 | 89% | 75% | 82% |
| Spearman correlation | 1–21 | 0.92 * | 0.78 * | 0.85 |

* Correlation is significant at the 0.01 level (2-tailed).

Table 7 reveals the inter-rater agreement level for every of the four levels of response employed in the MQAI. Table 8 presents the reliability data by appraisal type and includes the number of questions evaluated within each component. With respect to the questions assessed in audits and questionnaires, averages of 69.84% and 73%, respectively, corresponded to a high agreement ($\geq$75%) between the raters. The aggregated results of the inter-rater agreement level for each of the tool types are shown in Table 9. The weighted Kappa values for the four scales varied based on the tool type, and the K values for the audits were in the moderate to good range. Concerning the questionnaires, the K values ranged from fair/moderate to very good. The overall inter-rater reliability for the audits and questionnaires were 70% and 73%, respectively.

**Table 7.** Inter-rater degree of agreement for each of the tool types and levels of answers.

| Based on the Assessments of Three Audits | | | | |
| --- | --- | --- | --- | --- |
| | No Assessment | Simple Assessment | Partial Assessment | Complete Assessment |
| No assessment | 80% | 13.3% | 6.7% | 0.0%% |
| Simple assessment | 19.2% | 65.4% | 15.4% | 0.0%% |
| Partial assessment | 0.0% | 21.4% | 71.4% | 7.1% |
| Complete assessment | 0.0% | 0.0% | 37.5% | 62.5% |
| Based on the Assessments of Three Questionnaires | | | | |
| | No Assessment | Simple Assessment | Partial Assessment | Complete Assessment |
| No assessment | 100% | 0.0% | 0.0% | 0.0% |
| Simple assessment | 21.7% | 69.6% | 8.7% | 0.0% |
| Partial assessment | 0.0% | 70% | 30% | 0.0% |
| Complete assessment | 0.0% | 0.0% | 60% | 40% |

**Table 8.** Inter-rater degree of the agreement for each of the six tools.

| | | Audit 1 | | Audit 2 | | Audit 3 | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | | High [a] | Low [b] | High | Low | High | Low |
| Sidewalk | 1–2 | 2 | 0 | 1 | 1 | 1 | 1 |
| Land use and destinations | 3–5 | 2 | 1 | 1 | 2 | 1 | 2 |
| Street facilities | 6–17 | 11 | 1 | 7 | 5 | 8 | 4 |
| Road attributes | 18–21 | 4 | 0 | 3 | 1 | 3 | 1 |
| | | Questionnaire 1 | | Questionnaire 2 | | Questionnaire 3 | |
| | | High | Low | High | Low | High | Low |
| Sidewalk | 1–2 | 2 | 0 | 2 | 0 | 2 | 0 |
| Land use and destinations | 3–5 | 3 | 0 | 1 | 2 | 1 | 2 |
| Street facilities | 6–17 | 9 | 3 | 7 | 5 | 10 | 2 |
| Road attributes | 18–21 | 3 | 1 | 3 | 1 | 3 | 1 |

[a] Number of items with percent agreement $\geq$ 0.75. [b] Number of items with percent agreement < 0.75.

**Table 9.** Aggregated results for the inter-rater degree of agreement for the assessed tools.

|  |  | Audits | | Questionnaires | |
| --- | --- | --- | --- | --- | --- |
|  |  | Weighted Kappa | 95% CI | Weighted Kappa | 95% CI |
| Sidewalk | 1–2 | 0.57 | 0.02 to 1.00 | 1.00 | 1.00 to 1.00 |
| Land use and destinations | 3–5 | 0.45 | 0.13 to 0.77 | 0.59 | 0.21 to 0.96 |
| Street facilities | 6–17 | 0.59 | 0.37 to 0.80 | 0.62 | 0.47 to 0.76 |
| Road attributes | 18–21 | 0.79 | 0.50 to 1.00 | 0.79 | 0.60 to 0.97 |

Strength of agreement = K < 0.20: Poor; 0.21 < K ≤ 0.40: Fair; 0.41 < K ≤ 0.60: Moderate; 0.61 < K ≤ 0.80: Good; 0.81 < K ≤ 1.00: Very good.

## 5. Discussion

The baseline agreement level for the overall instrument was 82% for persons with a background in urban planning and urban design with respect to the team leader. The land use and destinations, street facilities, and road attributes scored the agreement levels in the range of 79–88%. The sidewalk scale had the lowest value, that is, a 75% agreement level. The main reason for this is that this scale includes only two items; therefore, missing an item will have a larger influence on the agreement level.

The improvement of the final version of MQAI compared to the pilot version was demonstrated through testing the final version with two raters with planning and public health backgrounds. This improvement might be related to adding instructions and the items' details. A simple check on the reliability results shows that the Kappa value is different for the same scale in audits and questionnaires. For example, the sidewalk attained a lower K value in audits than questionnaires. A possible explanation for this is the inherent difference of assessment in audits and questionnaires besides the dearth of knowledge in a specific field of proficiency. During testing of the MQAI instrument, the team noted that raters had difficulty choosing the 'partial' response. However, the raters did not experience any difficulty in assigning other response categories. The interpretation skill of raters was further significantly improved through in-depth training and supervision.

The results also showed that 'poor' tools are easier to assess than regular tools. The total scoring for a 'poor' tool by raters was very similar to that of the team leader. Based on the classification of the tools proposed in this study, the 'poor' tool represents a tool that considers a few numbers of urban design factors. Hence, the raters were required to easily score items as 'no' or 'simple'.

Both researchers in practice and academia can employ the MQAI to select the most suitable walkability assessment tool. The walkability assessment tools help decision-makers to identify shortcomings in the living environments. Decision-makers then conclude about the improvement strategies for a living environment with undesirable walking conditions. These strategies may include financial and cultural aspects, which may impact the everyday life of the residents. It is vital that a sufficient amount of investments be allocated to an area with inadequate walking conditions. A better walking infrastructure encourages people to walk and, in turn, increases the overall walking level of residents in a neighborhood. Thus, choosing a suitable walkability assessment tool that assesses the walking environment accurately is of great interest. Moreover, this can impact the plans for improving the walking conditions in a neighborhood indirectly. The employment of MQAI enables practitioners to (1) classify the walkability assessment tools, (2) select the most suitable one, and finally (3) identify walkability shortcomings within neighborhoods using the selected tool.

Researchers in academia also can benefit from the MQAI. Researchers in the domains of urban planning, transport planning, and public health need a comprehensive tool for assessing the walkability condition in a certain area and link this condition with the overall walking level in that area. Typically, this relationship is assessed using traditional statistical methods. However, the abundance of walkability assessment tools, in both the forms of audits and questionnaires, makes it challenging for these researchers to pick the most

appropriate one, which can truly reflect the walking condition within a certain area. Thus, the MQAI can help them choose the most comprehensive tool that can capture the details of the walking environment and find the associations of this environment and overall walking and physical activity levels.

**6. Conclusions**

In recent decades, walkability assessment tools have been developed to assess the suitability of a walking environment for pedestrians. These tools used numerous environmental factors in order to assess the built environments. To date, several reviews were published on walkability assessment tools, and they highlighted challenges faced by extant studies [11,69–71]. However, there is a paucity of a system for assessing the walkability assessment tools based on walking needs. The present study developed and tested an instrument to appraise walkability assessment tools based on walking needs. The main goal of the proposed instrument is to assess whether the walkability assessment tools consider the walking needs and urban design-related factors. This tool can serve as a decision-making system for researchers and practitioners to select the most appropriate assessment tool for evaluating the walking environment.

The present instrument can be used for meta-analyses and systematic reviews. This instrument is easy to use for planners and public health experts. The MQAI can aid practitioners and researchers in selecting the tool to assess the pedestrian environments in both the neighborhood and street scale based on their priorities. The instrument considers the majority of the walking needs to assess the existing tools. However, the planners can select the required items based on their priorities and adjust the proposed MQAI based on their selected items. Additionally, the instrument can serve as a base to develop future walkability assessment tools. The MQAI can be utilized to decide whether the design of a new walkability assessment tool adheres to the walking needs of diverse pedestrian groups. The MQAI did not perform the reliability and validity tests on virtual assessment tools. However, to keep abreast with new technological advancements, this tool also can be employed to assess the virtual assessment tools, which were recently released. Additionally, the methodology employed in this study can be followed to develop similar tools for assessing the virtual walkability/bikeability tools. The MQAI can also inspire future decision-making tools to select the best assessment tools that involve physical environment indicators.

**Appendix A. Measurement Quality Appraisal Instrument (MQAI)**

**Instrument Description**

- Please answer all questions pertaining to the depth of evaluation of each scale.

- To respond to the questions, one may choose one of four levels, i.e., 'No assessment', 'Simple assessment', 'Partial assessment', and 'Complete assessment'.
- Your judgment for choosing a response from 'No/Simple/Partial/Complete' should follow a three-step process. Initially, you are required to determine if the answer is 'No' (the indicator of interest is not assessed) or 'Simple'/'Partial'/'Complete' (i.e., the indicator(s) of interest is/are assessed). Second, you need to choose 'Simple' if the tool only assesses the availability of an indicator without considering the quality. Third, if the answer is 'Partial'/'Complete' (the tool considers the availability in addition to the quality), you need to decide if it is a 'Partial' (the tool does not consider all quality-related factors) or 'Complete' (the tool considers all quality-related factors). If you need to assess the tool many times to judge, there is a high probability that the response is 'Partial'.
- A different approach to determine if it is a 'Partial' or 'Complete' is by checking the scale below. As presented, the endpoints of the scale are marked with 'Complete' and 'No'. Hence, the 'Partial' response is the whole space within the 'Simple' and 'Complete'.
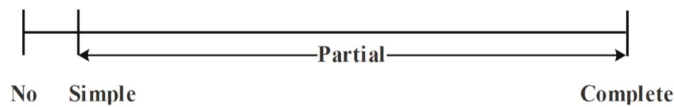


**Figure A1.** Graphical scale for determining the right response.

---

**Mathematical Calculation**

---

Mathematically, the NSAT score is defined as follows:

$MQAI\% = 100 \times \frac{\sum_{i=1}^{21} P_i\, W_i}{12}$

Here, $MQAI\%$ = strength of the tool of interest to assess the environmental factors, $P_i$ = point given by the rater to the indicator of interest, $W_i$ = relative weight of each indicator, 12 = total achievable points by each tool ($12 = \sum_{i=1}^{21} 3 \times W_i$).

---

**MQAI% Interpretation**

---

**Table A1.** Interpretation of the assessment result.

| MQAI% | Strength of Appraisal | Description |
|---|---|---|
| $80 \le MQAI\% \le 100$ | Very good | The tool considers a large number of urban design factors |
| $60 \le MQAI\% < 80$ | Good | The tool considers an appropriate number of urban design factors |
| $40 \le MQAI\% < 60$ | Regular | The tool considers an acceptable number of urban design factors |
| $20 \le MQAI\% < 40$ | Poor | The tool considers very few urban design factors |
| $0 \le MQAI\% < 20$ | Awful | The tool considers an insignificant number of urban design factors |

**Assessment Items**
**Sidewalk**

**1.     How does the tool assess the sidewalk network?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the availability of the sidewalk network along the street | 0 |
| Simple assessment | The tool only assesses the availability of the sidewalk network along the street | 1 |
| Partial assessment | The tool assesses the availability of the sidewalk along the street and one of the following issues of the sidewalk network: (1) completeness; (2) width of the sidewalk network along the street; and (3) presence/number of barriers along the sidewalk | 2 |
| Complete assessment | The tool assesses the availability of the sidewalk along the street and two to three of the following issues of the sidewalk network: (1) completeness; (2) width of the sidewalk network along the street; and (3) presence/number of barriers (obstacles) along the sidewalk | 3 |

**2.     How does the tool assess path conditions?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the condition of the path | 0 |
| Simple assessment | The tool simply assesses the path condition * in the study area | 1 |
| Partial assessment | The tool assesses the path condition, and one of the following path condition issues: (1) material ** used; and (2) slope *** | 2 |
| Complete assessment | The tool assesses the path condition and all of the following path condition issues: (1) material used; and (2) slope | 3 |

* Poor (several bumps, cracks, holes, and weeds), moderate (a few bumps, cracks, holes, and weeds), good (very few bumps, cracks, holes, and weeds), under repair. ** Flat segmented concrete slabs, paving stones, rustic natural stones, and Portuguese mosaic, slippery material (smooth ceramic tiles), rough material (hydraulic tiles, interlocked blocks, flattened concrete), regular, firm, antiskid, and ant vibration material (high strength paving). *** Flat or gentle, moderate slope, steep slope.

**Land Use and Destinations**

**3.     How does the tool assess the mixture of land use?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the mixture of land uses and activities | 0 |
| Simple assessment | The tool simply assesses the availability of various land uses and activities in the study area | 1 |
| Partial assessment | The tool assesses the availability of land uses and activities in the study area and determines the number of each activity such as residential, retail/commercial, office, public, and/or industrial | 2 |
| Complete assessment | The tool assesses the availability of land uses and activities in the study area and determines the number of each activity and overall desirable land use planning | 3 |

4.   **How does the tool assess the undesirable land uses** (e.g., dilapidated buildings, abandoned buildings, and rights of way of utilities and rail)?

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the undesirable land uses | 0 |
| Simple assessment | The tool simply assesses the availability of various undesirable land uses in the study area | 1 |
| Partial assessment | The tool assesses the availability of undesirable land uses in the study area and determines the number of each undesirable land use | 2 |
| Complete assessment | The tool assesses the availability of undesirable land uses in the study area and determines the number of each undesirable land use and the overall undesirable land use planning | 3 |

5.   **How does the tool assess the destinations** (e.g., local facilities, parks, public transport, services, shops, vehicle parking facilities, and bike parking facilities)?

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the destinations | 0 |
| Simple assessment | The tool simply assesses the availability of various destinations in the study area | 1 |
| Partial assessment | The tool assesses the availability of destinations in the study area and determines the number of destinations OR distance to destinations | 2 |
| Complete assessment | The tool assesses the availability of destinations in the study area and determines the number of destinations AND distance to destinations | 3 |

**Street Facilities**

6.   **How does the tool assess the signage?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the signage | 0 |
| Simple assessment | The tool simply assesses the availability of signage along the path | 1 |
| Partial assessment | The tool assesses the availability of signage along the path and one to three of the following issues in the signs: (1) simplicity of graphics and phrase; (2) contrast to the background; (3) placement of signs in the furnishing zone; (4) height of the signs; and (5) presentation of information in alternative formats | 2 |
| Complete assessment | The tool assesses the availability of signage along the path and four to five of the following issues in the signs: (1) simplicity of graphics and phrase; (2) contrast to the background; (3) placement of signs in the furnishing zone; (4) height of the signs; and (5) presentation of information in alternative formats | 3 |

**7.    How does the tool assess the signals?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the signals | 0 |
| Simple assessment | The tool simply assesses the availability of signals along the path | 1 |
| Partial assessment | The tool assesses the availability of signals along the path and one of the following issues in the signals: (1) counter timer for the signal poles; (2) audible signal for the signal poles; and (3) accessibility of the signal buttons to the wheelchair users | 2 |
| Complete assessment | The tool assesses the availability of signals along the path and two to three of the following issues in the signals: (1) counter timer for the signal poles; (2) audible signal for the signal poles; and (3) accessibility of the signal buttons to the wheelchair users | 3 |

**8.    How does the tool assess the drinking fountains?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the drinking fountains | 0 |
| Simple assessment | The tool simply assesses the availability of drinking fountains | 1 |
| Partial assessment | The tool assesses the availability of drinking fountains and one of the following issues in the drinking fountains: (1) distance between the drinking fountains; (2) height of drinking fountains; and (3) placement of water fountains in the furnishing zone | 2 |
| Complete assessment | The tool assesses the availability of drinking fountains and two to three of the following issues in the drinking fountains: (1) distance between the drinking fountains; (2) height of drinking fountains; and (3) placement of water fountains in the furnishing zone | 3 |

**9.    How does the tool assess the bollards?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the bollards | 0 |
| Simple assessment | The tool simply assesses the availability of bollards in the study area | 1 |
| Partial assessment | The tool assesses the availability of bollards and one of the following issues in the bollards: (1) space between bollards; (2) placement of bollards in the curb extension; and (3) height of bollards | 2 |
| Complete assessment | The tool assesses the availability of bollards and two to three of the following issues in the bollards: (1) space between the bollards; (2) placement of bollards in the curb extension; and (3) height of bollards | 3 |

**10.  How does the tool assess landscape and trees?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess landscape and trees | 0 |
| Simple assessment | The tool simply assesses the availability of landscape and trees in the study area | 1 |
| Partial assessment | The tool assesses the availability of landscape and trees and one of the following issues in the landscape and trees: (1) vertical clearance of tree branches; (2) placement of trees in the furnishing zone; and (3) distance between trees | 2 |
| Complete assessment | The tool assesses the availability of landscape and trees and two to three of the following issues in the landscape and trees: (1) vertical clearance of tree branches; (2) placement of trees in the furnishing zone; and (3) distance between trees | 3 |

**11.  How does the tool assess buffers?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the buffers | 0 |
| Simple assessment | The tool simply assesses the availability of buffers and barriers along the street such as on-street parking | 1 |
| Partial assessment | The tool assesses the availability of buffers and barriers along the street such as on-street parking and type of buffer(s) OR width of a buffer | 2 |
| Complete assessment | The tool assesses the availability of buffers and barriers along the street such as on-street parking and type of buffer(s) AND width of a buffer | 3 |

**12.  How does the tool assess benches and sitting areas?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the bench and sitting areas | 0 |
| Simple assessment | The tool simply assesses the availability of bench and sitting areas | 1 |
| Partial assessment | The tool assesses the availability of bench and sitting areas, and one to two of the following issues in the landscape and trees: (1) placement of benches in the furnishing zone; (2) distance from the curb; (3) space for parking a wheelchair or stroller; and (4) distance between the benches | 2 |
| Complete assessment | The tool assesses the availability of benches and sitting areas, and three to four of the following issues in the landscape and trees: (1) placement of benches in the furnishing zone; (2) distance from curb; (3) space for parking a wheelchair or stroller; and (4) distance between the benches | 3 |

**13.  How does the tool assess surveillance?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the surveillance | 0 |
| Simple assessment | The tool simply assesses the availability of surveillance in the study area | 1 |
| Partial assessment | The tool assesses the surveillance and active * OR passive ** surveillance | 2 |
| Complete assessment | The tool assesses the surveillance and active AND passive surveillance | 3 |

* CCTV and security patrols. ** Active frontages, façade solid-void ratio, windows, verandas, and gardens.

**14. How does the tool assess the items related to the disabled?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the items related to the disabled * | 0 |
| Simple assessment | The tool assesses one of the following issues in the items related to disabled individuals: (1) accessible drinking fountain; (2) accessible toilet; (3) tactile pavement; (4) curb cut; (5) accessible signage and signals; and (6) elevator next to the sky-bridge | 1 |
| Partial assessment | The tool assesses two to four of the following issues in the items related to disabled individuals: (1) accessible drinking fountain; (2) accessible toilet; (3) tactile pavement; (4) curb cut; (5) accessible signage and signals; (6) elevator next to sky-bridge | 2 |
| Complete assessment | The tool assesses five to six of the following issues in the items related to disabled individuals: (1) accessible drinking fountain; (2) accessible toilet; (3) tactile pavement; (4) curb cut; (5) accessible signage and signals; and (6) elevator next to the sky-bridge | 3 |

* Accessible drinking fountain, accessible toilet, tactile pavement, curb cut, accessible signage and signals, and elevator next to sky-bridge.

**15. How does the tool assess streetscape characters?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the streetscape characters * | 0 |
| Simple assessment | The tool assesses the availability of one of the following streetscape characters: (1) architectural elements; (2) historic or unique architecture; (3) presence of public space; (4) outdoor-dining areas; (5) abandoned buildings; (6) rundown buildings; and (7) vacant buildings | 1 |
| Partial assessment | The tool assesses the availability of two to five of the following streetscape characters: (1) architectural elements; (2) historic or unique architecture; (3) presence of public space; (4) outdoor-dining areas; (5) abandoned buildings; (6) rundown buildings; and (7) vacant buildings | 2 |
| Complete assessment | The tool assesses the availability of six to seven of the following streetscape characters: (1) architectural elements; (2) historic or unique architecture; (3) presence of public space; (4) outdoor-dining areas; (5) abandoned buildings; (6) rundown buildings; and (7) vacant buildings | 3 |

* Architectural elements, historic or unique architecture, presence of public space, outdoor dining areas, abandoned buildings, rundown buildings, vacant buildings.

**16. How does the tool assess driveways?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess driveways | 0 |
| Simple assessment | The tool only assesses the availability of the driveways | 1 |
| Partial assessment | The tool assesses the availability of the driveways and driveway width * OR availability of warning facilities ** | 2 |
| Complete assessment | The tool assesses the availability of the driveways and driveway width AND availability of warning facilities | 3 |

* More than a garage, equal to a garage, and less than a garage. ** Special paving, signs, auditory warning, and mirrors.

**17.** **How does the tool assess the transit points?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the transit points | 0 |
| Simple assessment | The tool simply assesses the availability of transit points along the path | 1 |
| Partial assessment | The tool assesses the proximity to transit points * OR accessibility of transit stations ** in the study area | 2 |
| Complete assessment | The tool assesses the proximity to transit points AND accessibility of transit stations in the study area | 3 |

* The proximity of transit stations to popular landmarks such as squares, towers, and malls. ** Connectivity and continuity of walkways to transit stations.

**Road Attributes**

**18.** **How does the tool assess traffic calming features?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess traffic calming features * | 0 |
| Simple assessment | The tool assesses the availability of one of the following traffic calming features: (1) roundabouts; (2) medians; (3) curb bulb-outs; (4) traffic signals; and (5) speed humps | 1 |
| Partial assessment | The tool assesses the availability of two to three of the following traffic calming features: (1) roundabouts; (2) medians; (3) curb bulb-outs; (4) traffic signals; and (5) speed humps | 2 |
| Complete assessment | The tool assesses the availability of four to five of the following traffic calming features: (1) roundabouts; (2) medians; (3) curb bulb-outs; (4) traffic signals; and (5) speed humps | 3 |

* Roundabouts, medians, curb bulb-outs, traffic signals, and speed humps.

**19.** **How does the tool assess road attributes?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess road attributes * | 0 |
| Simple assessment | - | 1 |
| Partial assessment | The tool assesses the number of lanes OR street width | 2 |
| Complete assessment | The tool assesses the number of lanes AND street width | 3 |

* Number of lanes and street width.

**20.** **How does the tool assess qualitative characteristics?**

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess traffic qualitative characteristics * | 0 |
| Simple assessment | The tool assesses one of the following issues with respect to the qualitative characteristics: (1) cleanliness; (2) lighting; (3) shading; and (4) color | 1 |
| Partial assessment | The tool assesses two of the following issues with respect to the qualitative characteristics: (1) cleanliness; (2) lighting; (3) shading; and (4) color | 2 |
| Complete assessment | The tool assesses three to four of the following issues with respect to the qualitative characteristics: (1) cleanliness; (2) lighting; (3) shading; and (4) color | 3 |

* Cleanliness/graffiti, lighting, shading, and color.

### 21. How does the tool assess the network design?

| Answer | Description | Point |
|---|---|---|
| No assessment | The tool does not assess the network design | 0 |
| Simple assessment | The tool simply assesses one of the flowing network design factors: (1) pattern of street network *; (2) directness of walkways/routes **; (3) length of blocks; (4) alternative routes ***; and (5) the presence and frequency of pedestrian crossings along the network | 1 |
| Partial assessment | The tool simply assesses two of the flowing network design issues: (1) pattern of street network; (2) directness of walkways/routes; (3) length of blocks; (4) alternative routes; and (5) the presence and frequency of pedestrian crossings along the network | 2 |
| Complete assessment | The tool simply assesses three to four of the flowing network design issues: (1) pattern of street network; (2) directness of walkways/routes; (3) length of blocks; (4) alternative routes; and (5) the presence and frequency of pedestrian crossings along the network | 3 |

* Grid or cul-de-sac. ** The number of directional changes. *** Number and type of alternative routes available between the origin and destination.

## References

1. Hatamzadeh, Y. Working commuters' tendency toward a travel pattern with potentially more walking: Examining the relative influence of personal and environmental measures. *Res. Transp. Econ.* **2021**, *86*, 101060. [CrossRef]
2. Oxley, J.; O'Hern, S.; Burtt, D.; Rossiter, B. Falling while walking: A hidden contributor to pedestrian injury. *Accid. Anal. Prev.* **2018**, *114*, 77–82. [CrossRef] [PubMed]
3. Peng, H.; Ma, X.; Chen, F. Examining Injury Severity of Pedestrians in Vehicle–Pedestrian Crashes at Mid-Blocks Using Path Analysis. *Int. J. Environ. Res. Public Health* **2020**, *17*, 6170. [CrossRef] [PubMed]
4. Alfonzo, M.A. To walk or not to walk? The hierarchy of walking needs. *Environ. Behav.* **2005**, *37*, 808–836. [CrossRef]
5. Mehta, V. Walkable streets: Pedestrian behavior, perceptions and attitudes. *J. Urban.* **2008**, *1*, 217–245. [CrossRef]
6. Handy, S.L.; Clifton, K.J. Evaluating Neighborhood Accessibility: Possibilities and Practicalities. *J. Transp. Stat.* **2001**, *4*, 67–78.
7. Litman, T. Evaluating accessibility for transportation planning. *Vic. Transp. Policy Inst. Vic. Can.* **2012**, *1*, 1–50.
8. Litman, T. *Evaluating Transportation Equity*; Victoria Transport Policy Institute: Victoria, BC, Canada, 2013.
9. Singleton, P.A. *A Theory of Travel Decision-Making with Applications for Modeling Active Travel Deman*; Portland State University: Portland, OR, USA, 2013.
10. Majid, M.R.; Nordin, A.; Johar, F.; Tifwa, H. Travel Emission Profile of Iskandar Malaysia Neighbourhoods from Pre-1980s to 2000s. *IOP Conf. Ser. Earth Environ. Sci.* **2014**, *18*, 12161. [CrossRef]
11. Aghaabbasi, M.; Moeinaddini, M.; Zaly Shah, M.; Asadi Shekari, Z.; Kermani, M.A. Evaluating the capability of walkability audit tools for assessing sidewalks. *Sustain. Cities Soc.* **2018**, *37*, 475–484. [CrossRef]
12. Alfonzo, M.; Boarnet, M.G.; Day, K.; McMillan, T.; Anderson, C.L. The Relationship of Neighbourhood Built Environment Features and Adult Parents' Walking. *J. Urban. Des.* **2008**, *13*, 29–51. [CrossRef]
13. Evenson, K.R.; Sotres-Alvarez, D.; Herring, A.H.; Messer, L.; Laraia, B.A.; Rodríguez, D.A. Assessing urban and rural neighborhood characteristics using audit and GIS data: Derivation and reliability of constructs. *Int. J. Behav. Nutr. Phys. Act.* **2009**, *6*, 44. [CrossRef]
14. Haans, A.; de Kort, Y.A.W. Light distribution in dynamic street lighting: Two experimental studies on its effects on perceived safety, prospect, concealment, and escape. *J. Environ. Psychol.* **2012**, *32*, 342–352. [CrossRef]
15. Park, Y.; Garcia, M. Pedestrian safety perception and urban street settings. *Int. J. Sustain. Transp.* **2020**, *14*, 860–871. [CrossRef]
16. Ulak, M.B.; Kocatepe, A.; Yazici, A.; Ozguven, E.E.; Kumar, A. A stop safety index to address pedestrian safety around bus stops. *Saf. Sci.* **2021**, *133*, 105017. [CrossRef]
17. Shumi, S.; Zuidgeest, M.; Martinez, J.; Efroymson, D.; van Maarseveen, M. Understanding the Relationship Between Walkability and Quality-of-Life of Women Garment Workers in Dhaka, Bangladesh. *Appl. Res. Qual. Life* **2015**, *10*, 263–287. [CrossRef]
18. Fahed, J.; Kinab, E.; Ginestet, S.; Adolphe, L. Impact of urban heat island mitigation measures on microclimate and pedestrian comfort in a dense urban district of Lebanon. *Sustain. Cities Soc.* **2020**, *61*, 102375. [CrossRef]
19. Kim, H.; Lee, K.; Kim, T. Investigation of pedestrian comfort with wind chill during winter. *Sustainability* **2018**, *10*, 274.
20. Lindelöw, D.; Svensson, Å.; Sternudd, C.; Johansson, M. What limits the pedestrian? Exploring perceptions of walking in the built environment and in the context of every-day life. *J. Transp. Health* **2014**, *1*, 223–231. [CrossRef]
21. Tiwari, R. Designing a safe walkable city. *Urban. Des. Int.* **2014**, *20*, 12–27. [CrossRef]

22. Zakaria, J.; Ujang, N. Comfort of Walking in the City Center of Kuala Lumpur. *Procedia Soc. Behav. Sci.* **2015**, *170*, 642–652. [CrossRef]

23. Cerin, E.; Chan, K.W.; Macfarlane, D.J.; Lee, K.Y.; Lai, P.C. Objective assessment of walking environments in ultra-dense cities: Development and reliability of the Environment in Asia Scan Tool–Hong Kong version (EAST-HK). *Health Place* **2011**, *17*, 937–945. [CrossRef] [PubMed]

24. Talavera-Garcia, R.; Soria-Lara, J.A. Q-PLOS, developing an alternative walking index: A method based on urban design quality. *Cities* **2015**, *45*, 7–17. [CrossRef]

25. Sdoukopoulos, A.; Verani, E.; Nikolaidou, A.; Gavanas, N.; Pitsiava-Latinopoulou, M.; Mikiki, F.; Mademli, E.; Pallas, C. Development and implementation of walkability audits in Greek medium-sized cities: The case of the Serres' city centre. *Transp. Res. Procedia* **2017**, *24*, 337–344. [CrossRef]

26. Samarasekara, G.N.; Fukahori, K.; Kubota, Y. Environmental correlates that provide walkability cues for tourists: An analysis based on walking decision narrations. *Environ. Behav.* **2011**, *43*, 501–524. [CrossRef]

27. Millington, C.; Ward Thompson, C.; Rowe, D.; Aspinall, P.; Fitzsimons, C.; Nelson, N.; Mutrie, N.; Collaboration, S.P.-t.S.P.A.R. Development of the Scottish Walkability Assessment Tool (SWAT). *Health Place* **2009**, *15*, 474–481. [CrossRef]

28. Ferrer, S.; Ruiz, T.; Mars, L. A qualitative study on the role of the built environment for short walking trips. *Transp. Res. Part. F: Traffic Psychol. Behav.* **2015**, *33*, 141–160. [CrossRef]

29. Aghaabbasi, M.; Moeinaddini, M.; Shah, M.Z.; Asadi-Shekari, Z. A new assessment model to evaluate the microscale sidewalk design factors at the neighbourhood level. *J. Transp. Health* **2017**, *5*, 97–112. [CrossRef]

30. Nasrin, S.; Afifah, F. Walkability assessment tool for a developing country. *IOP Conf. Ser. Mater. Sci. Eng.* **2019**, *527*, 12073. [CrossRef]

31. Albers, P.N.; Wright, C.; Olwoch, J. Developing a south African pedestrian environment. *South. Afr. J. Sci.* **2010**, *106*, 1–8.

32. Clifton, K.J.; Livi Smith, A.D.; Rodriguez, D. The development and testing of an audit for the pedestrian environment. *Landsc. Urban. Plan.* **2007**, *80*, 95–110. [CrossRef]

33. Pikora, T.J.; Bull, F.C.; Jamrozik, K.; Knuiman, M.; Giles-Corti, B.; Donovan, R.J. Developing a Reliable Audit Instrument to Measure the Physical Environment for Physical Activity. *Am. J. Prev. Med.* **2002**, *23*, 187–194. [CrossRef]

34. Troped, P.J.; Cromley, E.K.; Fragala, M.S.; Melly, S.J.; Hasbrouck, H.H.; Gortmaker, S.L.; Brownson, R.C. Development and reliability and validity testing of an audit tool for trail/path characteristics: The Path Environment Audit Tool (PEAT). *J. Phys. Act. Health* **2006**, *3*, S158. [CrossRef] [PubMed]

35. Kihl, M.; Brennan, D.; Gabhawala, N.; List, J.; Mittal, P. Livable Communities: An Evaluation Guide. Available online: http://assets.aarp.org/rgcenter/il/d18311_communities.pdf (accessed on 28 June 2020).

36. Saelens, B.E.; Sallis, J.F.; Black, J.B.; Chen, D. Neighborhood-based differences in physical activity: An environment scale evaluation. *Am. J. Public Health* **2003**, *93*, 1552–1558. [CrossRef] [PubMed]

37. Litman, T. *Well Measured Developing Indicators for Sustainable and Livable Transport Planning*; Victoria Transport Policy Institute: Victoria, BC, Canada, 2019; Volume 19.

38. Tchinda, P.E.; Kim, S.-N. The paradox of "eyes on the street": Pedestrian density and fear of crime in Yaoundé, Cameroon. *Sustainability* **2020**, *12*, 5300. [CrossRef]

39. Sapawi, R.; Said, I. Application of Rasch Model in Constructing Walkability Indices for Urban Neighborhood Area. *Asian J. Environ. Behav. Stud.* **2013**, *4*, 1–14. [CrossRef]

40. Cubukcu, E. Walking for Sustainable Living. *Procedia Soc. Behav. Sci.* **2013**, *85*, 33–42. [CrossRef]

41. Hernandez, D. Vulnerable Populations and the Built Environment. *Am. J. Public Health* **2013**, *103*, E32–E34. [CrossRef]

42. Cui, J.; Allan, A.; Taylor, M.A.; Lin, D. Developing Shanghai underground pedestrian system under urbanization: Mobility, functionality and equity. *J. Archit. Urban.* **2012**, *36*, 283–297. [CrossRef]

43. MacNeil, L. *Steps to a Walkable Community: A Guide for Citizens, Planners, and Engineers*; National Academies of Sciences Engineering Medicine: Washington, DC, USA, 2012.

44. Matan, A.; Newman, P. Jan Gehl and new visions for walkable Australian cities. *Spec. Ed. A Future Beyond Car* **2012**, *17*, 30–41.

45. Funk, C. *Walkability of Transit-Oriented Development: Evaluating the Pedestrian Environment of Metro Vancouver's Regional City Centres*; Queen's University: Kingston, ON, Canada, 2012.

46. Southworth, M. Designing the walkable city. *J. Urban. Plan. Dev.* **2005**, *131*, 246–257. [CrossRef]

47. Van Cauwenberg, J.; Van Holle, V.; Simons, D.; Deridder, R.; Clarys, P.; Goubert, L.; Nasar, J.; Salmon, J.; De Bourdeaudhuij, I.; Deforche, B. Environmental factors influencing older adults' walking for transportation: A study using walk-along interviews. *Int J. Behav Nutr Phys. Act.* **2012**, *9*, 85. [CrossRef]

48. Asadi-Shekari, Z.; Moeinaddini, M.; Zaly Shah, M. Disabled pedestrian level of service method for evaluating and promoting inclusive walking facilities on urban streets. *J. Transp. Eng.* **2013**, *139*, 181–192. [CrossRef]

49. Samarasekara, G.; Fukahori, K.; Kubota, Y. *Walkability Evaluation of Streetscapes: Development of Prediction Equations for Walking Needs of Tourists*; University of Moratuwa: Moratuwa, Sri Lanka, 2013; pp. 110–117.

50. Crews, D.E.; Zavotka, S. Aging, disability, and frailty: Implications for universal design. *J. Physiol. Anthropol.* **2006**, *25*, 113–118. [CrossRef]

51. Foster, S.; Giles-Corti, B. The built environment, neighborhood crime and constrained physical activity: An exploration of inconsistent findings. *Prev. Med.* **2008**, *47*, 241–251. [CrossRef]

52. Azemati, H.R.; Bagheri, M.; Hosseini, S.B.; Maleki, S.N. An assessment of pedestrian networks in accessible neighborhoods: Traditional neighborhoods in Iran. *Int. J. Archit. Eng. Urban. Plan.* **2011**, *21*, 52–59.
53. Karim, H.A.; Azmi, D.I. Convenience and Safety of Walking Experience in Putrajaya Neighbourhood Area. *Procedia Soc. Behav. Sci.* **2013**, *101*, 318–327. [CrossRef]
54. Slater, S.J.; Nicholson, L.; Chriqui, J.; Barker, D.C.; Chaloupka, F.J.; Johnston, L.D. Walkable communities and adolescent weight. *Am. J. Prev. Med.* **2013**, *44*, 164–168. [CrossRef]
55. Harkey, D.L.; Zegeer, C.V. *PEDSAFE: Pedestrian Safety Guide and Countermeasure Selection System*; Federal Highway Administration: Washington, DC, USA, 2004; p. 330.
56. GDOT. *Pedestrian and Streetscape Guide*; GDOT: Monticello, IL, USA, 2003; pp. 1–222.
57. Keat, L.K.; Yaacob, N.M.; Hashim, N.R. Campus Walkability in Malaysian Public Universities: A Case-Study of Universiti Malaya. *J. Malays. Inst. Plan.* **2016**, *14*, 101–114. [CrossRef]
58. O'Connor, D.; Borscheid, M.; Reid, O. Northside Partnership "Healthy Community" Pilot Area Walkability Audit. *Proc. Ir. Transp. Res. Netw.* **2015**. [CrossRef]
59. Kolbe-Alexander, T.L.; Pacheco, K.; Tomaz, S.A.; Karpul, D.; Lambert, E.V. The relationship between the built environment and habitual levels of physical activity in South African older adults: A pilot study. *BMC Public Health* **2015**, *15*, 518. [CrossRef]
60. Lee, B.J.; Jang, T.Y.; Wang, W.; Namgung, M. Design criteria for an urban sidewalk landscape considering emotional perception. *J. Urban. Plan. Dev.* **2009**, *135*, 133–140. [CrossRef]
61. Krambeck, H.V. *The Global Walkability Index*; Massachusetts Institute of Technology: Cambridge, MA, USA, 2006.
62. Kim, S.; Choi, J.; Kim, Y. Determining the sidewalk pavement width by using pedestrian discomfort levels and movement characteristics. *KSCE J. Civ. Eng.* **2011**, *15*, 883–889. [CrossRef]
63. Landis, B.W.; Vattikuti, V.R.; Ottenberg, R.M.; McLeod, D.S.; Guttenplan, M. Modeling the roadside walking environment: Pedestrian level of service. *Transp. Res. Rec. J. Transp. Res. Board* **2001**, *1773*, 82–88. [CrossRef]
64. Sarkar, S. Qualitative evaluation of comfort needs in urban walkways in major activity centers. *Transp. Q.* **2003**, *57*, 39–59.
65. Galanis, A.; Eliou, N. Evaluation of the pedestrian infrastructure using walkability indicators. *WSEAS Trans. Environ. Dev.* **2011**, *7*, 385–394.
66. Monteiro, F.B.; Campos, V.B. A proposal of indicators for evaluation of the urban space for pedestrians and cyclists in access to mass transit station. *Procedia Soc. Behav. Sci.* **2012**, *54*, 637–645. [CrossRef]
67. Rahimiashtiani, Z.; Ujang, N. Pedestrian Satisfaction with Aesthetic, Attractiveness and Pleasurability: Evaluating the Walkability of ChaharaghAbbasi Street in Isfahan, Iran. *ALAM CiptaInt. J. Sustain. Trop. Des. Res. Pract.* **2013**, *6*, 13–22.
68. Clifton, K.J.; Livi, A.; Rodríguez, D.A. Pedestrian Environment Data Scan (PEDS) Tool. Available online: http://planningandactivity.unc.edu/RP1.htm (accessed on 6 February 2020).
69. Emery, J.; Crump, C.; Bors, P. Reliability and validity of two instruments designed to assess the walking and bicycling suitability of sidewalks and roads. *Am. J. Health Promot.* **2003**, *18*, 38–46. [CrossRef]
70. Evenson, K. *PIN3 Neighborhood Audit Instrument*; MIDSS: Galway, Ireland, 2009.
71. Boarnet, M.G.; Day, K.; Alfonzo, M.; Forsyth, A.; Oakes, M. The Irvine—Minnesota inventory to measure built environments: Reliability tests. *Am. J. Prev. Med.* **2006**, *30*, 153–159. [CrossRef]
72. Brownson, R.C.; Ramirez, L.K.B.; Hoehner, C.M.; Cook, R.A. *Analytic Audit Tool and Checklist Audit Tool*; BEL: Trois-Rivières, QC, Canada, 2003; pp. 1–9.
73. Hoehner, C. *Active Neighborhood Checklist*; BEL: Trois-Rivières, QC, Canada, 2011.
74. Cain, K.L.; Millstein, R.A.; Geremia, C.M. *Microscale Audit of Pedestrian Streetscapes (MAPS): Data Collection & Scoring Manual*; University California San Diego: San Diego, CA, USA, 2012. Available online: http://sallis.ucsd.edu/measures/maps (accessed on 6 February 2020).

*Article*

# Connectivity in Superblock Street Networks: Measuring Distance, Directness, and the Diversity of Pedestrian Paths

**Martin Scoppa * and Rim Anabtawi**

Department of Architectural Engineering, United Arab Emirates University, Abu Dhabi P.O. Box 15551, United Arab Emirates; rim.anabtawi@uaeu.ac.ae
* Correspondence: martin.scoppa@uaeu.ac.ae

**Abstract:** Superblocks are a common urban development strategy used in cities of the United Arab Emirates and the larger Gulf region. In planning new neighborhoods, these cities utilize superblocks structured using various street network designs. Despite their key role in shaping its main transportation network, the connectivity of these designs has not been frequently studied. This paper addresses this research gap, analyzing ten different superblock designs, and focusing on their internal and external connectivity properties. Internal connectivity is studied by measuring connections between plots in the superblocks. External connectivity is measured from plots to the superblocks' corners, the points from which to access surrounding areas. Connectivity is measured in terms of distance, directness, and route diversity. The results show that strong similarities exist across the studied designs, particularly in terms of travel distances. Differences are found in terms of efficiency and, most notably, route diversity. Findings are discussed in relation to walkability, the costs associated to each design given network length variations, and the importance of creating rich and diverse street systems that support open-ended exploration. While based on a sample of ideal cases and in need of validation with built cases, this paper outlines a method by which to evaluate and compare superblock network design alternatives.

**Keywords:** sustainable urban form; urban networks analysis; street connectivity; Arab Gulf urbanization

## 1. Introduction

The role that city form plays in building more sustainable cities has been intensely investigated in the past decades [1–4]. In this work, numerous urban form descriptors linked to sustainability were recurrently discussed, including compactness, density, and land use mix and diversity [5]. From a transportation research point of view, these descriptors are often organized under the conceptual framework known as the 3Ds, denoting density, diversity and, importantly given the focus of this paper, design [6]. Since its publication, this framework was instrumental in the development of numerous studies on sustainable transportation, outlining how in compact, dense, and diverse cities, origins and destinations locate closer together, making walking and cycling viable, and making the operation of transit systems more efficient. Thus, in sustainable transportation research, city form is seen as a means by which to effect modal shifts, reducing the use of automobiles and, consequently, reducing fossil fuels consumption, air pollution, and greenhouse gas emissions [7–11].

In terms of design, researchers have often focused on street network connectivity, analyzing the networks themselves, and associated elements such as block sizes and lengths, intersection types, and overall road pattern descriptions such as curvilinear and gridiron [12–14]. However, despite the large number of articles that studied street connectivity as a key component linking urban form to more sustainable transportation modes, not much work has concentrated on superblocks and superblocks-built cities. In fact, with some recent exceptions, few studies addressed the connectivity of these urban street systems,

even when constituting a key component in the urbanization strategy of many countries worldwide, and particularly the Middle East [15–19].

Superblocks can be initially and summarily described as large tracts of land bounded by arterial roads, whose land use planning is strongly connected to the Perry's neighborhood planning unit concept [20–22], and to planning the principles outlined by key figures of the modern movement in architecture [23]. In the United Arab Emirates—and in Abu Dhabi in particular—superblocks are key to its urban development, as is the case in countries of the Gulf Council Cooperation (GCC), which includes Bahrain, Kuwait, Qatar, Oman, and Saudi Arabia. However, despite their widespread use, much remains to be understood regarding their connectivity properties.

This article addresses this research gap, focusing on understanding how different superblock network designs connect residents to one another, as well as to surrounding areas. More precisely, several metrics are applied to measure the connections between plots inside the superblocks, addressing their internal connectivity properties, as well as connections between plots and the superblocks corners—points from which to access surrounding areas—so addressing their external connectivity. In this latter case, this paper foregrounds the need to better understand how superblocks integrate with one another, acting as modules in a city building strategy and not as isolated communities.

Connectivity is examined using three metrics. These are distance, route directness, and a measure of route diversity. Using these metrics, three different but related questions about the internal and external connectivity of different superblocks network designs are addressed. First, how metrically close to one another are plots in the different designs studied, and how far are these plots from the superblocks' corners? Second, and noting that not only distance, but also the availability of direct routes between origins and destinations affects pedestrian access [24–26], the question is: how direct are the routes connecting residents of superblocks to one another, and to the superblock corners? Lastly, the third question asks: how many alternative routes, i.e., how much route diversity, is available to residents traversing the superblocks in search of internal destinations and corners? This last metric addresses the extent to which different networks provide alternative routes to pedestrians, allowing paths to mix and overlap, increasing the potential for social encounter and economic opportunity [27].

When planners and designers are confronted with the decision of which network design to adopt, there is not much research that can support their decision-making process. Addressing, how long, how direct, and how diverse pedestrian routes are in different designs, as studied in this paper, provides information that could assist in the evaluation of design alternatives. With street forming being the long-term framework over which cities grow, and with streets taking a key role in supporting more sustainable transportation modes, these are seen as timely questions to address.

## 2. Literature Review

### 2.1. Abu Dhabi and the Endurance of Superblock Planning

When studying superblocks and superblocks planned cities, Abu Dhabi represents an outstanding case of their application. Since the beginning of its urbanization drive in the mid-1960s, this city has consistently applied superblocks as the main strategy for its development. Historically coinciding with the peak of the implementation of the modern movement city planning propositions, Abu Dhabi, as many other cities in the region (notably C. Doxiadis' planned cities like Islamabad, Pakistan, 1959; Baghdad, Iraq, 1955; and Riyadh, Saudi Arabia, 1968), adopted the notion of neighborhood unit and efficient motorized transportation as the guiding principles for its development [20,22,28,29]. As a result, Abu Dhabi presents a grid of arterial roads whose spacing varies in different areas of the city, but which tend to enclose rectangular superblocks whose sides span several hundred meters. Today, superblocks in Abu Dhabi accommodate different building types, densities, and land use mixes, showcasing the ability of superblocks to adapt and respond to different development goals and to changing urban growth dynamics. These range from

higher densities and land use mixes in the central, older districts, to the characteristically low-density residential neighborhoods developed in the past two decades in the city's periphery. Particularly in these latter areas, large aggregations of identical superblocks accommodate the city's need for growth and expansion, shaping the city's suburbs. Figure 1 shows, in the first row, the use of superblocks in Abu Dhabi, noting the variety of network types and built form that characterizes the city center (Figure 1A). It shows, as well, the repetition of patterns and lower densities found in the new neighborhoods in the periphery (Figure 1B,C). Examples of their application in the region, and in the planning of extensive areas are shown in the second row (Figure 1D,E).
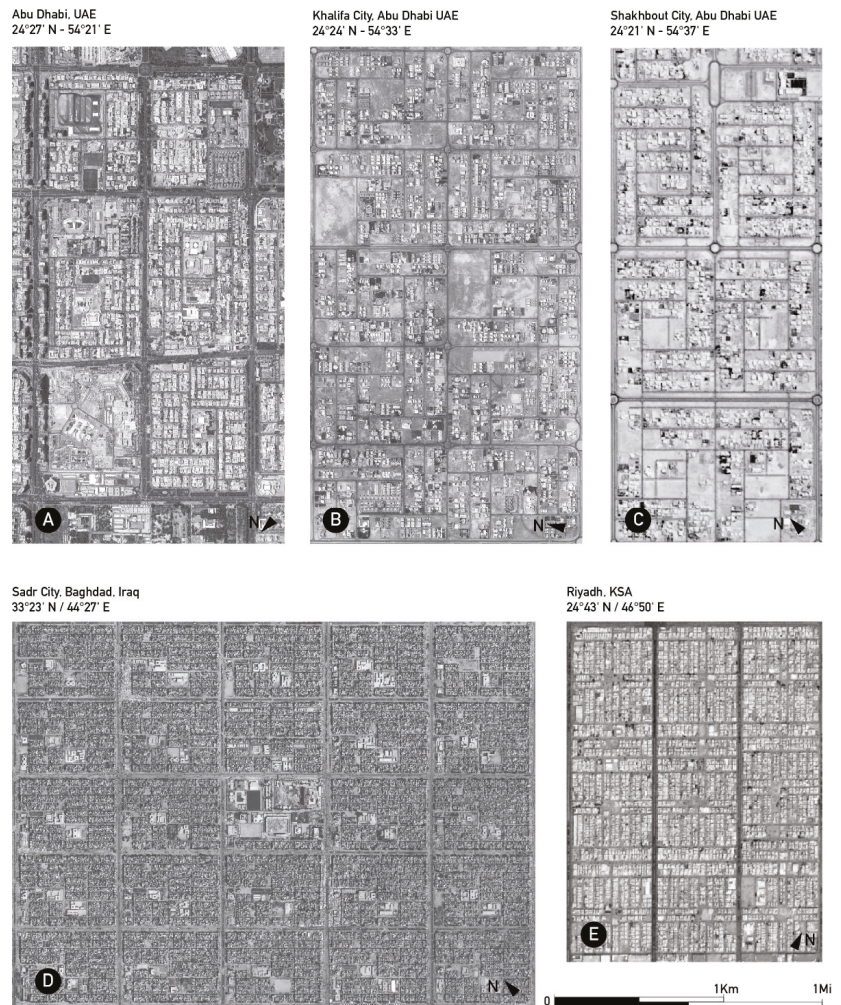


**Figure 1.** (**A**) shows Abu Dhabi's downtown area, where land use mix and densities are higher, and superblocks' designs mix. (**B**,**C**) show how repetitive designs structure suburban areas. Images of Baghdad, Iraq (**D**) and Riyadh, KSA (**E**), illustrate the widespread use of superblocks in the region.

A closer study of Abu Dhabi's planning history reveals that it was not until 1968 that the straight roads and superblocks that characterize the city today started to be built [28,30]. However, once adopted, superblocks proved to be the dominant urban planning strategy,

and endured over time. In fact, the original approach to planning the city using superblocks was regularly reaffirmed in the revisions that followed, the latest being the Master Directive Plan by consulting firm Atkins in the 1990s. Although modifications such as wider rights of way, and larger plots were introduced, the superblock was retained as the basic module by which to grow the city and develop its neighborhoods [31]. As a result of this long-term application of superblocks, Abu Dhabi proves to be a valuable source of superblock designs for this study.

### 2.2. Measures of Street Connectivity

The interest in better understanding street connectivity and its ability to support more sustainable mobility modes, especially walking, resulted in numerous metrics being proposed for its quantification. A relatively recent review highlights that various connectivity metrics are applied in practice and research, and concludes that no standard approach to its measurement exists [32]. Still, among the numerous metrics proposed, a major distinction can be made between what can be termed per-area connectivity metrics, and a network-based analysis of street connectivity.

In terms of the former, these have generally taken the form of densities, such as block, street, and intersections density, all of which are highly correlated to one another [12]. Besides being utilized in numerous studies on sustainable transportation and neighborhood design, these metrics have also been favored in regulatory instruments and practice given that they can be calculated, and legislated, with relative ease [26,32,33]. However, while useful given these advantages, aggregate metrics can also obscure connectivity variations within study areas, as demonstrated by Peponis and colleagues [34]. Further, per area metrics were also found to be susceptible to manipulation, or able to be "gamed", meeting established standards even when connectivity is low [33]. Lastly, per-area metrics are unable to handle origins and destinations. Thus, connectivity properties affecting the decision to walk, such as distance and directness, are not accurately measured using per area metrics [9,35–37].

In terms of network-based analyses, the most extensively developed sets of metrics used in urban studies include space syntax [38,39], metric and directional reach [34], and multiple centrality assessment [40,41]. The metrics these authors proposed have been instrumental in quantifying topological adjacency and centrality variations in street systems, providing a foundation by which to better understand key properties of cities, such as the distribution of pedestrian traffic, land use location patterns, and the emergence and consolidation of urban centers [42–45]. However, by placing the focus on configuration and centrality, these approaches have not directly addressed travel distances between sets of origins and destinations, such as plots, specific intersections, or land uses of interest. Therefore, while being a valuable reference, providing the most advanced methods to analyze urban street networks, these metrics do not directly and efficiently address the questions of this study.

Positioned between aggregate measures and the work on networks described in the previous paragraph, planning and transportation scholars have often studied urban street networks in terms accessibility. By measuring the separation between origins, usually residential plots, and various types of destinations, often commercial land uses, parks, bus stops, and educational facilities, among others, researchers evaluated whether and how the design of neighborhoods' street networks affect transportation mode choice, especially walking trips [46–49]. Among the metrics used in these studies, simple metric distance to destinations, and the modifications of it, such as pedestrian route directness (PRD) [50] have often been used. These types of metrics are especially relevant for this study, given their focus on connections between specific origins and destinations pairs, and their application in studies linking walkability and street network design. Further details about this type of metrics, and how they were used in this study, are discussed in the following section.

## 3. Materials and Methods

The analyses presented in this paper were conducted using standard representations of streets networks used in urban and transportation planning. That is, the superblocks' street networks are represented and studied as sets of nodes and links, with nodes representing origins and/or destinations for trips, and links representing the street centerlines over which travel occurs. The ten street networks studied are presented in Figure 2, along with a description of their general characteristics.



**Superblock 1 (SB1)**
Smaller rectangular perimeter blocks and internal diagonals for enhanced solar access. Numerous connections to boundary roads.

**Superblock 2 (SB2)**
Small rectangular blocks enclose a larger central block for community land uses. Interlocked blocks and resemblance to SB9. Numerous connections to boundary roads.

**Superblock 3 (SB3)**
Numerous inward facing cul de sacs enclose a larger central block for community land uses. Connections to boundary roads are extremely limted with only 1 on each of the short sides.

**Superblock 4 (SB4)**
Outward facing diagonals structure top and bottom portions. Larger block for community land uses divide top and bottom. Numerous connections to boundary roads.

**Superblock 5 (SB5)**
Deformed grid with curved blocks sides arranged in interlocking pattern to discourage through-traffic. Connections to the peripheral roads are limited to about 2 per side.

**Superblock 6 (SB6)**
Small rectangular blocks arranged in interlocking pattern to dicourage through traffic. Intersparsed larger blocks for community land use. Numerous connections to boundary roads.

**Superblock 7 (SB7)**
Perfect rectangular grid. No interlocking. No block size variations and thus no hierarchy or distinction. All roads connect to the boundary roads.

**Superblock 8 (SB8)**
Interlocking gridiron pattern placed as an 'island' at the center of the superblock. Connections to boundary roads limited to 1 per side.

**Superblock 9 (SB9)**
Interlocking pattern of large blocks. Square block at the center for community land uses. Connections to the periphery are limted to 2 per side.

**Superblock 10 (SB10)**
Deformed grid with central diagonally-placed block. Interlocking pattern to discourage through-traffic. Number of connections to the peripheral roads vary between 1 and 5.

**Figure 2.** The sample of ten network designs used in this paper. All superblocks have the same size and number of plots. All designs were developed from existing and frequently used cases found in the city of Abu Dhabi, UAE.

These networks were derived from real cases found in Abu Dhabi—often in its suburbs—and were slightly adjusted to make fair comparisons possible among them. Specifically, all designs were formatted to have the exact same size (860 m × 590 m), with

dimensions that correspond to the average size of superblocks found in new neighborhood developments in this city. Further, non-orthogonal superblocks were squared, so to have parallel sides and boundary roads that meet at 90-degree angles. In sum, by controlling and removing shape distortions, and by keeping their sizes constant, network distances could be fairly measured and evaluated across the sample. In terms of origins and destinations, 100 plots were randomly distributed along the superblock streets. Keeping the origins and destinations constant permitted, as well, fair comparisons among the different networks studied, and simplifying of the interpretation of the results.

Lastly, the corners of the superblocks were also placed as nodes in the network, noting that, in most cases, these are key points from which it is possible to access the adjacent superblocks/neighborhoods. The analysis of how long, direct, and diverse the routes are to the corners complemented the analysis of the internal connectivity of the different designs in the sample. Figure 2 shows how plots were distributed over the ten street networks' designs studied while their general network and block subdivision characteristics are presented in Table 1.

**Table 1.** Descriptive statistics of the superblock designs studied. Highest and lowest values in bold.

| Network Design ID | Total Road Length (mts.) | Average Block Face Length (mts.) | Intersections | | | Total Blocks | Average Block Area (ha.) |
|---|---|---|---|---|---|---|---|
| | | | 3-Way (T-Junctions) | 4-Way (Crossroads) | Total | | |
| SB1 | 10,593 | 105.9 | 30 | **15** | 45 | 31 | 1.63 |
| SB2 | **11,378** | 76.9 | **76** | 4 | **80** | **43** | **1.18** |
| SB3 | 10,473 | **73.8** | 53 | 4 | 57 | 12 | 4.22 |
| SB4 | 9766 | 113.6 | 42 | 3 | 45 | 25 | 2.02 |
| SB5 | 7693 | 150.8 | 22 | 4 | 26 | 16 | 3.16 |
| SB6 | 8962 | 121.1 | 30 | 7 | 37 | 23 | 2.21 |
| SB7 | 6390 | 206.1 | 10 | 6 | 16 | 12 | 4.22 |
| SB8 | 7490 | 182.6 | 20 | **0** | 20 | 11 | 4.61 |
| SB9 | **5769** | **262.2** | 12 | **0** | **12** | **7** | **7.24** |
| SB10 | 7480 | 162.6 | 24 | 2 | 26 | 15 | 3.38 |

The data presented in Table 1 provide important additional information regarding the designs studied. One of the main differences found between the different designs relates to the street length needed to structure each of them. Further variations across the different designs are found in terms of block lengths, the number and type of intersections, as well as the number of internal blocks that the street networks define. As expected, the bigger the road length, the smaller the blocks and the higher the number of intersections.

Table 1 shows that SB2 and SB9 tend to concentrate, respectively, most of the maxima and minima across the sample of studied designs. Interestingly, these two designs share some common traits, such as the large central block with T-intersections at its four corners. These two network designs show the extremes to which superblocks in Abu Dhabi are fractioned or aggregated. Values range from 43 blocks of 1.18 hectares on average in SB2, to 7 blocks of 7.24 hectares in SB9. Average block faces, on the other hand, vary from 73 to 262 m long, while intersections range from 80 to only 12, and a doubling of the road length in SB2 (11.7 km) when compared to SB9 (5.7 km). These extreme road length differences foreground that the costs involved in building and maintaining roads can vary quite substantially between different designs.

### 3.1. Addressing Internal and External Connectivity

With the superblock street networks built, the analyses focused on the ability of the different designs to perform two different but related tasks. These were, first, the ability to connect residents to one another, facilitating intra-neighborhood connectivity and pedestrian movement. In this sense, these analyses addressed the notion of superblocks as well-defined communities where destinations are accessible within a 5 min walk, as outlined in the original neighborhood planning unit (NPU) concept [20]. The second task addressed the ability of the different designs to provide residents of the studied superblocks with access to adjacent areas. In this case, each plot was defined as an origin for trips, and

the four corners of the superblock were set as the destinations. Corners are, in most cases, the locations where pedestrian infrastructure such as traffic lights and zebra crossings are found, thus supporting safe crossings into adjacent superblocks. In sum, the analyses conducted addressed both the internal or inter-plot connectivity, as well as the external or inter-block connectivity.

In terms of the analyses, intra-neighborhood connections were studied by measuring connectivity and routes between each individual plot and its 99 neighbors, a process that was repeated for each plot in each superblock, totaling 9900 trips once the all-plots-to-all-plots analyses were completed for a given design. External connectivity, the ability of residents to reach points from which they can cross into surrounding areas, was studied by setting all 100 plots in the superblocks as origins for trips and measuring the characteristics of the routes connecting the plots to the four corners, i.e., the destinations. As the analyses were completed, each plot completed a total of four trips, one to each corner, so a total of 400 trips to the corners were analyzed in each superblock design.

### 3.2. Measuring Distance, Directess, and Diversity

The internal and external connectivity of the ten network designs was quantified using three different metrics. The first focused on a key, though often overlooked, metric in street network analyses: distance. Specifically, the analysis of travel distances was conducted using ESRI's ArcMap Network Analyst, recording the shortest trip lengths between origins and destinations. Using this software's origin-destination (OD) cost matrix, calculations were performed using a proprietary multiple-origin, multiple-destination algorithm based on Dijkstra's [51] shortest path algorithm. These analyses answer the question of how far apart from each other are, on average, the plots in each of the studied designs. Further, graphic details about the calculation of this metric are presented in Figures 3 and 4, while Equation (1) below shows a formal definition of this metric.

$$Distance\ [i]\ =\ \frac{1}{n}\sum_{j\neq i}^{n}d(i,j) \tag{1}$$

where *Distance* [i] is the shortest network distance $d$ from origin plot $i$ to all destinations $j$, with $j$ being plots or corners depending on whether the analyses are of internal or external connectivity, and $n$ is the total number of destinations reached.

The second metric focused on the efficiency by which each of the studied networks connects the origins and destinations sets. The metric used in this case was pedestrian route directness, or PRD, and was also calculated using ArcMaps's Network Analyst. In this case, the shortest network routes obtained in the previous analyses were divided by the length of straight lines that connect origins and destinations. Used frequently in sustainable planning regulatory instruments, such as Abu Dhabi's own Estidama Sustainability Rating System [52], this ratio is easy to calculate and interpret indicating, as a percentage, how much longer than the shortest possible route is the actual travel distance. A formal definition is introduced in Equation (2), while a graphic representation of this metric is introduced in Figures 3 and 4.

$$Route\ Directness\ [i] = \frac{1}{n}\sum_{j\neq i}^{n}d_{i,j}/d_{i,j}^{Eucl} \tag{2}$$

where *Route Direcness* [i] is the directness value of origin plot $i$; $d_{ij}$ is the shortest network distance from origin plot $i$ to all destinations $j$, with $j$ being plots or corners depending on whether the analyses are of internal or external connectivity; $d_{i,j}^{Eucl}$ is the Euclidean or crow-fly distance from origin plot $i$ to all destinations $j$, and $n$ is the total number of destinations reached.

Once calculated, the PRD results obtained for each plot and in each design were evaluated using a PRD test [12,17,50,53,54]. This test permits the quantification of the

efficiency of the studied designs by identifying the number (or percentage) of plots that have values above or below a given threshold number. Previous studies have set this threshold at 1.6, while a value of 1.5 has been outlined in the connectivity guidelines of Abu Dhabi's Estidama [55]. Considering this, the two thresholds were studied and the total number of plots that meet these thresholds were reported and interpreted.

Lastly, an additional metric was computed to evaluate route diversity in the different superblock designs. This property of the networks was addressed using the urban network analysis (UNA) toolbox's redundancy index [56,57]. This index considers that while a single shortest route connects an origin and a destination, additional longer routes could also be considered as viable alternatives to gain a better understanding of the potential offered by urban street layouts. In setting up the analysis of redundancy, a detour ratio needed to be established. This value is used to determine the extra length that is permitted to be travelled between an origin and a destination. A detour ratio of 20%, as used in this study, considered that routes that are up to 20% longer than the shortest one are valid options, and their lengths were thus measured.

A mathematical formulation of the redundancy index, based on the results of the empirical testing of the metric is presented in Equation (3).

$$Redundancy\ Index\ [O,D] = \frac{1}{d_{\min(O,D)}} \sum_{path} d(path; O, D) \tag{3}$$

where *Redundancy Index* $[O,D]$ denotes the redundacy index between an origin $O$ and a destination $D$; $d_{\min(O,D)}$ is the shortest path connecting $O$ and $D$; and $d(path; O, D)$ is the length of paths connecting $O$ and $D$. In this paper, this sum is restricted to paths that obey this condition: $d(path; O, D)$ in $\left(d_{\min(O,D)},\ 1.2\,d_{\min(O,D)}\right)$. This index thus, expresses the diversity of paths as a ratio between all available paths within a 20% detour distance, and the shortest path. A value of 2, for example, would indicate a doubling of the shortest route experience, while a value of 1 would indicate that no alternative routes are available. Figures 3 and 4, show a graphic representation of this metric for the internal and external connectivity cases.

The results of the redundancy calculations are, in this paper, interpreted and discussed in terms of the diversity of routes provided by each design. This interpretation is preferred, given that the notion of redundancy in traffic analysis and infrastructure system management, is linked not only to route availability, but also spare capacity [58,59]. Considering that spare and carrying capacities are not a concern of this study, and that the metric authors foreground route choice as a valued quality of urban environments in terms of the everyday experiences they provide [60], the use of diversity is thus preferred.

Further, a valuable theoretical basis for the need to address route diversity in urban networks can be found in Jane Jacobs's discussion of the need for small blocks in cities [27]. The insightful discussion presented in this chapter highlights the ills of "self-isolating streets" and the "long sterile promenades" that are characteristic of superblock projects. Smaller blocks and denser road networks, in contrast, are discussed as key elements that bring life and vibrancy to the city streets. As blocks get smaller, route alternatives increase, along with the opportunity for city dwellers to mix their paths. In addition, the potential number of users of any given street would also increase, providing businesses a larger pool of potential customers. Measuring route diversity, thus, could provide insights regarding the potential for social interaction, as well as economic opportunity, that each street network provides.

## INTERNAL CONNECTIVITY (INTER-PLOT)

**Metric Distance**  **Route Directness**  **Route Diversity**



**Key**

● Origin
○ Destination

0      250 meters

Trips from 1 origin
to 1 destination.
Solid black line indicates
route length measured.
Radiating lines illustrate
the other 98 routes
studied. All origin
destination pairs were
measured. (9,900 trips)

Lines indicate network
(solid) and Euclidean
(dashed) distances.
The ratio of these two
distances is PRD.
Calculations addressed
all origin-destination
pairs. (9,900 trips)

Lines indicate shortest
(solid) and alternative
(dashed) routes (up to
20% longer than shortest
possible route). The ratio
of all routes to the shortest
one is the diversity index.
Calculations addressed
all origin-destination
pairs. (9,900 trips)

**Figure 3.** Internal connectivity analyses: a graphic explanation of the three metrics studied in the article. Note that in connecting each of the 100 plots in the sample to every other plot, a total of 9900 trips were evaluated.

## EXTERNAL CONNECTIVITY (INTER-BLOCK)

**Metric Distance**  **Route Directness**  **Route Diversity**



**Key**

● Origin
○ Destination

0      250 meters

Trips from 1 origin
to the 4 corners.
Solid black lines indicate
route length measured.
Radiating lines illustrate
the other 396 routes
studied. All origin
corner pairs were
measured. (400 trips)

Lines indicate network
(solid) and Euclidean
(dashed) distances.
The ratio of these two
distances is PRD.
Calculation addressed
all origin-corner
pairs. (400 trips)

Lines indicate shortest
(solid) and alternative
(dashed) routes (up to
20% longer than shortest
possible route).
Calculation addressed
all origin-corner
pairs. (400 trips)

**Figure 4.** External connectivity analyses: a graphic explanation of the three metrics studied in the article. Note that in connecting each of the 100 plots in the sample to each of the superblock's four corners, a total of 400 trips were evaluated.

## 4. Results

### 4.1. Addressing Trip Length and Metric Properties of the Routes

Measuring travel distance from each plot to all its neighboring plots in the superblock provided answers to two key questions. First, how far from one another do plots tend to be in the different designs studied? Second, and perhaps more importantly, are there significant variations in travel distances between plots in the different network designs studied?

The results presented in Table 2 indicate that even though network designs are substantially different, plots tend to be, on average, within 576.47 m from one another (SD = 102.71). Within the sample, superblock design 7 (SB7), the perfect orthogonal grid, provides for the closest average proximity between neighbors with inter-plot trips distances averaging 502 m.

**Table 2.** Descriptive statistics of the all-plots-to-all-plots metric distance analyses (internal connectivity) in columns 1 to 5. Results of the analyses of metric distance from plots to the four corners of the superblock (external connectivity) are presented in successive columns, 6 to 10. Highest and lowest values are in bold.

| Network Design ID | Avg. Distance between Plots (mts) | Std. Deviation Distance between Plots (mts.) | Min. Distance between Plots (mts.) | Max. Distance between Plots (mts.) | Avg. Distance to Corners. (mts) | Std. Deviation Distance to Corners (mts.) | Min. Distance to Corners (mts.) | Max. Distance to Corners (mts.) |
|---|---|---|---|---|---|---|---|---|
| SB1 | 562.27 | 73.3 | 447 | 771 | 713.51 | 23.36 | 685.77 | **717.15** |
| SB2 | 592.46 | 68.5 | 466 | 751 | 724.84 | 6.72 | 720.9 | 754.97 |
| SB3 | **715.92** | 145.6 | **533** | **1140** | **948.84** | 147.52 | 721.94 | **1318.06** |
| SB4 | 600.03 | 54.1 | 516 | 728 | 695.50 | 26.3 | **635.1** | 766 |
| SB5 | 551.54 | 83.4 | 399 | 728 | 719.26 | 8.07 | 711.3 | 742.79 |
| SB6 | 538.00 | 81.9 | 376 | 734 | 723.69 | >1 | 722.69 | 724.65 |
| SB7 | **502.11** | 92.3 | **373** | **708** | 724.66 | >1 | **724.65** | 724.65 |
| SB8 | 528.88 | 60.6 | 442 | 767 | 781.46 | 32.12 | 707.96 | 870.63 |
| SB9 | 619.12 | 89.5 | 488 | 790 | 724.67 | >1 | 724.64 | 724.67 |
| SB10 | 554.44 | 73.8 | 418 | 742 | 719.82 | 12.7 | 699.32 | 756.31 |

On the other hand, in superblock design 3 (SB3), a largely introverted network design with many cul-de-sacs, inter-plot trips average 716 m. Following SB3, the next highest inter-plot distances are found in network designs SB9 and SB4, while the remaining superblocks' average distances are below 600 m. Maximum trip lengths tend to stay relatively constant, and in the 700 m range, except for SB3 at 1140 m. These maximum distances describe the distances from the worst located plot in each of the studied superblocks. At the other end, minimum trip lengths show more variability, although SB3 still features the longest inter-plot distances with 533 m, about 40% longer that the minimum trip lengths provided by SB7 and SB6. In other words, the best located plot in SB3 is located 533 m away from all other plots, while the best located plots in SB7 and SB6 are, respectively, 373 and 376 m away from all other plots.

The second step addressed external connectivity, by checking how metrically far away are the superblock corners from each plot in the sample of superblocks. In this case, the results are quite striking. The values of trips to the corners of seven out of 10 superblocks designs indicate that their networks provide access to the corners within a narrow band of values ranging from 713 to 724 m. The exception cases are designs SB3, SB4, and SB8. In the case of design SB4, this has the lowest values in the sample, a likely product of the diagonals that characterize the outward oriented design of this superblock. These diagonals effectively connect the plots to the corners, reducing the overall travel distances and resulting in this being the only design with values below 700 m. At the other end, the highest values are found in SB3, already noted as the most introverted design in the sample, and in SB8 which also features an introverted street design pattern. Lastly, it is worth noting that the standard deviation of the length of trips to the corners shows several cases with values under 1 m.

### 4.2. Addressing Route Efficiency and Directness Properties of the Routes

Route efficiency was addressed by following the PRD test method, as earlier described. The tests were conducted using the two established thresholds, thus also evaluating the sensibility of the test. Once again, internal and external connectivity were tested, and the results are presented in Table 3. In the case of internal connectivity, the tested thresholds indicate that there exist variations in the efficiency of the routes connecting plots to one another depending on the network design. Extreme cases are SB3 and SB7, showcasing, respectively, the lowest and highest numbers of plots passing the test, regardless of the threshold used. More precisely, all plots pass the test—irrespective of the threshold used— in the case of SB7, while only a single plot passes the test in SB3 when the more demanding threshold is used. When the threshold is relaxed, only eight plots pass the test in SB3. The remaining cases vary, with SB9 performing quite low in both tests, followed by SB4. The remaining cases oscillate between 40 and 60 plots passing when the threshold of 1.5 is used, and above 70 in all cases when the 1.6 threshold is used. These results indicate that in contrast to the relatively homogeneous performances observed when studying metric distances, different superblock designs provide quite extreme differences in terms of the route efficiencies enjoyed by their occupants. They also show that the choice of threshold could affect the interpretation of results.

**Table 3.** Descriptive statistics of the all-plots-to-all-plots pedestrian route directness (PRD) and results of the PRD test (internal connectivity efficiency) in columns 1 to 4. Results of the analyses of directness between plots and the four corners of the superblock (external connectivity efficiency) are presented in successive columns. Highest and lowest values in bold.

| Network Design ID | All Plots to All Plots Directness (Internal Connectivity) | | | All Plots to Four Corners Distance Directness (External Connectivity) | | |
|---|---|---|---|---|---|---|
| | Avg. PRD Value | Passing Plots Threshold 1.5 | Passing Plots Threshold 1.6 | Avg. PRD Value | Passing Plots Threshold 1.5 | Passing Plots Threshold 1.6 |
| SB1 | 1.51 | 53 | 79 (+26) | 1.26 | 100 | 100 |
| SB2 | 1.49 | 54 | 84 (+30) | 1.27 | 100 | 100 |
| SB3 | **2.01** | **1** | **8 (+7)** | **1.97** | **20** | **26** |
| SB4 | 1.62 | 19 | 41 (+22) | **1.26** | 100 | 100 |
| SB5 | 1.48 | 58 | 79 (+21) | 1.27 | 100 | 100 |
| SB6 | 1.52 | 44 | 73 (+29) | 1.29 | 100 | 100 |
| SB7 | **1.35** | **100** | **100** | 1.28 | 100 | 100 |
| SB8 | 1.47 | 61 | 72 (+11) | 1.60 | 51 | 70 |
| SB9 | 1.65 | 22 | 38 (+16) | 1.28 | 100 | 100 |
| SB10 | 1.51 | 47 | 70 (+23) | 1.28 | 100 | 100 |

When studying the connectivity of plots to the corners, thus addressing the possibility of crossing over to surrounding areas, most superblock designs perform quite well, with all plots meeting the limits of the two thresholds tested. The exceptions are SB3 and SB8, characteristically introverted designs where the connection of the internal network to the peripheral roads (and thus the corners) occurs at only a few locations. It is only in these two cases that the average value of PRD exceeds 1.3, and the only two cases where the number of plots passing the PRD test is less than the totality. It is worth noting that PRD values approximating 1.3 are characteristic of regular grids, that is, orthogonal grids where roads tend to intersect at 90-degree angles.

### 4.3. Addressing Route Diversity

The last series of analyses involved the measurement of route diversity. As in the previous two sets of analyses, internal and external connections were evaluated. The results, presented in Table 4, indicate that there are wide variations in terms of what different designs can offer, if the study of connectivity between origin–destination pairs is relaxed to include routes that are up to—in this paper's case—20% longer than the shortest possible ones. The results indicate that plots in SB2 can access neighboring plots with an

exposure of up to 3.25 times the length of the shortest path. At the lower end, SB9 provides a much more limited experience and range of opportunities for route alternatives, with only 1.58 times the length of the shortest path accessible, if up to 20% longer routes are allowed when connecting origins and destinations. Across the ten design samples, five designs show diversity index values below two, indicating that in all these cases, the potential for alternative route building exists, although it does not reach a doubling of what is provided by the shortest route. Four cases show that the shortest route length is at least doubled, while a high tripling of values is only found in one case, the previously discussed SB2.

**Table 4.** Descriptive statistics of the all-plots-to-all-plots route diversity analyses in columns 1 to 5. Results of the analyses of route diversity from plots to the four corners are presented in successive columns. Highest and lowest values are in bold.

| Network Design ID | All Plots to All Plots Route Diversity (Internal Connectivity) | | | | All Plots to Four Corners Route Diversity (External Connectivity) | | | |
|---|---|---|---|---|---|---|---|---|
| | Avg. Diversity Index | Std. Deviation | Min. Diversity Index | Max. Diversity Index | Avg. Diversity Index | Std. Deviation | Min. Diversity Index | Max. Diversity Index |
| SB1 | 2.14 | 0.44 | 1.52 | 3.39 | 2.74 | 0.41 | 1.75 | 3.83 |
| SB2 | **3.25** | 0.62 | 1.93 | 4.62 | **3.94** | 0.78 | 2.49 | 5.69 |
| SB3 | 2.29 | 0.25 | 1.82 | 2.8 | 2.48 | 0.6 | 1.03 | 3.56 |
| SB4 | 2.30 | 0.35 | 1.73 | 3.5 | 2.55 | 0.64 | 1.52 | 4.22 |
| SB5 | 1.87 | 0.35 | 1.42 | 2.8 | 2.72 | 0.48 | 1.99 | 3.61 |
| SB6 | 2.18 | 0.42 | 1.61 | 3.3 | 3.32 | 0.54 | 1.71 | 4.31 |
| SB7 | 1.76 | 0.29 | 1.3 | 2.68 | 2.62 | 0.24 | 2.13 | 3.33 |
| SB8 | 1.72 | 0.22 | 1.4 | 2.26 | **1.91** | 0.48 | 1.04 | 2.83 |
| SB9 | **1.58** | 0.19 | 1.23 | 2.02 | **1.66** | 0.48 | 1 | 2.73 |
| SB10 | 1.88 | 0.29 | 1.36 | 2.9 | 2.57 | 0.55 | 1.37 | 3.61 |

When looking at access to the superblock corners, there are similar variations, although route diversity numbers tend to be larger across the sample and only two cases, SB8 and SB9, show values below 2. When considering the diversity of routes to the corners, the results indicate that most designs offer alternative routes adding up to two-and-a-half times, or more, the length of the shortest possible one.

*4.4. Network Properties and Block Subdivision Characteristics in Relation to Connectivity*

With the results of the three sets of analyses completed it is now possible to outline observations that link these connectivity analyses to the general characteristics of the network designs. A summary of the analyses performed is presented to Table 5. In the first columns, the network design characteristics show that the length of the road systems varies quite substantially across the sample. Expectedly, and as earlier noted, this network length increase is associated with a higher number of intersections, and a higher number of blocks, and consequently smaller blocks with shorter faces. Following these columns, the results of the internal and external connectivity analyses are reintroduced. Table 5 gives a comprehensive overview of the characteristics and performance of each of the designs studied.

It is worth reviewing Table 5 and noting the large differences in the length, and consequently in the building costs, that characterize these different networks. The lengths of streets in the studied networks vary from slightly more than 11 km in SB 1, 2, and 3 to about almost exactly half of that length in SB9's 5.7 km. The remaining designs fill the middle ground in terms of total roads' lengths. Table 5 serves as a comprehensive reference capturing the network characteristics and connectivity properties of all the studied designs. However, the relationship between differences in road length, intersections, and block sizes, and the results of the distance, directness, and diversity analyses, can be more clearly understood by studying their correlation. These are presented in Table 6.

**Table 5.** Summary of superblock general characteristics and results of the three analyses performed.

| Network Design ID | Superblock Network Design General Characteristics | | | | | All Plots to All Plots (Internal Connectivity) | | | All Plots to Four Corners | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Total Road Length (mts.) | Avg. Block Face Length (mts.) | Total Inter-Sections | Total Blocks | Avg. Block Area (Ha.) | Avg. Inter-Plot Distance (mts.) | Avg. Inter-plot PRD [1] | Avg. Diversity Index | Avg. Distance to Corners (mts.) | Avg. PRD to Corners | Avg. Diversity Index |
| SB1 | 10,593 | 106.1 | 45 | 31 | 1.63 | 562.27 | 1.51 (53/79) | 2.14 | 713.51 | 1.26 | 2.74 |
| SB2 | 11,378 | 77.3 | 80 | 43 | 1.18 | 592.46 | 1.49 (54/84) | 3.25 | 724.84 | 1.27 | 3.94 |
| SB3 | 10,473 | 74.3 | 57 | 12 | 4.22 | 715.92 | 2.01 (1/8) | 2.29 | 948.84 | 1.97 | 2.48 |
| SB4 | 9766 | 114.2 | 45 | 25 | 2.02 | 599.78 | 1.62 (9/41) | 2.30 | 695.50 | 1.26 | 2.55 |
| SB5 | 7693 | 151.7 | 26 | 16 | 3.16 | 551.54 | 1.48 (58/79) | 1.87 | 719.26 | 1.27 | 2.72 |
| SB6 | 8962 | 121.6 | 37 | 23 | 2.21 | 538.00 | 1.52 (44/73) | 2.18 | 723.69 | 1.29 | 3.32 |
| SB7 | 6390 | 206.7 | 16 | 12 | 4.22 | 502.11 | 1.35 (100) | 1.76 | 724.66 | 1.28 | 2.62 |
| SB8 | 7490 | 183.4 | 20 | 11 | 4.61 | 528.88 | 1.47 (61/72) | 1.72 | 781.46 | 1.60 | 1.91 |
| SB9 | 5769 | 262.3 | 12 | 7 | 7.24 | 619.12 | 1.65 (22/38) | 1.58 | 724.67 | 1.28 | 1.66 |
| SB10 | 7480 | 163 | 26 | 15 | 3.38 | 554.44 | 1.51 (47/70) | 1.88 | 719.82 | 1.28 | 2.57 |

[1] Values in parentheses show number of passing plots with PRD thresholds of 1.5 and 1.6, respectively.

**Table 6.** Analysis of correlations between measures' network characteristics and connectivity metrics based on Pearson's correlation coefficient. Significant correlations are highlighted in bold.

| | | Avg. Inter-Plot Distance | Avg. Inter-Plot PRD | Avg. Inter-Plot Diversity Index | Avg. Distance to Corners | Avg. PRD to Corners | Avg. Diversity Index to Corners |
|---|---|---|---|---|---|---|---|
| **Road length** | R-Squared | 0.17 | 0.11 | **0.72** | 0.059 | 0.046 | **0.436** |
| | Prob > F | 0.228 | 0.33 | **0.0017** | 0.496 | 0.5515 | **0.0376** |
| **Intersections** | R-Squared | 0.22 | 0.11 | **0.92** | 0.061 | 0.035 | **0.552** |
| | Prob > F | 0.168 | 0.346 | **0.0001** | 0.4915 | 0.6018 | **0.0138** |
| **Block Size** | R-Squared | 0.141 | 0.121 | **−0.63** | −0.009 | 0.081 | **−0.483** |
| | Prob > F | 0.2845 | 0.3233 | **0.006** | 0.3674 | 0.4225 | **0.0256** |

The analyses indicate that road lengths, intersections, and block sizes are only significantly associated to internal and/or external route diversity values. More precisely, correlations are strong and positive in the case of road lengths and, particularly, intersections. As expected, correlations between internal and external route diversity and block size are negative, foregrounding how route diversity increases as blocks get smaller. Distances between plots, and plots and corners are in all cases weak and not significant. Lastly, average values of PRD are also found to be weakly associated to road length, intersections, and block sizes.

## 5. Discussion

Ten superblock designs were tested to better understand their internal and external connectivity characteristics. In the case of internal connectivity analyses, the focus was placed on their ability to support connections between residents of the superblocks. These analyses thus examined how different designs support the formation of well-connected and walkable communities which could, in turn, support the formation of vibrant neighborhood life. Further analyses concentrated on accessing the corners of the superblocks, addressing the ease with which residents of superblocks could, by foot, cross over to and access adjacent areas. In doing this, external connectivity analyses addressed the need to consider superblocks as modules of a city building strategy and not as isolated units.

Results of the analyses indicate, first, that trips within the ten superblocks network designs studied tend to remain within walkable distances. Regardless of the design adopted, the average distance from a plot to its neighbors was found to be, in most cases, within a 500 to 600 m range. When trips to the corners, so the possibility of reaching surrounding areas, are considered, the results show that corners are located at walkable, and quite constant, average distances from the plots. In this case, at approximately 720 m. Several observations can be derived from these results.

The first is that superblocks—in this paper sized to approximate average sizes found in Abu Dhabi—provide for walkable distances, especially when considering the standard 5 and 10-min walking ranges used in transportation studies [61] and present in the original

NPU concept [20]. These distances correspond to quarter and half mile radii, or 400 and 800 m, respectively. However, the most notable finding is that this relatively constant distribution of trips' distances, for internally and externally oriented trips, is obtained with widely different road network designs. Designs that have, as well, widely different road lengths. These results highlight that the extension of the road networks does not significantly affect the length of trips that connect superblocks' residents to one another. Further, network designs have no significant effect on the length of trips linking residents to the corners, the points from which they can access surrounding areas.

The study of directness complemented the results of the metric distance analyses, highlighting the role that the network design has on trip efficiency. The results indicate, first, that internal connectivity efficiency can vary quite substantially depending on the network design used. In this case, two PRD test thresholds were evaluated, and the results indicate that the different networks provide differently in terms of the directness with which other plots could be accessed. First, it was found that the perfect grid (SB7) is the most efficient street system of all those studied, with all plots passing the test regardless of the threshold used. Second, SBs 1, 2, 5, 6, and 8, were found to have about 50% of their plots passing the test when the more demanding threshold was used. These values logically increased as the threshold was slightly relaxed. Lastly, superblocks SB3, 4, and 9, were the worst performing ones, with values below 50 regardless of the threshold used. When analyzing the efficiency of trips to the corners, most designs performed remarkably well (all plots passed the tests) except for SB3 and SB8.

The analysis of directness also allows for several important observations. First, the results indicate that routes between plots and the corners tend to be quite efficient. Except for cases where internal roads have limited connections to the boundary arterials—SB3 and SB8—plots can reach corners with direct routes. It is also clear that the efficiency of routes to the corners is not affected by the two thresholds tested. Lastly, and importantly, designs that have widely different road lengths, tend to provide highly efficient routes to the corners. Internal connectivity efficiency, on the other hand, presents a more complex scenario. In the first place, there exist substantial variations in the efficiency of the routes between plots, touching on extremes: note SB3 and SB7. Secondly, results were found to be sensitive to threshold variations. Still, designs with different road lengths provide substantially different internal trips efficiencies. For example, SB9, one of the worst performing cases in terms of internal connectivity efficiency, performs better than SB3 and SB4, designs with almost double the street length. In another key example, all the plots of the regular grid of SB7, one of the shortest networks, reach all other plots with PRD values below the two thresholds studied.

Finally, the study of the route diversity, based on the measurement of redundancy, shows that the richness of routes and potential for path overlap and open-ended exploration offered by the different designs can be markedly different. From the dearth of alternatives offered by SB9, to the richness of paths offered by the dense grid of SB2. Trips connecting plots to one another, as well as trips to the corners, tend to be more diverse as the availability of streets, as described by the length of the street network, increases.

While based on a sample of ideal conditions, and in need of validation with real cases, these results provide insightful information for neighborhood planning and design, as well as decision making. More specifically, the results indicate that when confronted with the decision to adopt one design over another, distances between plots and between plots and corners, tend to remain within walkable ranges in all the designs studied. If trip distances were the only walkability criteria applied for selecting a particular design, decisions could gravitate towards the costs associated to road building, knowing that access would not be compromised. The same criteria could be adopted in the case of the efficiency of routes to the corners. Knowing that most designs provide efficient access to the corners, the decision making could also be tied to the costs associated with road building. Still, the results indicate that internal connectivity efficiency can vary quite substantially and that test thresholds, road lengths, and importantly the design of the network play key

roles in the efficiency of internal trips. In this case, more research is needed to provide reliable conclusions.

Lastly, the diversity of routes offered, as measured by the redundancy index, is the only factor directly and significantly associated with the road lengths of the studied designs. This is the case both in terms of internal connections, and in terms of connections between plots and the superblock corners. While this is, in retrospect, an expected result, the fact that it was quantified and associated to specific network designs makes the study of route diversity in superblocks particularly rewarding. It provides a simple and interpretable quantitative means to evaluate the potential for route-building of each network.

As research on superblocks continues, questions regarding the maximization of desirable characteristics for superblock networks, such as increased route diversity, proximity, and efficiency, while minimizing the costs of road building, are expected to be addressed. However, the results of this paper contribute, at this point, to enlarge recent research on superblocks' street networks by including metrics that were not previously discussed, applied, or investigated. Indeed, while syntactic properties of superblocks were recently studied [16,19], and so were route directness, walking sheds, and betweenness centrality [17,18], the evaluation of route distance and diversity add new information to the ongoing research on superblocks. Further, current discussions about the use of superblocks in Chinese cities [62], as well as the study of the adaptation of Cerda's plan in Barcelona to accommodate sustainable mobility modes by aggregating several blocks into superblocks [63,64], highlight the ongoing need to better understand superblocks as a well-established urban planning strategy. The study of Middle Eastern cities' superblocks and their connectivity contributes to this global discussion.

In closing, it should finally be noted that the methods and results presented in this paper could already inform planning practice. More specifically, decisions regarding which design to adopt could be more sharply addressed by considering their connectivity and walkability, as well as the costs associated with their construction. Further, the metrics and methods used in this study could be easily replicated as they were performed using standard planning software, such as geographic information systems (GIS) and computer aided design (CAD) systems.

## 6. Conclusions

Cities in the United Arab Emirates, as well as numerous other cities in the Gulf Cooperation Council (GCC), adopted superblocks as the backbone of their urban development strategy. Built and planned following modernist principles and in a context of increased motorization, they provided a solution to the fast-paced urbanization needs that these cities faced. In Abu Dhabi, superblocks have historically supported the city's growth and expansion and, notably, they continue to do so today. This is particularly the case in the city suburbs, which are often built through the aggregation of numerous identical superblocks. However, the current and pressing need to reduce energy consumption and greenhouse gas emissions, along with renewed notions about the role that urban form plays in building more sustainable cities, calls for a re-examination of this enduring approach to urban planning and development.

This paper contributed to this task by looking at the connectivity of superblocks' street networks and their ability to accommodate walking trips. Findings indicate that distances are walkable, and routes often direct—particularly to corners—in all the studied designs. In contrast, the availability of alternative paths differed across the sample, and was found to be linked to the total road length of the different street systems. If cost minimization prevails in the decision-making process, network designs with the least amount of streets could be favored when planning new neighborhoods. Walking distances and directness would not be greatly affected if this alternative is preferred. A more involved cost–benefit approach, on the other hand, would be appropriate if path diversity is considered. In this latter case, the benefits could be associated to the concepts quite sharply outlined, long

ago, by Jane Jacobs, such as the fluidity of use and the mixing of paths that dense street networks with small blocks support.

Clearly, it is not through street network design alone that vibrant streets and walkable communities and cities are shaped. Key urban form variables, such as density and land use diversity and mix, would also need to be carefully evaluated when planning cities where walking, and eventually cycling, are viable transportation options. Cultural and climatic factors should, as well, be carefully considered. However, despite the noted challenges, it is worth noting that Abu Dhabi's community planning guidelines, as well as its sustainability rating framework Estidama, currently call for a transition towards more livable communities where walking, cycling, and public transportation use are supported by well-connected street networks [52,65,66]. The results of this paper are thus expected to contribute to both research and practice on walkability and sustainable mobility, in Abu Dhabi in particular, and in superblock-planned cities in general.

**Author Contributions:** Conceptualization, M.S.; methodology, M.S.; software, M.S. and R.A.; validation, M.S. and R.A.; formal analysis, M.S. and R.A.; investigation, M.S. and R.A.; resources, M.S. and R.A.; data curation, M.S.; writing—original draft preparation, M.S.; writing—review and editing, M.S.; visualization, M.S. and R.A.; supervision, M.S.; project administration, M.S.; funding acquisition, M.S. All authors have read and agreed to the published version of the manuscript.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

# References

1. Burton, E.; Jenks, M.; Williams, K. *Achieving Sustainable Urban Form*; Routledge: London, UK, 2013.
2. Echenique, M.H.; Hargreaves, A.J.; Mitchell, G.; Namdeo, A. Growing cities sustainably: Does urban form really matter? *J. Am. Plan. Assoc.* **2012**, *78*, 121–137. [CrossRef]
3. Newman, P.W.G.; Kenworthy, J.R. Gasoline consumption and cities: A comparison of US cities with a global survey. *J. Am. Plan. Assoc.* **1989**, *55*, 24–37. [CrossRef]
4. Ewing, R.; Bartholomew, K.; Winkelman, S.; Walters, J.; Chen, D.; McCann, B.; Goldberg, D. *Growing Cooler: The Evidence on Urban Development and Climate Change*; Urban Land Institute: Washington, DC, USA, 1997.
5. Jabareen, Y. Sustainable Urban Forms. *J. Plan. Educ. Res.* **2006**, *26*, 38–52. [CrossRef]
6. Cervero, R.; Kockelman, K. Travel demand and the 3Ds: Density, diversity, and design. *Transp. Res. Part D Transp. Environ.* **1997**, *2*, 199–219. [CrossRef]
7. Hamidi, S.; Ewing, R.; Preuss, I.; Dodds, A. Measuring Sprawl and Its Impacts: An Update. *J. Plan. Educ. Res.* **2015**, *35*, 35–50. [CrossRef]
8. Moudon, A.V.; Hess, P.M.; Snyder, M.C.; Stanilov, K. Effects of Site Design on Pedestrian Travel in Mixed-Use, Medium-Density Environments. *Transp. Res. Rec.* **1997**, *1578*, 48–55. [CrossRef]
9. Saelens, B.E.; Handy, S.L. Built environment correlates of walking: A review. *Med. Sci. Sports Exerc.* **2008**, *40*, S550–S566. [CrossRef]
10. Saelens, B.E.; Sallis, J.F.; Frank, L.D. Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Ann. Behav. Med.* **2003**, *25*, 80–91. [CrossRef]
11. Ewing, R.; Cervero, R. Travel and the Built Environment. *J. Am. Plan. Assoc.* **2010**, *76*, 265–294. [CrossRef]
12. Dill, J. Measuring Network Connectivity for Bicycling and Walking. In Proceedings of the 83rd Annual Meeting of the Transportation Research Board, Washington, DC, USA, 11–15 January 2004; pp. 11–15.
13. Frank, L.D.; Sallis, J.F.; Saelens, B.E.; Leary, L.; Cain, K.; Conway, T.L.; Hess, P.M. The development of a walkability index: Application to the Neighborhood Quality of Life Study. *Br. J. Sports Med.* **2010**, *44*, 924–933. [CrossRef]
14. Koohsari, M.J.; Sugiyama, T.; Mavoa, S.; Villanueva, K.; Badland, H.; Giles-Corti, B.; Owen, N. Street network measures and adults' walking for transport: Application of space syntax. *Health Place* **2016**, *38*, 89–95. [CrossRef]
15. Feng, C.; Peponis, J. Pathways to creating differentiated grids: Types, benefits and costs. *Environ. Plan. B Urban Anal. City Sci.* **2021**. [CrossRef]
16. Feng, C.; Peponis, J. The definition of syntactic types: The generation, analysis, and sorting of universes of superblock designs. *Environ. Plan. B Urban Anal. City Sci.* **2020**, *47*, 1031–1046. [CrossRef]

17. Scoppa, M.; Bawazir, K.; Alawadi, K. Walking the superblocks: Street layout efficiency and the sikkak system in Abu Dhabi. *Sustain. Cities Soc.* **2018**, *38*, 359–369. [CrossRef]
18. Scoppa, M.; Bawazir, K.; Alawadi, K. Straddling boundaries in superblock cities. Assessing local and global network connectivity using cases from Abu Dhabi, UAE. *Transp. Res. Part A Policy Pract.* **2019**, *130*, 770–782. [CrossRef]
19. Peponis, J.; Feng, C.; Green, D.; Haynie, D.; Kim, S.H.; Sheng, Q.; Vialard, A.; Wang, H. Syntax and parametric analysis of superblock patterns. *J. Space Syntax* **2015**, *6*, 109–141.
20. Perry, C.A. City planning for neighborhood life. *Soc. Forces* **1929**, *8*, 98–100. [CrossRef]
21. Mehaffy, M.W.; Porta, S.; Romice, O. The "neighborhood unit" on trial: A case study in the impacts of urban morphology. *J. Urban. Int. Res. Placemaking Urban Sustain.* **2015**, *8*, 199–217. [CrossRef]
22. Patricios, N.N. The neighborhood concept: A retrospective of physical design and social interaction. *J. Archit. Plan. Res.* **2002**, 70–90.
23. Southworth, M.; Ben-Joseph, E. *Streets and the Shaping of Towns and Cities*; Island Press: Washington, DC, USA, 2013.
24. Stigell, E.; Schantz, P. Methods for determining route distances in active commuting—Their validity and reproducibility. *J. Transp. Geogr.* **2011**, *19*, 563–574. [CrossRef]
25. Berrigan, D.; Pickle, L.W.; Dill, J. Associations between street connectivity and active transportation. *Int. J. Health Geogr.* **2010**, *9*, 1–18. [CrossRef] [PubMed]
26. Handy, S.; Paterson, R.G.; Butler, K.S. *Planning for Street Connectivity: Getting from Here to There*; APA Planning Advisory Service Reports; APA Planning Advisory Service: Chicago, IL, USA, 2003; pp. 1–75.
27. Jacobs, J. *The Death and Life of Great American Cities*; Random House: New York, NY, USA, 1961; 458p.
28. Abdo, R.; Batzel, G.M. Planning the modern Arab city: The case of Abu Dhabi. In *Reconsidering Jane Jacobs*; Routledge: Chicago, IL, USA, 2011; pp. 105–118.
29. Davidson, C.M.; Vanderwalle, D. *Abu Dhabi: Oil and Beyond*; Columbia University Press: New York, NY, USA, 2009.
30. Hashim, A.R.B. *Planning Abu Dhabi: An Urban History*; Routledge: London, UK, 2018.
31. Alawadi, K.; Benkraouda, O. What happened to Abu Dhabi's urbanism? The question of regional integration. *J. Urban Des.* **2018**, *23*, 367–394. [CrossRef]
32. Marshall, W.E.; Garrick, N.W. Community design and how much we drive. *J. Transp. Land Use* **2012**, *5*, 5–20. [CrossRef]
33. Stangl, P.; Guinn, J.M. Neighborhood design, connectivity assessment and obstruction. *Urban Des. Int.* **2011**, *16*, 285–296. [CrossRef]
34. Peponis, J.; Bafna, S.; Zhang, Z. The Connectivity of Streets: Reach and Directional Distance. *Environ. Plan. B Plan. Des.* **2008**, *35*, 881–901. [CrossRef]
35. Handy, S. Methodologies for exploring the link between urban form and travel behavior. *Transp. Res. Part D Transp. Environ.* **1996**, *1*, 151–165. [CrossRef]
36. Steiner, R.L. Traditional Neighborhood Shopping Districts: Patterns of Use and Modes of Access. Ph.D. Thesis, University of California, Berkeley, CA, USA, 1996.
37. Handy, S.L. Regional versus local accessibility: Neo-traditional development and its implications for non-work travel. *Built Environ.* **1992**, *18*, 253–267.
38. Hillier, B. *A Configurational Theory of Architecture*; Cambridge University Press: Cambridge, UK, 1996.
39. Hillier, B.; Hanson, J. *The Social Logic of Space*; Cambridge University Press: Cambridge, UK, 1989.
40. Porta, S.; Crucitti, P.; Latora, V. The Network Analysis of Urban Streets: A Primal Approach. *Environ. Plan. B Plan. Des.* **2006**, *33*, 705–725. [CrossRef]
41. Crucitti, P.; Latora, V.; Porta, S. Centrality measures in spatial networks of urban streets. *Phys. Rev. E-Stat. Nonlinear Soft Matter Phys.* **2006**, *73*. [CrossRef]
42. Scoppa, M.D.; Peponis, J. Distributed attraction: The effects of street network connectivity upon the distribution of retail frontage in the city of buenos aires. *Environ. Plan. B Plan. Des.* **2015**, *42*, 354–378. [CrossRef]
43. Wang, F.; Antipova, A.; Porta, S. Street centrality and land use intensity in Baton Rouge, Louisiana. *J. Transp. Geogr.* **2011**, *19*, 285–293. [CrossRef]
44. Porta, S.; Latora, V.; Wang, F.; Rueda, S.; Strano, E.; Scellato, S.; Cardillo, A.; Belli, E.; Cardenas, F.; Cormenzana, B. Street centrality and the location of economic activities in Barcelona. *Urban Stud.* **2012**, *49*, 1471–1488. [CrossRef]
45. Hillier, B.; Penn, A. Dense civilisations: The shape of cities in the 21st century. *Appl. Energy* **1992**, *43*, 41–66. [CrossRef]
46. Hess, P.; Moudon, A.; Snyder, M.; Stanilov, K. Site Design and Pedestrian Travel. *Transp. Res. Rec.* **1999**, *1674*, 9–19. [CrossRef]
47. Song, Y.; Knaap, G.J. Measuring urban form: Is portland winning the war on sprawl? *J. Am. Plan. Assoc.* **2004**, *70*, 210–225. [CrossRef]
48. Moudon, A.V.; Lee, C.; Cheadle, A.D.; Garvin, C.; Johnson, D.; Schmid, T.L.; Weathers, R.D.; Lin, L. Operational definitions of walkable neighborhood: Theoretical and empirical insights. *J. Phys. Act. Health* **2006**, *3*, S99–S117. [CrossRef]
49. Lee, C.; Moudon, A.V. The 3Ds+ R: Quantifying land use and urban form correlates of walking. *Transp. Res. Part D Transp. Environ.* **2006**, *11*, 204–215. [CrossRef]
50. Hess, P.M. Measures of connectivity [Streets: Old paradigm, new investment]. *Places* **1997**, *11*, 137.
51. Dijkstra, E.W. A note on two problems in connexion with graphs. *Numer. Math.* **1959**, *1*, 269–271. [CrossRef]

52. Abu Dhabi Urban Planning Council. *The Pearl Rating System for Estidama: Community Rating System*; Abu Dhabi Urban Planning Council: Abu Dhabi, United Arab Emirates, 2010.
53. Stangl, P. The pedestrian route directness test: A new level-of-service model. *Urban Des. Int.* **2012**, *17*, 228–238. [CrossRef]
54. Stangl, P. Overcoming flaws in permeability measures: Modified route directness. *J. Urban. Int. Res. Placemaking Urban Sustain.* **2017**, *12*, 1–14. [CrossRef]
55. Abu Dhabi Urban Planning Council. The Pearl Community Rating System. 2010. Available online: https://pages.dmt.gov.ae/-/media/DE1617B2A0634AC58B42DB511E18ECF4.ashx?newTab=1 (accessed on 25 May 2021).
56. Sevtsuk, A. Analysis and Planning of Urban Networks. In *Encyclopedia of Social Network Analysis and Mining*; Springer: New York, NY, USA, 2014.
57. Sevtsuk, A.; Mekonnen, M.; Kalvo, R.; Amindarbari, R. (Eds.) Redundant Paths for Urban Network Analysis. Presented at the 2014 ESRI GeoDesign Summit, Redlands, CA, USA, 29–30 January 2014.
58. Immers, B.; Bleukx, A.; Stada, J.; Tampere, C.; Yperman, I. Robustness and resilience of road network structures. In Proceedings of the NECTAR Cluster Meeting on Reliability of Networks, Amsterdam, The Netherlands, 19–20 March 2004.
59. Javanbarg, M.B.; Takada, S. Redundancy model for water supply systems under earthquake environments. In Proceedings of the 5th International Conference on Seismology and Earthquake Engineering, Tehran, Iran, 13–16 May 2007.
60. Sevtsuk, A. *Urban Network Analysis. Tools for Modeling Pedestrian and Bicycle Trips in Cities*; Harvard Graduate School of Design: Cambridge, MA, USA, 2018.
61. El-Geneidy, A.; Grimsrud, M.; Wasfi, R.; Tétreault, P.; Surprenant-Legault, J. New evidence on walking distances to transit stops: Identifying redundancies and gaps using variable service areas. *Transportation* **2014**, *41*, 193–210. [CrossRef]
62. Kan, H.Y.; Forsyth, A.; Rowe, P. Redesigning China's superblock neighbourhoods: Policies, opportunities and challenges. *J. Urban Des.* **2017**, *22*, 757–777. [CrossRef]
63. Scudellari, J.; Staricco, L.; Vitale Brovarone, E. Implementing the Supermanzana approach in Barcelona. Critical issues at local and urban level. *J. Urban Des.* **2020**, *25*, 675–696. [CrossRef]
64. Brovarone, E.V.; Scudellari, J.; Staricco, L. Planning the transition to autonomous driving: A policy pathway towards urban liveability. *Cities* **2021**, *108*, 102996. [CrossRef]
65. Abu Dhabi Urban Planning Council. *Abu Dhabi Community Facility Planning Standards*; Abu Dhabi Urban Planning Council: Abu Dhabi, United Arab Emirates, 2020.
66. Abu Dhabi Urban Planning Council. *Abu Dhabi Urban Street Design Manual*; Abu Dhabi Urban Planning Council: Abu Dhabi, United Arab Emirates, 2014.

# Hybrid Bayesian Network Models to Investigate the Impact of Built Environment Experience before Adulthood on Students' Tolerable Travel Time to Campus: Towards Sustainable Commute Behavior

**Yu Chen** [1,2,*], **Mahdi Aghaabbasi** [3,*], **Mujahid Ali** [4,*], **Sergey Anciferov** [5], **Linar Sabitov** [6,7], **Sergey Chebotarev** [8], **Karina Nabiullina** [6], **Evgeny Sychev** [5], **Roman Fediuk** [9,10] and **Rosilawati Zainol** [3]

[1] School of Architecture, Hunan University, Changsha 410012, China
[2] School of Architecture and Urban Planning, Hunan City University, Yiyang 413002, China
[3] Centre for Sustainable Urban Planning and Real Estate (SUPRE), Department of Urban and Regional Planning, Faculty of Built Environment, University of Malaya, Kuala Lumpur 50603, Malaysia; rosilawatizai@um.edu.my
[4] Department of Civil and Environmental Engineering, Universiti Teknologi Petronas, Seri Iskandar 32610, Malaysia
[5] Department of Mechanical Equipment, Belgorod State Technological University Named after V.G. Shukhov, 308012 Belgorod, Russia; anciferov.sergey@gmail.com (S.A.); evgeniy.sychov.015@gmail.com (E.S.)
[6] Department of Structural and Design Engineering, Kazan Federal University, 420008 Kazan, Russia; sabitov-kgasu@mail.ru (L.S.); karina.nabiullina@kpfu.ru (K.N.)
[7] Department of Energy Supply of Enterprises, Construction of Buildings and Structures, Kazan State Power Engineering University, 420066 Kazan, Russia
[8] Moscow State University of Technologies and Management (FCU), 109004 Moscow, Russia; chebotarev.sergei@gmail.com
[9] Far Eastern Federal University, 690922 Vladivostok, Russia; fedyuk.rs@dvfu.ru
[10] Peter the Great St. Petersburg Polytechnic University, 195251 St. Petersburg, Russia
[*] Correspondence: chenyu2020123@163.com (Y.C.); mahdi@um.edu.my (M.A.); mujahid_19001704@utp.edu.my (M.A.)

**Abstract:** This present study developed two predictive and associative Bayesian network models to forecast the tolerable travel time of university students to campus. This study considered the built environment experiences of university students during their early life-course as the main predictors of this study. The Bayesian network models were hybridized with the Pearson chi-square test to select the most relevant variables to predict the tolerable travel time. Two predictive models were developed. The first model was applied only to the variables of the built environment, while the second model was applied to all variables that were identified using the Pearson chi-square tests. The results showed that most students were inclined to choose the tolerable travel time of 0–20 min. Among the built environment predictors, the availability of residential buildings in the neighborhood in the age periods of 14–18 was the most important. Taking all the variables into account, distance from students' homes to campuses was the most important. The findings of this research imply that the built environment experiences of people during their early life-course may affect their future travel behaviors and tolerance. Besides, the outcome of this study can help planners create more sustainable commute behaviors among people in the future by building more compact and mixed-use neighborhoods.

**Keywords:** tolerable travel time; university students; built environment; early life-course; Bayesian network; machine learning

## 1. Introduction

Travel time (TT) is viewed as a necessary university-related activity and functions as a link between home and university campus. For each student, travel to campus differs

in distance and complexity. This complexity may be increased if certain activities which link the travel and family are incorporated (e.g., the school operated or residential location decisions when spouses in households pursue careers) according to Wheatley [1]. This travel time spent can be regarded as both "productive" and a "waste of time".

Several studies identified the associations between duration of travel and individuals' well-being, including stress, comfort and satisfaction, and health [2–5]. In addition, several studies assessed the relationship between the TT and all daily activities and work duration [6,7]. Many factors, including sociodemographic, household characteristics, and travel mode, may influence TT [8]. In addition, many academic studies analyzed the reciprocity between built environment (BE) attributes and the TT [9–14].

As regards to the university students, there are some studies which considered TT as a function of students' commute mode choice [15–19]. However, no available study has ever assessed the influence of BE factors on the students' TT, and also, no study has ever considered the tolerable travel time (TTT) of university students considering the effects of BE variables. While a sizeable number of literature considered the effects of BE attributes on TT of the general population, studies on university students that exhibited different travel behaviors from the general population are still lacking [20–22].

The concept of TTT was developed by Milakis et al. [23]. This concept was established based on various theories related to commuting time, which include satisfaction [24,25], consideration sets [26], the travel time budget [27,28], and ideal travel time [29]. Milakis, Cervero, Van Wee, and Maat [23] employed semi-structured interviews to explore the primary characteristics of acceptable travel time (ATT). The study supported the validity of the concept of ATT through their findings and showed that the ATT may be varied for people with different sociodemographic attributes and travel modes. According to this concept, people presumably consider an ATT in their trips and decision-making processes regarding destinations. This concept views ATT as a behavioral threshold that is defined by the process of utilitarianism (i.e., intrinsic and derived utility). Intrinsic utility refers to the travel-related advantages (or disadvantages), while the advantages concerning activity at a journey destination are referred to as the derived utility. The concept splits the timeline of a one-way trip into three main periods in terms of total utility changes: (1) growth, (2) tolerance, and (3) decay. In the growth phase, both intrinsic and derived utility witnessed total utility increase. In the tolerance phase, the total utility yet increases, but slower than before until it touches the ATT (maximum level). Compared to the growth period, intrinsic utility is reduced and derived utility rises, but at a slower rate. Eventually, in the decay period, the total utility decreases because of the rapid decrease in intrinsic utility coupled with slow growth in derived utility.

The TTT, in fact, is the duration between the ideal travel time and ATT. In simple terms, TTT refers to the maximum amount of one-way TT that an individual tolerates [30]. If the actual TTs of a commuter reach or exceed the tolerance thresholds, the commuter is keen to decrease his/her travel time by making some changes, including, but not limited to, residential, job locations, or travel modes. The literature acknowledged the negative effects of exceeding the TTT thresholds. These impacts may be increasing stress levels, demanding excessive energy, and consuming time which may limit the time available for other daily activities [31–37].

There has been growing acknowledgement that travel behaviors are habitual [38–40] and these behaviors may become debilitated when disturbed by a contextual adjustment [39]. These contexts may comprise the environment where behavior occurs, such as social, physical, spatial, and time cues. Moreover, major life events (e.g., change in employment) may change the travel behaviors of individuals over time [41]. To date, several studies have examined the impacts of changes in life events and residential locations on the travel behavior of individuals [42–47]. However, many of these studies focused on predicting the travel mode choice, and many other aspects of travel outcomes, such as TTT, were overlooked. Furthermore, other phenomena that occurred to individuals in the past have received less attention. For example, no study has considered the associations between

adults' travel behavior outcomes and their living environments and BE experiences during childhood and adolescence. Several studies pointed out that previous living environments of people may influence their future behaviors that are related to commuting, such as their adaption and tolerance of crowding or their concern over the environment [48–50]. More importantly, lifetime habits, such as physically active lifestyle, can be developed during the early childhood years [51]. Thus, there may be a relationship between the BE experiences and the early life-course of people and their future travel behaviors, such as their TTT.

The aim of this study is to identify associations between the childhood BE experiences of university students and their current TTT to campus. This investigation extends the literature in two main ways. To begin with, it adds to the growing body of knowledge about tolerable travel time in developing countries. Second, this study evaluates the significance of different built environments (during childhood and now) and sociodemographic factors in determining students' tolerable travel time to campus. It also shows that childhood built environment experiences have associations with the students' tolerable travel time to campus, corroborating the sparse data in the literature.

## 2. Knowledge Gaps and Research Questions

While some non-academic reports on average commute time of employees to work are available in Malaysia [52], no academic study has considered the average or tolerable travel time of students living off-campus to the universities' campus. Therefore, this present study endeavors to identify what factors of travel time resolutely affect the TTT of university students to their campuses.

Among university students, off-campus students typically experience various mobility challenges, including travel between home and campus, as well as trips linked to non-study activities [19]. For example, off-campus students may require more commute time for campus-related trips than their on-campus peers. Alternatively, these students can use this prolonged commute time to study and develop networks and social bonds. Moreover, these students usually face challenges in finding suitable travel alternatives (on the condition that their car/motorcycle is unavailable) for attending sessions programmed for the early hours of morning, late hours of night, or days other than working days. So far, only a few investigations have exclusively appraised the commute patterns of off-campus university students and examined difficulties connected to the transportation they encountered [19,53].

The literature review also provided evidence that people who experience life events are more inclined to travel behavior alterations. Past research on life events and travel behavior alternations have mostly focused on a particular or restricted variety of life experiences. Conversely, and to the best of the authors' knowledge, no study has examined the influence of built environment experiences at the early life-course of the general population and specific populations (such as university students) on their future travel behaviors, particularly TTT. Therefore, the investigation conducted in the following sections attempted to discuss three principal research questions:

1.  What is the most probable TTT of off-campus university students to the campus?
2.  To what extent is off-campus university students' TTT to the campus associated with BE experiences during their childhood and adolescence?
3.  How are sociodemographic, household, residential, and travel mode characteristics associated with off-campus university students' TTT to the campus?

The collection of retrospective data from two universities in Malaysia is used in this study to address these questions using a two-step analysis structure. The details of data collection and analysis are discussed in the subsequent sections.

## 3. Research Design

This study adopted a retrospective research design. According to Behrens and Mistro [54], this design involves one-time surveys of people and asks participants to remember experiences or events that previously happened to them. The respondents for this present study are off-campus university students that were surveyed and asked to recall their

living environments during the age periods of 1–6, 7–13, and 14–18. The retrospective surveys are suitable for observations over long time spans. The literature suggests that the respondents can remember main life-process experiences and can also describe any of their essential characteristics, which enables the assessment of general alternations over more prolonged periods.

This present study evaluated the influence of BE experiences during childhood and adolescence. Using this design, this study thus examined the impact of BE experiences during childhood and adolescence on university students' TTT to campus. van de Coevering et al. [55] pointed out that the principal disadvantage of the retrospective design is that the examination of opinions and specification of everyday travel behavior are misleading. The authors thus adopted a comparably short time span and urged university students to show preferences for their tolerable travel time on a nominal measure. The survey particularly inquired about current inclinations and regarded these trends as steady throughout university time. The critical role of control variables on the study of BE and travel behavior is undeniable and these variables cannot be eliminated from the modelling procedure [55,56]. Therefore, this study has considered the effects of these variables in the second series of models to obtain a more rigorous research design. The possible effects of different influential factors on university students' TTT to campus are presented in Figure 1.
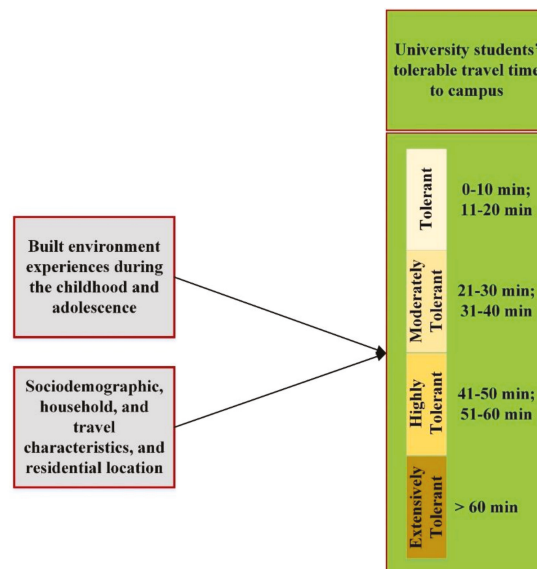


**Figure 1.** Schematic diagram of possible factors influencing university students' tolerable travel time to campus and the classifications of tolerable travel time used in this study.

### 3.1. Variables of Built Environment during the Early Life-Course of People

The impact of the built environment on university students' TTT to campus was investigated through the "5Ds" model. Initially, Cervero and Kockelman [57] developed the "3Ds" model which included density, diversity, and design to express the urban structure. Subsequently, Ewing et al. [58] combined two more dimensions, including destination accessibility and distance to transit, with the previous model and developed the "5Ds" model. The magnitude of land use for residence, work, and other goals is regarded as density. Diversity relates to the level of heterogeneity of land use. The properties of the street network and the walking environment quality are viewed as the design. Distance to transit refers to the accessibility to public transportation facilities. Finally, the measurement of ease of access to trip attractions is referred to as destination accessibility.

### 3.2. Survey and Data Collection

This present study used an online questionnaire survey in March and May 2020 to collect data regarding the TTT of off-campus university students in Malaysia. In comparison with paper-based questionnaires, the online option is more comfortable to complete by the respondents, without any geographical restrictions. This advantage of the online survey makes it a suitable instrument for studies which try to collect data in multiple locations during times in which movements are restricted (e.g., lockdown and quarantine). The respondents of this study were mainly from two public universities in two renowned tertiary education cities. The universities are A and B (for the sake of the blind review process, the case studies are removed from this manuscript). An email was sent to the students' email account in each university, which explained the aims of the study. Besides, the research team included the internet address of the questionnaire in the email. A reminder email was also sent to the students every two weeks to increase the response rate and balance the sample size.

The questionnaire comprised three main sections. The initial section examined the respondents' sociodemographic and household characteristics. The second part involved some questions regarding current residential location and the usual travel mode to the campus. The third part asked students to recall their living environment during two periods of age, namely 7–13 and 14–18. This part also assessed the attitudes of respondents towards their living environment during the mentioned age periods in the form of Likert scale measurement. Once the questionnaire was designed, the research team sent a full version of the questionnaire to a panel of experts, which included urban planners and transport planners. The panel was urged to give their feedback regarding the suitability and communicability of the questionnaire. Likewise, the panel was deemed fit to modify, add, or remove any item from the questionnaire. Minor changes were made to the questionnaire as a result of the experts' consultation. For instance, the time and distance scales for TTT and tolerable travel distance have become finer to avoid difficulties from the extremely large discretization of travel distance and TT.

Following the panel review, the research team conducted a pilot survey and collected 33 completed questionnaires. The research team also asked the respondents to express any difficulties or incommunicability they found in filling the questionnaire. This pilot survey resulted in some changes in the questionnaire. The main change was made to the age scale. Before the pilot study, the attitudes of the students towards their living environment were supposed to be assessed using three age periods, which are 1–6, 7–13, and 14–18. However, the respondents' feedback indicated that it was difficult for them to recall their living environment during the age period of 1–6. Besides, the primary analysis result also showed that responses related to this age range were inconsistent. Thus, the age period of 1–6 was removed from the age scale for all survey items, except the question which asked the respondents to indicate the type of settlement (city, village, and suburb) in which they have lived. Age scale of this question has not been changed because it was easy for the respondents to recall general rather than specific characteristics of the living environment. Consequently, the final version of the questionnaire included 49 questions. The questionnaire items that were utilized in this present research are presented in Table 1.

### 3.3. Analysis Approaches and Techniques

Multiple traditional statistical methods, including the multinomial logit, binary logit, and mixed logit models were frequently employed in studies related to transport for analyzing predictors of the university students' travel behaviors, particularly their mode choice [59–62]. The data related to travel behaviors are generally bulky and complicated, which makes the use of regression models challenging for studying predictors of the travel behaviors and patterns. These models typically assume that the associations between the variables are linear and consider the data without outliers [63–65]. However, these assumptions are hardly adequate for travel behavior data. Another daunting task, which can occur in regression models, is using the cross-product terms for distinguishing the

predictors because of interaction that happens in complicated configurations [66]. Moreover, according to Karlaftis and Golias [67] and Yan, Richards, and Su [66], regression models are often unable to efficiently handle differing categorical variables.

**Table 1.** Variables employed in this study.

| Variable | Description | Value |
|---|---|---|
| *Sociodemographic and household characteristics* | | |
| AGE | Respondent's age | (1) 19–24; (2) 25–30; (3) 31–36; (4) 37–42; (5) 43–48; (6) more than 48 |
| GEN | Respondent's gender | (1) male; (2) female |
| EDU | Highest education level of respondent | (1) primary; (2) secondary; (3) diploma; (4) bachelor's degree; (5) master's degree; (6) doctorate degree |
| HHCO | Count of household members | 1–9 |
| CHCO | Number of children in household | 0–4 |
| INC | Household income | (1) less than MYR 1000; (2) between MYR 1000 and MYR 2000; (3) between MYR 2000 and MYR 3000; (4) between MYR 3000 and MYR 6000; (5) between MYR 6000 and MYR 13,000: (6) more than MYR 13,000 |
| RACE | Respondent's race | (1) Malay; (2) Chinese; (3) Indian; (4) foreigner |
| PRVE | Vehicle ownership | (1) yes; (2) no |
| VECO | Count of household vehicles | 0–7 |
| *Residential and travel mode characteristics* | | |
| NETYCH | Description of the neighborhood in terms of type and characteristic | (1) residential only; (2) residential with some commercial buildings; (3) residential with some industrial facilities; (4) a commercial area with some residential; (5) an industrial area with some residential; (6) mixed residential and commercial |
| REHOMELOC | Top first reason for choosing current home location | (1) Cost/price of home; (2) home size and characteristics; (3) neighborhood characteristics; (4) home or lot size; (5) school district/system; (6) convenient for work; (7) convenient for school; (8) convenient for retail (shopping, entertainment, restaurants); (9) close to friends and family; (10) close to public transportation; (11) close to scenic locations (beach, lake, golf courses); (12) less traffic to school; (13) no other choices apply |
| UTMS | Usual travel mode to campus | (1) Private car; (2) private motorcycle; (3) public transportation; (4) walking/cycling; (5) metered taxi; (6) ride-sourcing |
| DISSC | Distance from home to campus | (1) 0–10 km; (2) 11–20 km; (3) 21–30 km; (4) 31–40 km; (5) 41–50 km; (6) 51–60 km; (7) more than 60 km |
| ADISSC | Acceptable distance from home to campus | (1) 0–10 km; (2) 11–20 km; (3) 21–30 km; (4) 31–40 km; (5) 41–50 km; (6) 51–60 km; (7) more than 60 km |
| *Living environment during childhood and adolescence* | | |
| KSETTLE | Settlement type during the age periods of 1–6, 7–13, and 14–18 | (1) city; (2) village; (3) suburb |
| PCIVILRS | Perception towards the size of settlement during the age periods of 7–13 and 14–18 | (1) very small; (2) small; (3) medium; (4) large; (5) very large |
| TPHOUSE | Type of house during the age periods of 7–13 and 14–18 | (1) bungalow; (2) detached/semi-detached; (3) shop houses; (4) flat (non-gated); (5) apartment (gated); (6) condominium (high rises) |
| *Living environment during childhood and adolescence—density* | | |
| 1DENSITY | The neighborhood I lived in had many shop lots in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 2DENSITY | The neighborhood I lived in had many offices in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |

**Table 1.** *Cont.*

| Variable | Description | Value |
|---|---|---|
| 3DENSITY | The neighborhood I lived in had many residential buildings in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 4DENSITY | The neighborhood I lived in had many entertainment facilities in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 5DENSITY | The neighborhood I lived in had many industrial facilities in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 6DENSITY | The neighborhood I lived in had some schools in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| Living environment during childhood and adolescence—diversity | | |
| 1DIVERSITY | My house was close to the shops in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 2DIVERSITY | My house was close to public offices in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 3DIVERSITY | My house was close to entertainment facilities in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 4DIVERSITY | My house was close to other residential buildings in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 5DIVERSITY | The school I attended was within walking distance of my house in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| Living environment during childhood and adolescence—design | | |
| 1DESIGN | The neighborhood I lived in had large block sizes in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 2DESIGN | The neighborhood I lived in had many intersections in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 3DESIGN | The neighborhood I lived in had a full sidewalk coverage along the street in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 4DESIGN | The neighborhood I lived in had many buildings that were set back from the sidewalks with an appropriate distance (there was a good distance between buildings and the sidewalks) in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 5DESIGN | The neighborhood I lived in had wide sidewalks in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 6DESIGN | The neighborhood I lived in had several pedestrian crossings in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 7DESIGN | The neighborhood I lived in had many trees and landscapes in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 8DESIGN | The neighborhood I lived in had many pedestrian-related facilities (e.g., water fountains and benches) in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |

**Table 1.** *Cont.*

| Variable | Description | Value |
|---|---|---|
| Living environment during childhood and adolescence—destination accessibility | | |
| 1ACCESSIBILITY | In the neighborhood I lived in, it was easy for me to access local stores in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 2ACCESSIBILITY | In the neighborhood I lived in, it was easy for me to access business districts in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 3ACCESSIBILITY | In the neighborhood I lived in, it was easy for me to access the primary/secondary school in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 4ACCESSIBILITY | In the neighborhood I lived in, it was easy for me to access the recreation facilities in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| Living environment during childhood and adolescence—distance to transit | | |
| 1DISTANCETOTRAN | In the neighborhood I lived in, my house was close to the bus stops in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 2DISTANCETOTRAN | In the neighborhood I lived in, my house was close to the taxi stops in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| 3DISTANCETOTRAN | My school was close to the taxi/bus stops in the age ranges of 7–13 and 14–18 | (1) strongly disagree; (2) disagree; (3) neutral; (4) agree; (5) strongly agree |
| Target variable | | |
| TTTTOSC | Tolerable travel time to campus | (1) 0–10 min; (2) 11–20 min; (3) 21–30 min; (4) 31–40 min; (5) 41–50 min; (6) 51–60 min; (7) more than 60 min |

To remedy the above shortcomings of regression models, this study employed non-parametric and machine learning (ML) techniques. These techniques refer to a procedure that makes use of preprocessing, input selection, and extraction and classification processes. The body of literature suggested that ML techniques such as Bayesian network (BN) are free of assumptions of variable distributions; thus, possessing prior probabilistic knowledge on university students' travel behavior and their TTT is not needed. The ML techniques are also effective in dealing with outliers and many categorical variables. Finally, these techniques efficiently extract knowledge from massive data [68–73]. Pearson chi-square test and BN have been successfully applied in a limited number of studies related to transport [74]. However, to the best of the authors' knowledge, no study has employed both Pearson chi-square test and BN in the study of the university students' travel behaviors and their TTT.

This present study used a two-step approach to analyze the data collected. The first step was to examine the association between the input variables and the target variable through Pearson chi-square tests. The variables with a value greater than 0.75 were selected as the most associated variables with the target variable and were selected to be used as the inputs of prediction models. Next, two BN models were developed to predict the university students' TTT to campus. While the first model was applied to those BE variables (during childhood and adolescence) that were selected in the input selection step, the second model was applied to all selected variables. Figure 2 shows the study process and framework.
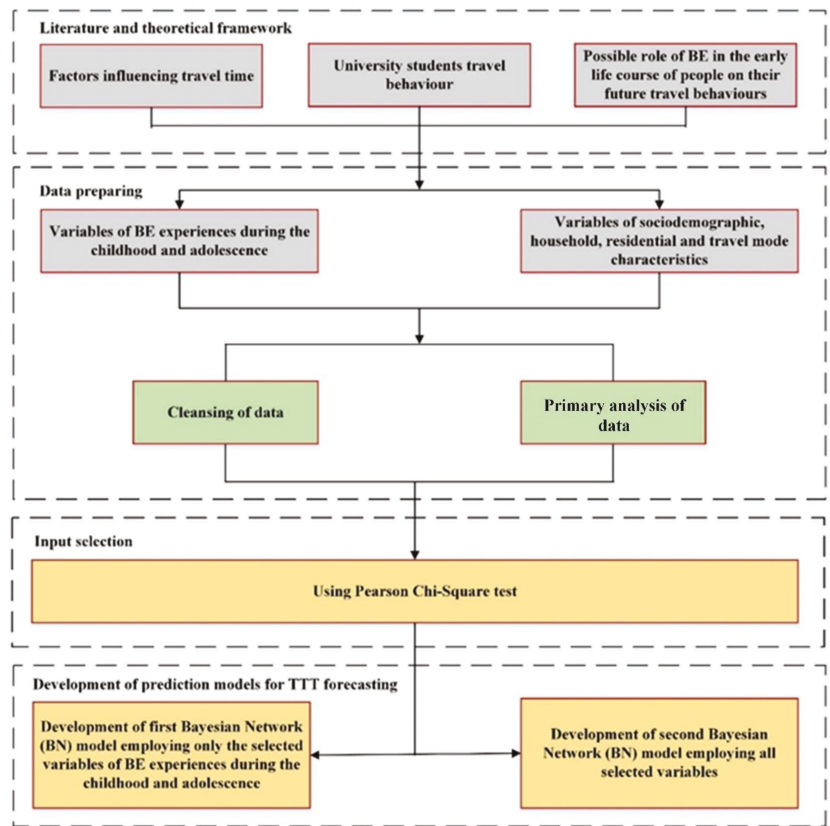
**Figure 2.** The framework of this study.

*3.4. Bayesian Network Model*

BN is a probabilistic network model that employs the probability theory and the graph system concurrently. The theory behind the BN analysis is the Bayesian probability. The analysis employs joint distributions and preceding distributions of each variable to measure a subsequent distribution for each variable of concern. Two principal parts of BN are probabilistic and graphical structures. A graph $K = (H, L)$ is defined by a collection of nodes $H = \{H_1, \ldots, H_p\}$ and a collection of edges $L \subseteq H \times H$. In a BN, the nodes $H$ denote the variables, and the edges $L$ signify the directed arrows, showing the conditional dependencies amongst these variables. Equation (1) manifests the probabilistic relationships between the nodes defined by a function of joint probability density $F(H)$.

$$F(H_1, \ldots, H_k) = \prod_{i=1}^{k} F(V_i | Parent(H_i)). \qquad (1)$$

The conditional probability tables reflect the aforementioned joint likelihood density function, developing the probabilistic BN composition. The BN graphical arrangement necessity possesses an acyclic character. In particular, a BN follows a directed acyclic graph formation. To be precise, there must not be any edge redirecting, including $H_i \rightarrow \ldots \rightarrow H_i$ for any $H_i$ and $H$. The edges reveal the mathematical dependencies among the nodes; however, the edge direction may not inevitably indicate a causality association. Between a pair of nodes linked together by an edge, the preceding and following nodes are named

parent and child, subsequently. To build the Bayesian networks, this study utilized the Markov blanket, which finds all the variables in the network that are essential to forecast the target variable. A simple structure of BN based on directed acyclic graph formation is shown in Figure 3.
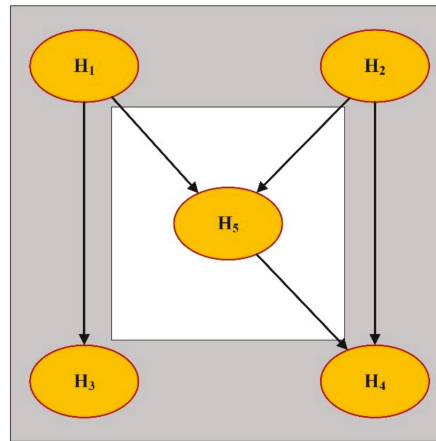


**Figure 3.** Simple BN network.

Generally, travel behavior datasets contain various parameters, and each parameter may have diverse classes. Besides, when new knowledge is accessible or needed, these datasets may remain constantly updated. Moreover, it is very common that travel behavior datasets are incomplete or possess missing values. Several studies acknowledged that the BN technique can deal with variables with various classes and undersampling data efficiently. Additionally, this technique can handle data that are deficient, fallacious, or dubious [75–77]. According to Tareeq and Inamura [78], the BN technique was considered proper to learn changeable behaviors (including the TTT under review) because it can effectively improve its network following the data specified or inserted into it.

The BN works excellently with a limited number of candidate variables [79]; thus, the Pearson chi-square tests were employed to reduce data dimensionality and select only the most relevant inputs. Pearson chi-square test is a non-parametric statistical test which is applied to sets of categorical data to assess how probable it was that any observed variation between the sets occurred by chance. This test is suitable for feature selection when the target variables of some inputs are categorical. Equation (2) shows the mathematical formulation of the Pearson chi-square test.

$$\chi^2 = \sum \frac{(A_r - A_e)^2}{A_e} \tag{2}$$

where the $A_r$ and $A_e$ are the real and expected frequencies of categories.

## 4. Results

This present study created a dataset that included 758 university students' travel data from two public universities in Malaysia. The dataset contains only the off-campus participants. As previously mentioned, this study aims to predict the tolerable travel time of the university students to the campus considering their past built environment experiences. On the TTT frequencies, 68.35% of students were tolerant, 3.69% were moderately tolerant, 18.99% were highly tolerant, and 8.97% were extensively tolerant. The age range of the majority of students was 19–24 (73.88%). This overrepresentation was believed to have stemmed from the fact that younger students were more capable of and interested in

participating in an online survey. Moreover, the older students might be involved in some family matters or might have had less free time, and thus, had much less time for filling in the online questionnaire. The study trends can be extrapolated to other university students because of the size and variety of this study. The sociodemographic characteristics and respondents' profiles are presented in Appendix A.

### 4.1. Input Selection

The associations of 74 input variables with the target variable (TTT to campus) were tested through the Pearson chi-square tests. This present study selected those variables with the value of 0.75 and above as the most associated and important inputs for predicting the students' TTT to campus. Thus, 38 input variables were selected (Table 2). Furthermore, these variables will be used to develop two predictive BN models. Among the total variables, distance from home to campus (DISSC) was the most important variable, while for variables of BE during the childhood and adolescence of students, the ease of access to the primary/secondary school in the age range of 7–13 (3ACCESSIBILITY713) was the most important variable and was followed by the ease of access to local stores in the same age period (1ACCESSIBILITY713).

**Table 2.** Input variables selected by the Pearson chi-square tests.

| Rank | Variable | Value | Rank | Variable | Value |
|------|----------|-------|------|----------|-------|
| 1 | DISSC | 1.00 | 20 | 2DISTANCETOTRAN713 | 0.91 |
| 2 | 3ACCESSIBILITY713 | 1.00 | 21 | HHCO | 0.90 |
| 3 | 1ACCESSIBILITY713 | 0.99 | 22 | 1DIVERSITY713 | 0.89 |
| 4 | 4DIVERSITY713 | 0.99 | 23 | PRVE | 0.87 |
| 5 | RACE | 0.98 | 24 | 5DESIGN1418 | 0.87 |
| 6 | 7DESIGN713 | 0.98 | 25 | 4DIVERSITY1418 | 0.86 |
| 7 | UTMWS | 0.98 | 26 | 7DESIGN1418 | 0.85 |
| 8 | 3ACCESSIBILITY1418 | 0.98 | 27 | 4DENSITY713 | 0.85 |
| 9 | 3DENSITY1418 | 0.98 | 28 | 2DIVERSITY1418 | 0.85 |
| 10 | 3DENSITY713 | 0.98 | 29 | KSETTLE16 | 0.85 |
| 11 | 1DIVERSITY1418 | 0.97 | 30 | 5DESIGN713 | 0.84 |
| 12 | AGE | 0.96 | 31 | 4ACCESSIBILITY713 | 0.81 |
| 13 | 6DESIGN713 | 0.94 | 32 | UTMTOSC1418 | 0.81 |
| 14 | GEN | 0.94 | 33 | UTMTOSC713 | 0.80 |
| 15 | 2DISTANCETOTRAN1418 | 0.93 | 34 | 1ACCESSIBILITY1418 | 0.80 |
| 16 | 4ACCESSIBILITY1418 | 0.93 | 35 | REHOMLOC | 0.79 |
| 17 | 6DENSITY713 | 0.93 | 36 | KSETTLE1418 | 0.79 |
| 18 | TPHOUSE713 | 0.92 | 37 | 2DIVERSITY713 | 0.78 |
| 19 | 6DENSITY1418 | 0.91 | 38 | EDU | 0.77 |

### 4.2. BN#1 Model Focusing on BE Attributes

The first BN model was developed using 27 BE variables that were chosen in the previous step. This model selected the 10 most important variables to predict the TTT of university students to campus. The training accuracy of this model was 97.47%. The BN#1 structure is presented in Figure 4. This diagram includes 11 variables, 10 predictors, and 1 target variable. The importance of each predictor is shown in Figure 5. As evidently shown, the availability of residential buildings in the neighborhood that respondents lived in, within the age period of 14–18, was the most critical predictor. This predictor was followed by the proximity of the house to shops in the age range of 14–18. The least essential predictor was the type of settlement in the age period of 14–18. As analytically revealed, settlement type in the age range of 1–6 was more critical than 14–18. From the age group perspective, settlement type was the only factor that was assessed by this study for the age range of 1–6. This predictor was selected as an essential predictor by the BN. For the age range of 7–13, four predictors were the most important, which are: (1) availability of residential buildings in the neighborhood, (2) availability of schools in the neighborhood, (3) availability of entertainment facilities in the neighborhood, and (4) proximity of the

house to shops. For the age range of 14–18, four predictors were the most important, which are: (1) availability of residential buildings in the neighborhood, (2) availability of schools in the neighborhood, (3) proximity of the house to shops, and (4) settlement type.
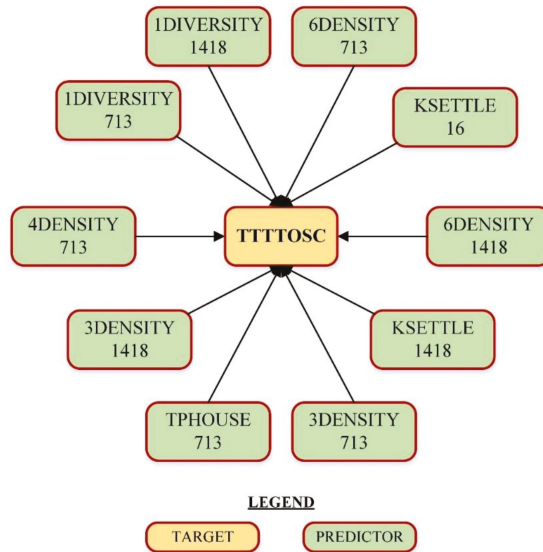


**Figure 4.** BN diagram to predict the TTT of university students to campus considering only the effects of BE variables.
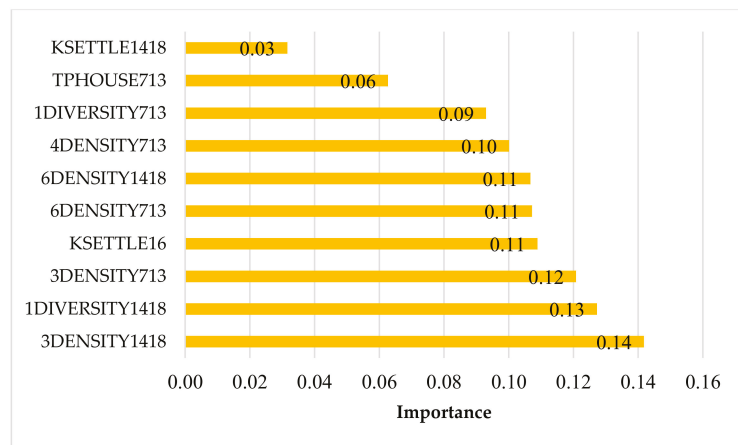


**Figure 5.** Importance of 14 variables to predict the TTT of students to campus considering only the BE features.

The BN#1 identified 76 conditional probabilities for each category of TTT, except TTT of 21–30 min. No TTT of 21–30 min was predicted by BN#1. To simplify the interpretation of the probabilities, only high probable TTTs (probability $\geq 0.75$) were reported for each category. The most frequent and influential value of each predictor that predicted each TTT is presented in Table 3.

**Table 3.** Conditional probabilities of high probable TTTs to campus derived from BN#1.

| Variable/Value | Frequency (%) | | | | | |
|---|---|---|---|---|---|---|
| | 0–10 | 11–20 | 31–40 | 41–50 | 51–60 | More than 60 |
| TPHOUSE713 | | | | | | |
| 1 | 10.3 | 40.7 | 0 | 16.7 | 40.0 | 40 |
| 2 | 75.9 | 33.3 | 50 | 50.0 | 60.0 | 60 |
| 3 | 0 | | 0 | 0 | 0 | 0 |
| 4 | 3.4 | 14.8 | 0 | 33.3 | 0 | 0 |
| 5 | 10.3 | 11.1 | 50 | 0 | 0 | 0 |
| 3DENSITY713 | | | | | | |
| 1 | 10.3 | 7.4 | 0 | 0 | 40.0 | 20.0 |
| 2 | 10.3 | 7.4 | 0 | 16.7 | 40.0 | 20.0 |
| 3 | 37.9 | 14.8 | 0 | 33.3 | 0 | 0 |
| 4 | 31.0 | 37.0 | 100 | 50.0 | 20.0 | 0 |
| 5 | 10.3 | 33.3 | 0 | 0 | 0 | 60.0 |
| 3DENSITY1418 | | | | | | |
| 1 | 10.3 | 7.4 | 0 | 0 | 20 | 20 |
| 2 | 6.9 | 11.1 | 0 | 0 | 60 | 0 |
| 3 | 27.6 | 11.1 | 0 | 33.3 | 0 | 0 |
| 4 | 41.4 | 44.4 | 100 | 66.7 | 0 | 20 |
| 5 | 13.8 | 25.9 | 0 | 0 | 20.0 | 60 |
| 4DENSITY713 | | | | | | |
| 1 | 13.8 | 18.5 | 0 | 16.7 | 0 | 0 |
| 2 | 44.8 | 25.9 | 0 | 83.3 | 40.0 | 60.0 |
| 3 | 31.0 | 25.9 | 50 | 0 | 20.0 | 40.0 |
| 4 | 10.3 | 29.6 | 50 | 0 | 40.0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6DENSITY713 | | | | | | |
| 1 | 0 | 0 | 0 | 16.7 | 0 | |
| 2 | 6.9 | 3.7 | 50 | 16.7 | 20.0 | 20.0 |
| 3 | 20.7 | 22.2 | 50 | 0 | 0 | 0 |
| 4 | 62.1 | 40.7 | 0 | 66.7 | 40.0 | 40.0 |
| 5 | 10.3 | 33.3 | 0 | 0 | 40.0 | 40.0 |
| 6DENSITY1418 | | | | | | |
| 1 | 0 | 0 | 50 | 16.7 | 0 | 0 |
| 2 | 6.9 | 3.7 | 0 | 16.7 | 20.0 | 20.0 |
| 3 | 17.2 | 18.5 | 50 | 0 | 0 | 0 |
| 4 | 62.1 | 51.9 | 0 | 50 | 40.0 | 40.0 |
| 5 | 13.8 | 25.9 | 0 | 16.7 | 40.0 | 40.0 |
| 1DIVERSITY713 | | | | | | |
| 1 | 0 | 3.7 | 0 | 0 | 0 | 40.0 |
| 2 | 13.8 | 14.8 | 0 | 33.3 | 0 | 0 |
| 3 | 17.2 | 14.8 | 50 | 33.3 | 20.0 | 0 |
| 4 | 55.2 | 48.1 | 50 | 33.3 | 40.0 | 40.0 |
| 5 | 13.8 | 18.5 | 0 | 0 | 40.0 | 20.0 |
| 1DIVERSITY1418 | | | | | | |
| 1 | 0 | 1 | 0 | 0 | 0 | 40.0 |
| 2 | 17.2 | 11.1 | 0 | 16.7 | 0 | 0 |
| 3 | 13.8 | 11.1 | 0 | 16.7 | 20.0 | 0 |
| 4 | 55.2 | 63.0 | 100 | 50.0 | 60.0 | 40.0 |
| 5 | 13.8 | 14.8 | 0 | 16.7 | 20.0 | 20.0 |

Figure 6 summarizes the TTTs according to important variables identified by BN#1. This study also calculated the *p*-value to identify those BE variables that may cause a significant difference in TTTs. Based on the calculations, the significant difference in TTTs was found only in 1DIVERSITY1418 (*p*-value = 0.011). This means that attitudes of students regarding the availability of shops near their houses during the ages of 14–18 resulted in a significant difference in TTT to campus. Figure 6 shows that students who had shops near their house tended to choose shorter TTTs.
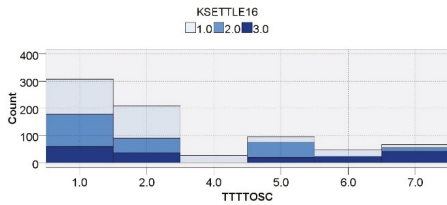


a. TTTTOSC vs. settlement type in the age period of 1–6.
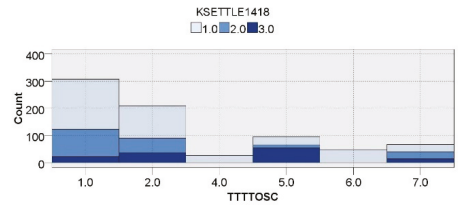
b. TTTTOSC vs. settlement type in the age period of 14–18.
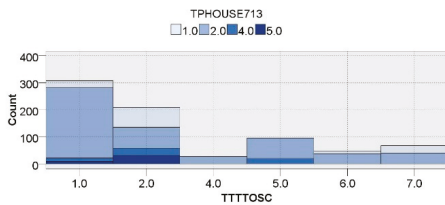
c. TTTTOSC vs. house type in the age period of 7–13.
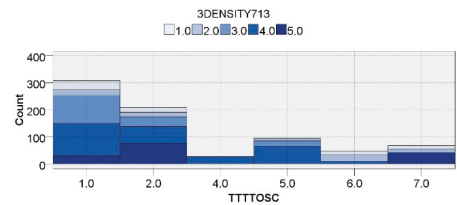
d. TTTTOSC vs. 3DENSITY7–13.
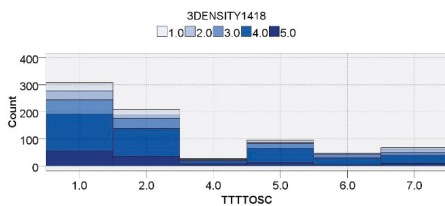
e. TTTTOSC vs. 3DENSITY14–18.
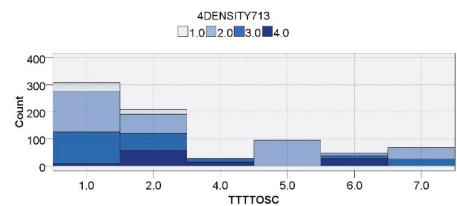
f. TTTTOSC vs. 4DENSITY7–13.
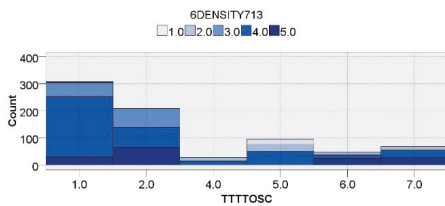
g. TTTTOSC vs. 6DENSITY7–13.
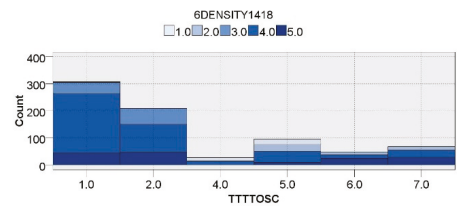
h. TTTTOSC vs. 6DENSITY14–18.

**Figure 6.** Histograms of TTT to school by important predictors of BN#1.

*4.3. BN#2 Model Considering the Control Variables and BE Variables*

The second BN model was developed using 38 variables. These variables included personal characteristics of the respondents, their household characteristics, variables related to the residential location, and travel mode choice. Eventually, the BN#2 selected 10 predictors as the most important and built the diagram based on these predictors (Figures 7 and 8). The training accuracy of this model was 81.01%. Apparently, among the BE variables, settlement type during the age periods of 1–6 and 14–18, as well as residential/house type during the age period of 7–13 were selected as the most important. Among the controlled variables, age, education level, race, usual travel mode to campus, and distance to campus were chosen as the most important.
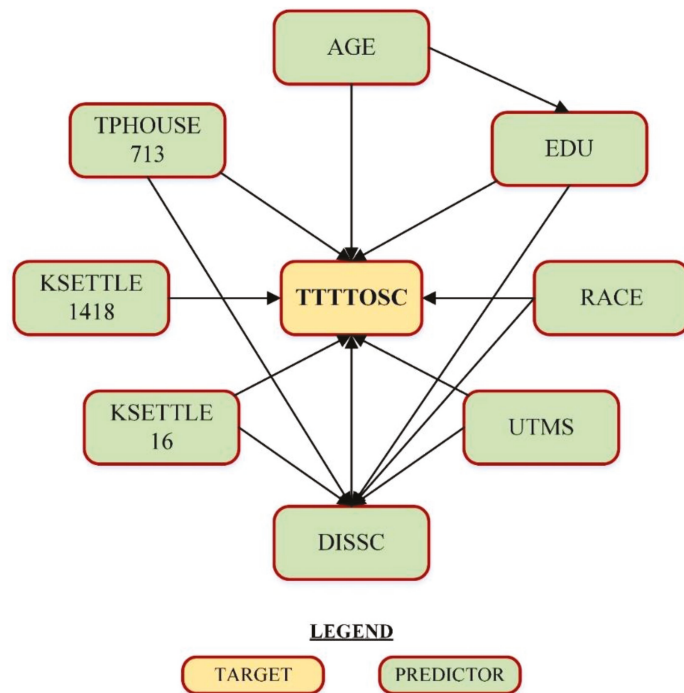


**Figure 7.** BN#2 diagram to predict the TTT of university students to campus considering the effects of control variables and the BE variables.

The BN#2 identified 53 conditional probabilities for each category of TTT, except TTT of 21–30 min. No TTT of 21–30 min was predicted by BN#2. To simplify the interpretation of the probabilities, only high probable TTTs (probability $\geq 0.75$) were reported for each category. The most frequent and influential value of each predictor for predicting each TTT is presented in Table 4.

A summary of TTTs by important control variables is presented in Figure 9. This study assessed whether any significant difference among TTTs exists regarding race, gender, education level, usual travel mode to campus, and distance to campus. Calculations obtained showed that differences in age and distance to campus significantly resulted in different TTTs (*p*-value = 0.008 and 0.000, respectively). The results indicated that the majority of younger students prefer to choose shorter TTTs. On the other hand, older students were inclined to select longer TTTs, such as 41–50 min. While the majority of students who lived closer to their school chose shorter TTTs, the students who lived far from

the school (more than 51 km) selected longer TTTs (more than 60 min). Figure 10 shows principal reasons for selecting the current houses by university students which provide a deeper insight into the factors that influenced university students' residential choices.
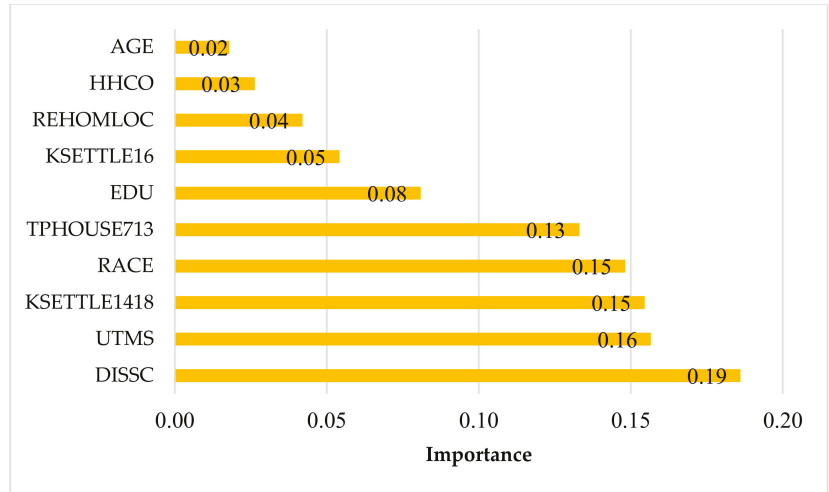


**Figure 8.** Importance of variables to predict the TTT of students to campus considering the effects of control variables and BE variables.

**Table 4.** Conditional probabilities of high probable TTTs to campus derived from BN#2.

| Variable/Value | Frequency (%) | | | | | |
|---|---|---|---|---|---|---|
| | 0–10 | 11–20 | 31–40 | 41–50 | 51–60 | More than 60 |
| AGE | | | | | | |
| 1 | 78.3 | 53.3 | 100 | 100 | 50.0 | 100 |
| 2 | 8.7 | 20.0 | 0 | 0 | 50.0 | 0 |
| 3 | 13.0 | 26.7 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 |
| EDU | | | | | | |
| 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 8.7 | 6.7 | 33.3 | 0 | 0 | 0 |
| 3 | 8.7 | 0 | 33.3 | 0 | 0 | 0 |
| 4 | 69.6 | 66.7 | 33.3 | 100.0 | 100.0 | 100.0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 13 | 26.7 | 0 | 0 | 0 | 0 |
| RACE | | | | | | |
| 1 | 52.2 | 53.3 | 66.7 | 0 | 0 | 75.0 |
| 2 | 30.4 | 26.7 | 33.3 | 33.3 | 0 | 25.0 |
| 3 | 4.3 | 0 | 0 | 33.3 | 100.0 | 0 |
| 4 | 13.0 | 20.0 | 0 | 33.3 | 0 | 0 |

**Table 4.** *Cont.*

| Variable/Value | Frequency (%) | | | | | |
|---|---|---|---|---|---|---|
| | 0–10 | 11–20 | 31–40 | 41–50 | 51–60 | More than 60 |
| UTMS | | | | | | |
| 1 | 52.2 | 53.3 | 66.7 | 100.0 | 100.0 | 50.0 |
| 2 | 26.1 | 0 | 0 | 0 | 0 | 0 |
| 3 | 13.0 | 20.0 | 33.3 | 0 | 0 | 50.0 |
| 4 | 8.7 | 26.7 | 0 | 0 | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 |
| DISSC | | | | | | |
| 1 | 87.0 | 33.3 | 0 | 0 | 0 | 0 |
| 2 | 13.0 | 46.7 | 33.3 | 66.7 | 50.0 | 50.0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 66.7 | 33.3 | 50.0 | 0 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 20.0 | 0 | 0 | 0 | 50.0 |
| KSETTLE16 | | | | | | |
| 1 | 39.1 | 66.7 | 100.0 | 0 | 50.0 | 25.0 |
| 2 | 39.1 | 13.3 | 0 | 66.7 | 0 | 25.0 |
| 3 | 21.7 | 20.0 | 0 | 33.3 | 50.0 | 50.0 |
| KSETTLE1418 | | | | | | |
| 1 | 56.5 | 60.0 | 100.0 | 0 | 100.0 | 25.0 |
| 2 | 34.8 | 13.3 | 0 | 33.3 | 0 | 50.0 |
| 3 | 8.7 | 26.7 | 0 | 66.7 | 0 | 25.0 |
| TPHOUSE713 | | | | | | |
| 1 | 8.7 | 40.0 | 0 | 0 | 0 | 25.0 |
| 2 | 78.3 | 26.7 | 66.7 | 33.3 | 100.0 | 75.0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 4.3 | 13.3 | 0 | 66.7 | 0 | 0 |
| 5 | 8.7 | 20.0 | 33.3 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 |
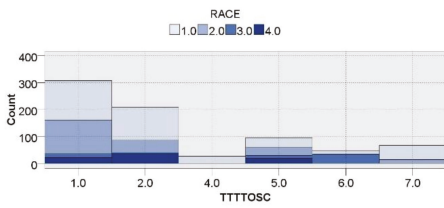
a. TTTTOSC vs. race.
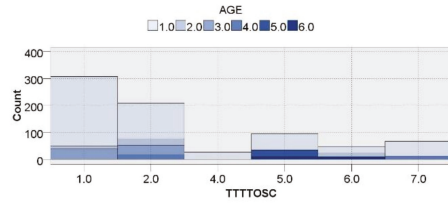


b. TTTTOSC vs. age.
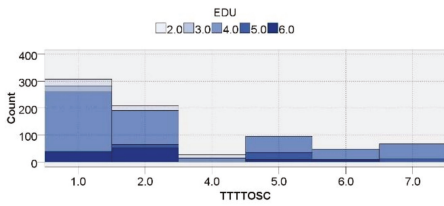


c. TTTTOSC vs. education level.
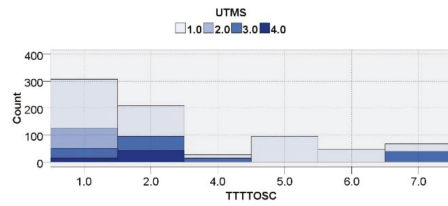


d. TTTTOSC vs. usual travel mode to campus.
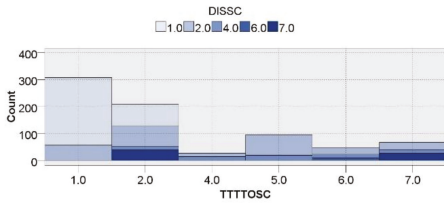


e. TTTTOSC vs. distance to campus.

**Figure 9.** Histograms of TTT to campus by important control predictors of BN#2.
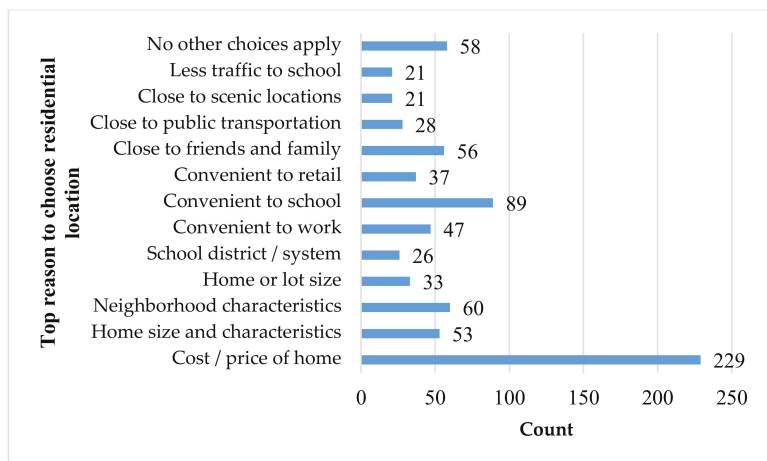


**Figure 10.** Top reasons to choose the residential location by the university students.

## 5. Discussion

Cumulatively, 68.33% of university students possessed TTT below 20 min to campus. This finding is in line with those of previous studies that revealed that ideal travel times below 20 min to different destinations were desirable for most of their respondents [80–82]. On the other hand, TTT found in this study differ from those described in He, Zhao, and He [30], Milakis and Van Wee [83], and Ye et al. [84], which showed that their participants' ideal commute time was mostly above 20 min. The possible reason for this contradiction could be the differences between the travel behavior and pattern of university students and other people [20,21].

The BE variables selected by the BN#1 model as the most important indicated that all variables related to the size, type, and composition of BEs may influence the TTT of university students to campus. These variables included those related to the settlement type, neighborhood density and diversity, and residential type. Till date, literature has confirmed the importance of current neighborhood attributes related to density and diversity in establishing the current travel behavior of commuters [85–88]. However, the results of this study are an immense and creative contribution to the body of literature that confirms that the past living environment experience of students in a diverse BE can affect their future travel behavior, particularly their TTT.

The first and second BN models showed that three BE attributes, including settlement type during the age period of 1–6, settlement type during the age period of 14–18, and apartment/house type during the age period of 7–13, are among the most influential factors of university students' TTT to campus. By retaining these large-scale BE variables in the BN#2, it can be indicatively explained that size and type attributes of BE may have more impact on the TTT of university students compared to the composition attributes. Moreover, the Pearson correlation tests did not find significant relationships between house type during the age period of 14–18 and the settlement type through the age period of 7–13 and TTT. However, this does not mean that the settlement and house type within these age periods do not influence the TTT of university students. Again, these variables may have less impact compared to peers of other age periods. These findings are unique in the sense that they provide insights into the importance of the role of built environment experiences during childhood and adolescence for analyzing university students' travel behavior. In addition, to the authors of this study's best knowledge, this is the first time that the impact of these kinds of experiences on university students' TTT to campus has been examined.

The BN#2 model did not adopt the BE variables related to diversity and density (which were selected as important variables by BN#1 model), to predict the TTT of university students to campus in the presence of control variables. This implies that a combination of sociodemographic attributes, trip characteristics, and non-composition BE attributes are more efficient variable sets for TTT forecasting of university students to campus. A possible explanation for this may emerge from the ability of people to recall larger characteristics of their living environment during their childhood and adolescence. Indeed, it is quite easy for people to remember the type of house and settlement in which they once lived.

The importance of BE variables for predicting the TTTs varied by age period. For example, for settlement type, the age periods of 1–6 and 14–18 were important while the age period of 7–13 was not. However, it is necessary to remark that this conclusion does not suggest that settlement type in the age period of 7–13 was not important at all but that it was less significant than other age periods for predicting the TTT of university students to campus. For those variables that were important in both the age periods of 7–13 and 14–18 (1DIVERSITY, 6DENSITY, and 3DENSITY), it could be argued that these variables would play a significant role in developing the future students' travel behavior and constantly affected the development of their travel habits and preferences. Arguably, availability of shops near the respondents' past houses and availability of residential buildings, entertainment facilities, and schools in the respondents' past neighborhoods may influence other future travel behaviors of people.

With regards to the controlled variables, race, age, and education level of students were selected as the critical sociodemographic variables to predict the TTT of university students to campus. Additionally, this present study identified the usual travel mode and distance of the school from home as important predictors of TTT of university students. However, no previous studies have assessed the impacts of such variables on TTT of university students. He, Zhao, and He [30] found the significant impacts of age, education level, and travel mode on tolerance threshold of commuting time of the general population to be important variables. Besides, the contribution of sociodemographic factors and travel mode to the TTT of the general population was confirmed in Páez and Whalen [89] and Redmond and Mokhtarian [90].

It was evidently shown that younger students tend to select shorter TTTs to campus. One important reason for this issue is that most survey participants (73.88%) belonged to the age cluster of 19–24 years. This is also in line with the fact that most UM and UTM students are in this age spectrum. Generally, younger students possess a weaker socioeconomic status compared to their older peers. They cannot buy a car and mostly use other active travel modes [91,92]. However, in this study, a majority of the students (73.62%) used private vehicles (car and motorcycle) to travel to campus. This result may be rooted in the high rate of vehicle ownership in Malaysia [93]. At the same time, 7.65% of the students adopted walking and cycling to campus, and their TTTs were 0–10 and 11–20 min. This finding was different from that of Milakis, Cervero, Van Wee, and Maat [23], Milakis and Van Wee [83], and Le et al. [94], that declared that people who walk or cycle had longer ATT than car users. On the other hand, the findings of this study regarding the lower TTTs of car users were in line with the same findings in the literature [23,83,94].

The analytical findings showed that most students who lived closer to the university experienced a shorter TTT. As explained earlier, most respondents were in the age range of 19–24. In Malaysia, many young students study at universities that are far from their hometowns. Besides, the majority of young students in public universities come from families with low socioeconomic status. Thus, these young students cannot afford to buy a house due to its high price, and they consider travel costs and choose to rent homes close to their campuses.

Certain implications for transport researchers and policy makers may be made from this present study. Findings presented in this study showed that the majority of university students had tendencies to experience shorter TTT to campus. Shorter TTTs may lead students to live in residential areas that are close to their campuses. This proximity of housing to the university may be a good opportunity for decision makers to implement sustainable transport solutions and provide sufficient facilities which could encourage students to use the active transport to campuses, such as sidewalks, bike paths, and bus stops. On the other hand, longer TTTs may lead students to live in housing in suburbs. Consequently, the students have to possess cars or motorcycles for travelling between the campus and residential areas if sufficient public transports are not available. Thus, the university decision makers should consider provision of a sufficient number of cheap housing units near the university campuses to decrease the need for using the private vehicles.

The findings of this present study also indicate that there is substantial homogeneity in the intrinsic preference for different TTTs and past BE experiences may create reference points for future travel behaviors and TTT of individuals. The findings also confirmed the undeniable intervention of BE in people's travel behaviors. Although using these factors for predicting future travel behaviors is still in its early stages, thus, urban and transport planners should include retrospective questions in their surveys to produce more accurate forecasts. Besides, researchers and policymakers should use longitudinal BE data and track the changes of BE over time and examine possible effects of these changes on individuals' future travel behaviors.

*Limitations*

In this present study, reference should be made to several limitations. First, university students may not represent the travel behavior of the general population in Malaysia. Thus, future studies can apply the same approach to different population groups to identify the impacts of the BE experiences during childhood or adolescence on their current travel behaviors. Second, this paper did not capture the TTTs of university students for leisure and shopping trips. Third, the study oversampled participants that have accessed the internet during the lockdown of COVID-19. The fourth constraining point of this study is that our study utilized self-report data. Trip observations may complement the self-report data in future investigations. This study obtained acceptable precision for BN models. However, larger datasets can be employed by further investigations to achieve greater accuracy. Fifthly, the participants of this survey were university students in Malaysia. However, in this country, the rate of vehicle ownership is very high. In addition, the overall condition of infrastructure supporting active transportation is poor. Therefore, the results of this study should be applied with caution to developed countries. Sixthly, this study considered a wide range of BE and sociodemographic factors; however, variables related to perceptions and habits were not included in this study. Thus, future studies can design surveys that include more variables to predict the TTT. Finally, the authors did not assess the BE experiences during the ages of 1-6 because it was challenging for students to remember the BE experiences. Thus, future studies can also include parents in their survey and ask them about the BE conditions when their child was 1–6 years old.

## 6. Conclusions

This present study used the Pearson chi-square technique and Bayesian network analysis to: (1) determine the most probable TTT of the off-campus university students to the campus; (2) investigate the association between off-campus university students' TTT to the campus and BE experiences during their childhood and teenage years; and (3) investigate the association between sociodemographic, household, residential, and travel mode characteristics of the off-campus university students' TTT to the campus.

A retrospective approach was adopted, which considered BE variables in the childhood and adolescent age periods to accompany sociodemographic, household characteristics, and current travel mode choice and residential location. The Pearson chi-square analysis identified 34 variables out of 74 candidate inputs. These variables were involved as predictors of the target (i.e., university students' TTT to campus) in BN analysis. Two BN models, including BN#1 and BN#2, were developed. The BN#1 applied only on BE variables. By developing this model, the availability of residential buildings in the neighborhoods that respondents lived in, during the age period of 14–18, was shown to be the most critical predictor of TTT of university students to campus. BN#2 was applied on all 34 variables. By running this second model, distance to campus was chosen as the most important. BE variables, including settlement type during the age period of 1–6 and 14–18 and house type in the age period of 7–13, were also identified as the most significant factors.

It is a challenging task to obtain information regarding the past living environment of university students and predict their future travel behavior based on these experiences. However, the results of this study can be instructive for urban and transport planners in the sense that built environment attributes can play an essential role during the whole life-course and the development of travel behaviors and patterns of individuals. To achieve more sustainable commute behavior in the future, planners and designers should consider more compact and mixed-use neighborhoods. In Malaysia, the rate of vehicle ownership is high. While several other factors, such as weather, low price of cars, and cheap parking are associated with this high vehicle ownership rate, advocating more sustainable behaviors may help the youths to minimize the usage of cars. Compact and dense living environments during the early life-course of people may be a desirable setting to shape their future habits. The authors of this study believe that the tendency of people to have shorter TTT could emerge from their experiences of previous living environments, especially during childhood

and adolescence. Additionally, experiencing shorter trip distances before adulthood might mean habituation to the higher usage of alternative modes such as public transport, walking, and cycling. During adulthood, the habits of using these modes may result in less flexibility and prevent people from dwelling in suburbs, as well as prevention from sprawling.

**Author Contributions:** Conceptualization, M.A. (Mahdi Aghaabbasi); formal analysis, M.A. (Mahdi Aghaabbasi); funding acquisition, Y.C., S.A., L.S., S.C., K.N., E.S. and R.F.; investigation, M.A. (Mahdi Aghaabbasi); methodology, M.A. (Mahdi Aghaabbasi); resources, Y.C., M.A. (Mujahid Ali), S.A., L.S., S.C., K.N., E.S., R.F. and R.Z.; supervision, R.Z.; writing—original draft, M.A. (Mahdi Aghaabbasi) and M.A. (Mujahid Ali). All authors have read and agreed to the published version of the manuscript.

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

## Appendix A

**Table A1.** Sociodemographic and household characteristics of respondents by their TTT.

| | Tolerability of Travel Time | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | Tolerate | | Moderate Tolerate | | Highly Tolerate | | Extensively Tolerate | |
| | N | % | N | % | N | % | N | % |
| **Age** | | | | | | | | |
| 19–24 | 392 | 75.7 | 28 | 100 | 85 | 59.0 | 55 | 80.9 |
| 25–30 | 35 | 6.8 | | | 14 | 9.7 | | |
| 31–36 | 72 | 13.9 | | | | | | |
| 37–42 | 19 | 3.7 | | | | | 13 | 19.1 |
| 43–48 | | | | | 25 | 17.4 | | |
| More than 48 | | | | | 20 | 13.9 | | |
| **Gender** | | | | | | | | |
| Male | 295 | 56.9 | 28 | 100 | 56 | 38.9 | 41 | 60.3 |
| Female | 223 | 43.1 | | | 88 | 61.1 | 27 | 39.7 |
| **Education level** | | | | | | | | |
| Primary | | | | | | | | |
| Secondary | 44 | 8.5 | 14 | 50 | | | | |
| Diploma | 22 | 4.2 | | | | | | |
| Bachelor's degree | 348 | 67.2 | 14 | 50 | 99 | 68.8 | 55 | 80.9 |
| Master's degree | 13 | 2.5 | | | 25 | 17.4 | 13 | 19.1 |
| Doctorate degree | 91 | 17.6 | | | 20 | 13.9 | | |
| **Income** | | | | | | | | |
| Less than MYR 1000 | 97 | 18.7 | 5 | 17.9 | 35 | 24.3 | 14 | 20.6 |
| Between MYR 1000 and MYR 2000 | 58 | 11.2 | 4 | 14.3 | 20 | 13.9 | 6 | 8.8 |
| Between MYR 2000 and MYR 3000 | 74 | 14.3 | 6 | 21.4 | 18 | 12.5 | 9 | 13.2 |
| Between MYR 3000 and MYR 6000 | 89 | 17.2 | 4 | 14.3 | 18 | 12.5 | 11 | 16.2 |
| Between MYR 6000 and MYR 13,000 | 145 | 28.0 | 5 | 17.9 | 41 | 28.5 | 22 | 32.4 |
| More than MYR 13,000 | 55 | 10.6 | 4 | 14.3 | 12 | 8.3 | 6 | 8.8 |
| **Race** | | | | | | | | |
| Malay | 269 | 51.9 | 28 | 100 | 48 | 33.3 | 54 | 79.4 |

**Table A1.** *Cont.*

| | Tolerability of Travel Time | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Tolerate | | Moderate Tolerate | | Highly Tolerate | | Extensively Tolerate | |
| | N | % | N | % | N | % | N | % |
| Chinese | 173 | 33.4 | | | 31 | 21.5 | 14 | 20.6 |
| Indian | 14 | 2.7 | | | 45 | 31.3 | | |
| Foreigner | 62 | 12.0 | | | 20 | 13.9 | | |
| Vehicle ownership | | | | | | | | |
| Yes | 54 | 79.4 | 14 | 50.0 | 80 | 55.6 | | |
| No | 14 | 20.6 | 14 | 50.0 | 64 | 44.4 | 68 | 100 |
| Vehicle count | | | | | | | | |
| 0 | 9 | 1.7 | 1 | 3.6 | 2 | 1.4 | 3 | 4.4 |
| 1 | 129 | 24.9 | 6 | 21.4 | 31 | 21.5 | 14 | 20.6 |
| 2 | 145 | 28.0 | 10 | 35.7 | 44 | 30.6 | 21 | 30.9 |
| >3 | 235 | 47.1 | 11 | 39.3 | 67 | 46.6 | 30 | 44.1 |
| Number of children | | | | | | | | |
| 0 | 351 | 67.8 | 17 | 60.7 | 90 | 62.5 | 47 | 69.1 |
| 1 | 71 | 13.7 | 7 | 25.0 | 28 | 19.4 | 10 | 14.7 |
| 2 | 59 | 11.4 | 3 | 10.7 | 13 | 9.0 | 6 | 8.8 |
| >3 | 37 | 7.2 | 1 | 3.6 | 13 | 9.0 | 5 | 7.4 |
| Number of people in household | | | | | | | | |
| 1 | 4 | 0.8 | 1 | 3.6 | 1 | 0.7 | 2 | 2.9 |
| 2 | 42 | 8.1 | 2 | 7.1 | 12 | 8.3 | 4 | 5.9 |
| 3 | 72 | 13.9 | 4 | 14.3 | 27 | 18.8 | 10 | 14.7 |
| >4 | 400 | 77.2 | 21 | 75.1 | 104 | 72.3 | 52 | 76.4 |

## References

1. Wheatley, D. Travel-to-work and subjective well-being: A study of UK dual career households. *J. Transp. Geogr.* **2014**, *39*, 187–196. [CrossRef]
2. Ye, R.; Titheridge, H. *Impact of Individuals' Commuting Trips on Subjective Well-Being-Evidence from Xi'an, China*; University College London: London, UK, 2015.
3. Oliveira, R.; Moura, K.; Viana, J.; Tigre, R.; Sampaio, B. Commute duration and health: Empirical evidence from Brazil. *Transp. Res. Part A Policy Pract.* **2015**, *80*, 62–75. [CrossRef]
4. Ali, M.; Dharmowijoyo, D.B.E.; de Azevedo, A.R.G.; Fediuk, R.; Ahmad, H.; Salah, B. Time-Use and Spatio-Temporal Variables Influence on Physical Activity Intensity, Physical and Social Health of Travelers. *Sustainability* **2021**, *13*, 12226. [CrossRef]
5. Ali, M.; Dharmowijoyo, D.B.; Harahap, I.S.; Puri, A.; Tanjung, L.E. Travel behaviour and health: Interaction of Activity-Travel Pattern, Travel Parameter and Physical Intensity. *Solid State Technol.* **2020**, *63*, 4026–4039.
6. Dijst, M.; Vidakovic, V. Travel time ratio: The key factor of spatial reach. *Transportation* **2000**, *27*, 179–199. [CrossRef]
7. Schwanen, T.; Dijst, M. Travel-time ratios for visits to the workplace: The relationship between commuting time and work duration. *Transp. Res. Part A Policy Pract.* **2002**, *36*, 573–592. [CrossRef]
8. Deka, D. The living, moving and travel behaviour of the growing American solo: Implications for cities. *Urban Stud.* **2014**, *51*, 634–654. [CrossRef]
9. Bolte, G.; Fromme, H.; Grp, G.M.E.S. Influence of Built Environment and Socioeconomic Position on Children's Physical Activity During Leisure Time and due to Mode of Travel. *Epidemiology* **2009**, *20*, S53. [CrossRef]
10. Ma, X.L.; Yang, J.; Ding, C.; Liu, J.F.; Zhu, Q. Joint Analysis of the Commuting Departure Time and Travel Mode Choice: Role of the Built Environment. *J. Adv. Transp.* **2018**, *2018*, 4540832. [CrossRef]
11. Manville, M. Travel and the Built Environment: Time for Change. *J. Am. Plan. Assoc.* **2017**, *83*, 29–32. [CrossRef]
12. Molina-Garcia, J.; Menescardi, C.; Estevan, I.; Martinez-Bello, V.; Queralt, A. Neighborhood Built Environment and Socioeconomic Status are Associated with Active Commuting and Sedentary Behavior, but not with Leisure-Time Physical Activity, in University Students. *Int. J. Environ. Res. Public Health* **2019**, *16*, 3176. [CrossRef]
13. Mouratidis, K. Built environment and leisure satisfaction: The role of commute time, social interaction, and active travel. *J. Transp. Geogr.* **2019**, *80*, 102491. [CrossRef]
14. Yu, L.; Xie, B.L.; Chan, E.H.W. Exploring impacts of the built environment on transit travel: Distance, time and mode choice, for urban villages in Shenzhen, China. *Transp. Res. Part E-Logist. Transp. Rev.* **2019**, *132*, 57–71. [CrossRef]
15. Wang, D.; Liu, Y. Factors influencing public transport use: A study of university commuters' travel and mode choice behaviours. In Proceedings of the State of Australian Cities Conference, Gold Coast, Australia, 9–11 December 2015.

16. Abdul Sukora, N.S.; Hassan, S.A. En route to a sustainable campus–an analysis of university students' travel patterns via 7 day travel diary. *J. Teknol.* **2014**, *70*, 9–16. [CrossRef]
17. Lundberg, B.; Weber, J. Non-motorized transport and university populations: An analysis of connectivity and network perceptions. *J. Transp. Geogr.* **2014**, *39*, 165–178. [CrossRef]
18. Whalen, K.E.; Páez, A.; Carrasco, J.A. Mode choice of university students commuting to school and the role of active travel. *J. Transp. Geogr.* **2013**, *31*, 132–142. [CrossRef]
19. Aghaabbasi, M.; Shekari, Z.A.; Shah, M.Z.; Olakunle, O.; Armaghani, D.J.; Moeinaddini, M. Predicting the use frequency of ride-sourcing by off-campus university students through random forest and Bayesian network techniques. *Transp. Res. Part A Policy Pract.* **2020**, *136*, 262–281. [CrossRef]
20. Garikapati, V.M.; You, D.; Pendyala, R.M.; Patel, T.; Kottommannil, J.; Sussman, A. Design, development, and implementation of a university travel demand modeling framework. *J. Transp. Res. Board* **2016**, *2563*, 105–113. [CrossRef]
21. Khattak, A.; Wang, X.; Son, S.; Agnello, P. Travel by university students in Virginia. *J. Transp. Res. Board* **2011**, *2255*, 137–145. [CrossRef]
22. Namgung, M.; Akar, G. Influences of neighborhood characteristics and personal attitudes on university commuters' public transit use. *J. Transp. Res. Board* **2015**, *2500*, 93–101. [CrossRef]
23. Milakis, D.; Cervero, R.; Van Wee, B.; Maat, K. Do people consider an acceptable travel time? Evidence from Berkeley, CA. *J. Transp. Geogr.* **2015**, *44*, 76–86. [CrossRef]
24. Simon, H.A. Rational choice and the structure of the environment. *Psychol. Rev.* **1956**, *63*, 129. [CrossRef]
25. Simon, H.A. A behavioral model of rational choice. *Q. J. Econ.* **1955**, *69*, 99–118. [CrossRef]
26. Wright, P.; Barbour, F. *Phased Decision Strategies: Sequels to an Initial Screening*; Graduate School of Business, Stanford University: Stanford, CA, USA, 1977.
27. Ryan, J.; Zahavi, Y. Stability of Travel Components Over Time. *Transp. Res. Rec.* **1980**, *750*, 19–26.
28. Zahavi, Y.; Talvitie, A. Regularities in Travel Time and Money Expenditures. In Proceedings of the 59th Annual Meeting of the Transportation Research Board, Washington, DC, USA, 21–25 January 1980.
29. Mokhtarian, P.L.; Salomon, I. How derived is the demand for travel? Some conceptual and measurement considerations. *Transp. Res. Part A Policy Pract.* **2001**, *35*, 695–719. [CrossRef]
30. He, M.; Zhao, S.; He, M. Tolerance threshold of commuting time: Evidence from Kunming, China. *J. Transp. Geogr.* **2016**, *57*, 1–7. [CrossRef]
31. Marchetti, C. Anthropological invariants in travel behavior. *Technol. Forecast. Soc. Chang.* **1994**, *47*, 75–88. [CrossRef]
32. Wener, R.E.; Evans, G.W.; Phillips, D.; Nadler, N. Running for the 7: 45: The effects of public transit improvements on commuter stress. *Transportation* **2003**, *30*, 203–220. [CrossRef]
33. Evans, G.W.; Wener, R.E. Rail commuting duration and passenger stress. *Health Psychol.* **2006**, *25*, 408. [CrossRef]
34. Novaco, R.W.; Gonzalez, O.I. Commuting and well-being. In *Technology and Psychological Well-Being*; Amichai-Hamburger, Y., Ed.; Cambridge University Press: Cambridge, UK, 2009; pp. 174–205.
35. Young, W.; Morris, J. Evaluation by individuals of their travel time to work. *Econometrica* **1978**, *46*, 403–426.
36. Hupkes, G. The law of constant travel time and trip-rates. *Futures* **1982**, *14*, 38–46. [CrossRef]
37. Ali, M.; de Azevedo, A.R.G.; Marvila, M.T.; Khan, M.I.; Memon, A.M.; Masood, F.; Almahbashi, N.M.; Shad, M.K.; Khan, M.A.; Fediuk, R.; et al. The Influence of COVID-19-Induced Daily Activities on Health Parameters—A Case Study in Malaysia. *Sustainability* **2021**, *13*, 7465. [CrossRef]
38. Clark, B.; Chatterjee, K.; Melia, S. Changes to commute mode: The role of life events, spatial context and environmental attitude. *Transp. Res. Part A Policy Pract.* **2016**, *89*, 89–105. [CrossRef]
39. Verplanken, B.; Walker, I.; Davis, A.; Jurasek, M. Context change and travel mode choice: Combining the habit discontinuity and self-activation hypotheses. *J. Environ. Psychol.* **2008**, *28*, 121–127. [CrossRef]
40. Gao, K.; Yang, Y.; Sun, L.; Qu, X. Revealing psychological inertia in mode shift behavior and its quantitative influences on commuting trips. *Transp. Res. Part F Traffic Psychol. Behav.* **2020**, *71*, 272–287. [CrossRef]
41. Lanzendorf, M. Mobility biographies: A new perspective for understanding travel behaviour. In Proceedings of the 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland, 10–15 August 2003.
42. Scheiner, J. Gendered key events in the life course: Effects on changes in travel mode choice over time. *J. Transp. Geogr.* **2014**, *37*, 47–60. [CrossRef]
43. Lanzendorf, M. Key events and their effect on mobility biographies: The case of childbirth. *Int. J. Sustain. Transp.* **2010**, *4*, 272–292. [CrossRef]
44. Van der Waerden, P. The influence of key events and critical incidents on transport mode choice switching behaviour: A descriptive analysis. In Proceedings of the Proceedings of 10th International Conference on Travel Behaviour Research, Lucerne, Switzerland, 10–15 August 2003.
45. Clark, B.; Chatterjee, K.; Melia, S.; Knies, G.; Laurie, H. Life events and travel behavior: Exploring the interrelationship using UK household longitudinal study data. *Transp. Res. Rec.* **2014**, *2413*, 54–64. [CrossRef]
46. Scheiner, J.; Holz-Rau, C. Changes in travel mode use after residential relocation: A contribution to mobility biographies. *Transportation* **2013**, *40*, 431–458. [CrossRef]

47. González, R.M.; Marrero, Á.S.; Cherchi, E. Testing for inertia effect when a new tram is implemented. *Transp. Res. Part A Policy Pract.* **2017**, *98*, 150–159. [CrossRef]
48. Yeh, A.G.; Yuen, B. Tall building living in high density cities: A comparison of Hong Kong and Singapore. In *High-Rise Living in Asian Cities*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 9–23.
49. Yeh, A.G.-O. *The Planning and Management of a Better High Density Environment*; Routledge: London, UK, 2000; p. 400.
50. Takayama, N.; Petrova, E.; Matsushima, H.; Furuya, K.; Ueda, H.; Mironov, Y.; Petrova, A.; Aoki, Y. Values, Concerns, and Attitudes Toward the Environment in Japan and Russia. *Urban Reg. Plan. Rev.* **2015**, *2*, 43–67. [CrossRef]
51. Deiner, P. *Inclusive Early Childhood Education: Development, Resources, and Practice*; Cengage Learning: Boston, MA, USA, 2012.
52. Michael Page. How Long Is Your Commute to Work? Available online: https://www.michaelpage.com.my/content/how-long-is-your-commute-to-work/ (accessed on 1 January 2020).
53. Mbara, T.C.; Celliers, C. Travel patterns and challenges experienced by University of Johannesburg off-campus students. *J. Transp. Supply Chain. Manag.* **2013**, *7*, 167700576. [CrossRef]
54. Behrens, R.; Mistro, R.D. Shocking habits: Methodological issues in analyzing changing personal travel behavior over time. *Int. J. Sustain. Transp.* **2010**, *4*, 253–271. [CrossRef]
55. van de Coevering, P.; Maat, K.; van Wee, B. Multi-period Research Designs for Identifying Causal Effects of Built Environment Characteristics on Travel Behaviour. *Transp. Rev.* **2015**, *35*, 512–532. [CrossRef]
56. de Azevedo, A.R.G.; Marvila, M.T.; Ali, M.; Khan, M.I.; Masood, F.; Vieira, C.M.F. Effect of the addition and processing of glass polishing waste on the durability of geopolymeric mortars. *Case Stud. Constr. Mater.* **2021**, *15*, e00662. [CrossRef]
57. Cervero, R.; Kockelman, K. Travel demand and the 3Ds: Density, diversity, and design. *Transp. Research. Part D Transp. Environ.* **1997**, *2*, 199–219. [CrossRef]
58. Ewing, R.; Greenwald, M.; Zhang, M.; Walters, J.; Feldman, M.; Cervero, R.; Thomas, J. *Measuring the Impact of Urban form and Transit Access on Mixed Use Site Trip Generation Rates—Portland Pilot Study*; US Environmental Protection Agency: Washington, DC, USA, 2009.
59. Rotaris, L.; Danielis, R.; Maltese, I. Carsharing use by college students: The case of Milan and Rome. *Transp. Res. Part A Policy Pract.* **2019**, *120*, 239–251. [CrossRef]
60. Nguyen-Phuoc, D.Q.; Amoh-Gyimah, R.; Tran, A.T.P.; Phan, C.T. Mode choice among university students to school in Danang, Vietnam. *Travel Behav. Soc.* **2018**, *13*, 1–10. [CrossRef]
61. Tezcan, H.O. Potential of carpooling among unfamiliar users: Case of undergraduate students at Istanbul Technical University. *J. Urban Plan. Dev.* **2016**, *142*, 04015006. [CrossRef]
62. Zhou, J. Proactive sustainable university transportation: Marginal effects, intrinsic values, and university students' mode choice. *Int. J. Sustain. Transp.* **2016**, *10*, 815–824. [CrossRef]
63. Stylianou, K.; Dimitriou, L.; Abdel-Aty, M. Big data and road safety: A comprehensive review. In *Mobility Patterns, Big Data and Transport Analytics*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 297–343.
64. Ali, M.; Room, S.; Khan, M.I.; Masood, F.; Ali Memon, R.; Khan, R.; Memon, A.M. Assessment of local earthen bricks in perspective of physical and mechanical properties using Geographical Information System in Peshawar, Pakistan. *Structures* **2020**, *28*, 2549–2561. [CrossRef]
65. Lv, Z.; Li, Y.; Feng, H.; Lv, H. Deep Learning for Security in Digital Twins of Cooperative Intelligent Transportation Systems. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 1–10. [CrossRef]
66. Yan, X.; Richards, S.; Su, X. Using hierarchical tree-based regression model to predict train-vehicle crashes at passive highway-rail grade crossings. *Accid. Anal. Prev.* **2010**, *42*, 64–74. [CrossRef] [PubMed]
67. Karlaftis, M.G.; Golias, I. Effects of road geometry and traffic volumes on rural roadway accident rates. *Accid. Anal. Prev.* **2002**, *34*, 357–365. [CrossRef]
68. Ahmad, I.; Basheri, M.; Iqbal, M.J.; Rahim, A. Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. *IEEE Access* **2018**, *6*, 33789–33795. [CrossRef]
69. Utkin, L.V.; Konstantinov, A.V.; Chukanov, V.S.; Kots, M.V.; Ryabinin, M.A.; Meldo, A.A. A weighted random survival forest. *Knowl.-Based Syst.* **2019**, *177*, 136–144. [CrossRef]
70. Khatami, A.; Khosravi, A.; Nguyen, T.; Lim, C.P.; Nahavandi, S. Medical image analysis using wavelet transform and deep belief networks. *Expert Syst. Appl.* **2017**, *86*, 190–198. [CrossRef]
71. Friedman, N.; Geiger, D.; Goldszmidt, M. Bayesian Network Classifiers. *Mach. Learn.* **1997**, *29*, 131–163. [CrossRef]
72. Barreto, E.D.; Stafanato, K.V.; Marvila, M.T.; de Azevedo, A.R.; Ali, M.; Pereira, R.M.; Monteiro, S.N. Clay Ceramic Waste as Pozzolan Constituent in Cement for Structural Concrete. *Materials* **2021**, *14*, 2917. [CrossRef]
73. Qian, Y.; Aghaabbasi, M.; Ali, M.; Alqurashi, M.; Salah, B.; Zainol, R.; Moeinaddini, M.; Hussein, E.E. Classification of Imbalanced Travel Mode Choice to Work Data Using Adjustable SVM Model. *Appl. Sci.* **2021**, *11*, 11916. [CrossRef]
74. Leilabadi, S.H.; Schmidt, S. In-depth Analysis of Autonomous Vehicle Collisions in California. In Proceedings of the 2019 IEEE Intelligent Transportation Systems Conference (ITSC), Auckland, NZ, USA, 27–29 October 2019.
75. Liu, T.; Zhang, Y.; Chen, J.; Shen, H. Discovery of Association Rule of Learning Action Based on Bayesian Network. In Proceedings of the 2018 9th International Conference on Information Technology in Medicine and Education (ITME), Hangzhou, China, 19–21 October 2018; pp. 466–470.

76. Susanti, S.P.; Azizah, F.N. Imputation of missing value using dynamic Bayesian network for multivariate time series data. In Proceedings of the 2017 International Conference on Data and Software Engineering (ICoDSE), Buenos Aires, Argentina, 20–28 May 2017; pp. 1–5.
77. Wu, Q.; Yang, C.; Gao, X.; He, P.; Chen, G. EPAB: Early Pattern Aware Bayesian Model for Social Content Popularity Prediction. In Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 17–20 November 2018; pp. 1296–1301.
78. Tareeq, S.M.; Inamura, T. A sample discarding strategy for rapid adaptation to new situation based on Bayesian behavior learning. In Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics, Bangkok, Thailand, 14–17 December 2009; pp. 1950–1955.
79. Prati, G.; Pietrantoni, L.; Fraboni, F. Using data mining techniques to predict the severity of bicycle crashes. *Accid. Anal. Prev.* **2017**, *101*, 44–54. [CrossRef] [PubMed]
80. O'Fallon, C.; Wallis, I. *A Wider Look at How Travellers Value the Quality and Quantity of Travel Time*; NZ Transport Agency Research Report 469; NZ Transport Agency: Wellington, New Zealand, 2012; p. 128.
81. Russell, M.L. *Travel Time Use on Public Transport: What Passengers Do and How It Affects Their Wellbeing*; University of Otago: Dunedin, New Zealand, 2012.
82. Russell, M.; Mokhtarian, P. How real is a reported desire to travel for its own sake? Exploring the 'teleportation' concept in travel behaviour research. *Transportation* **2014**, *42*, 333–345. [CrossRef]
83. Milakis, D.; Van Wee, B. "For me it is always like half an hour": Exploring the acceptable travel time concept in the US and European contexts. *Transp. Policy* **2018**, *64*, 113–122. [CrossRef]
84. Ye, R.; De Vos, J.; Ma, L. Analysing the association of dissonance between actual and ideal commute time and commute satisfaction. *Transp. Res. Part A Policy Pract.* **2020**, *132*, 47–60. [CrossRef]
85. Ewing, R.; Cervero, R. Travel and the built environment. *J. Am. Plan. Assoc.* **2010**, *76*, 265–294. [CrossRef]
86. Cervero, R. Built environments and mode choice: Toward a normative framework. *Transp. Res. Part D Transp. Environ.* **2002**, *7*, 265–284. [CrossRef]
87. Cervero, R.; Duncan, M. Walking, bicycling, and urban landscapes: Evidence from the San Francisco Bay Area. *Am. J. Public Health* **2003**, *93*, 1478–1483. [CrossRef] [PubMed]
88. Ewing, R.; Handy, S. Measuring the unmeasurable: Urban design qualities related to walkability. *J. Urban Des.* **2009**, *14*, 65–84. [CrossRef]
89. Páez, A.; Whalen, K. Enjoyment of commute: A comparison of different transportation modes. *Transp. Res. Part A Policy Pract.* **2010**, *44*, 537–549. [CrossRef]
90. Redmond, L.S.; Mokhtarian, P.L. The positive utility of the commute: Modeling ideal commute time and relative desired commute amount. *Transportation* **2001**, *28*, 179–205. [CrossRef]
91. Habib, K.N.; Weiss, A.; Hasnine, S. On the heterogeneity and substitution patterns in mobility tool ownership choices of post-secondary students: The case of Toronto. *Transp. Res. Part A Policy Pract.* **2018**, *116*, 650–665. [CrossRef]
92. Mitra, R.; Nash, S. Can the built environment explain gender gap in cycling? An exploration of university students' travel behavior in Toronto, Canada. *Int. J. Sustain. Transp.* **2018**, *13*, 138–147. [CrossRef]
93. Govindaraju, V.C. Greening the economy with low carbon energy system: Developments, policy initiatives and lessons from Malaysia. In *Investing on Low-Carbon Energy Systems*; Springer: Berlin/Heidelberg, Germany, 2016; pp. 111–129.
94. Le, H.T.K.; Buehler, R.; Fan, Y.; Hankey, S. Expanding the positive utility of travel through weeklong tracking: Within-person and multi-environment variability of ideal travel time. *J. Transp. Geogr.* **2020**, *84*, 102679. [CrossRef]

*Article*

# Applying Machine Learning to Explore Feelings about Sharing the Road with Autonomous Vehicles as a Bicyclist or as a Pedestrian

**Zohreh Asadi-Shekari [1,\*], Ismaïl Saadi [2,3,4] and Mario Cools [2,5,6]**

[1] Centre for Innovative Planning and Development, CIPD, Universiti Teknologi Malaysia, Skudai, Johor Bahru 81310, Johor, Malaysia

[2] Local Environment & Management Analysis (LEMA), Urban and Environmental Engineering (UEE), University of Liège, Allée de la Découverte 9, Quartier Polytech 1, 4000 Liege, Belgium; ismail.saadi@uliege.be (I.S.); mario.cools@uliege.be (M.C.)

[3] F.R.S.-FNRS, Rue d'Egmont 5, 1050 Brussels, Belgium

[4] IFSTTAR, COSYS-GRETTIA, University Gustave Eiffel, F-77454 Marne-la-Vallée, France

[5] Department of Informatics, Simulation and Modeling, KU Leuven Campus Brussels, Warmoesberg 26, 1000 Brussels, Belgium

[6] Faculty of Business Economics, Hasselt University, Agoralaan Gebouw D, 3590 Diepenbeek, Belgium

\* Correspondence: asadi41360@gmail.com

**Abstract:** The current literature on public perceptions of autonomous vehicles focuses on potential users and the target market. However, autonomous vehicles need to operate in a mixed traffic condition, and it is essential to consider the perceptions of road users, especially vulnerable road users. This paper builds explicitly on the limitations of previous studies that did not include a wide range of road users, especially vulnerable road users who often receive less priority. Therefore, this paper considers the perceptions of vulnerable road users towards sharing roads with autonomous vehicles. The data were collected from 795 people. Extreme gradient boosting (XGBoost) and random forests are used to select the most influential independent variables. Then, a decision tree-based model is used to explore the effects of the selected most effective variables on the respondents who approve the use of public streets as a proving ground for autonomous vehicles. The results show that the effect of autonomous vehicles on traffic injuries and fatalities, being safe to share the road with autonomous vehicles, the Elaine Herzberg accident and its outcome, and maximum speed when operating in autonomous are the most influential variables. The results can be used by authorities, companies, policymakers, planners, and other stakeholders.

**Keywords:** autonomous vehicles; vulnerable road users; public perception; machine learning; most effective variables

## 1. Introduction

Most of the studies related to public perceptions of autonomous vehicles focus on potential users. For example, Silberg et al. [1] conducted a survey in California, New Jersey, and found the elderly and young people (from 18 to 25 years old) as the most potential users. They also found that providing incentives, such as designated lanes, was an important factor for adopting autonomous vehicles. Some of these studies explored the real presence of autonomous vehicles as a mobility option. Begg [2] explored the opinions of transportation experts in the U.K. about the real presence of autonomous vehicles on public roads. The experts suggested 2025 for level 4 and 2040 for level 5 (level 0: no driving automation; level 1: driver assistance; level 2: partial driving automation; level 3: conditional driving automation; level 4: high driving automation; level 5: full driving automation). This study also proposed safety as an important factor. Safety-related factors,

such as physical threats, privacy, and trust, are among the important factors in this type of study [3].

Some studies considered the effects of technology on perceptions of autonomous vehicles. Young adults and men are two groups that are more interested in autonomous vehicles than other demographic groups [4,5] since they are more interested in using new technologies [6]. Kyriakidis et al. [7] conducted a survey in different countries and found a positive association between driving and new technologies, such as cruise control usage and willingness to buy autonomous vehicles. They also found that respondents would be willing to pay more to have fully automated vehicles. However, Seapine Software [8] found equipment failures, liability issues, and hacking issues as important concerns for the potential users.

Few studies focused on the potentially shared mobility that can be provided by autonomous vehicles. Autonomous vehicles can be easily adopted for shared mobility, but people still prefer to have a private autonomous vehicle [9]. However, Haboucha et al. [10] found that men in Israel prefer shared autonomous vehicles. This is in line with other studies that found that public perceptions of autonomous vehicles and related effective factors can vary among different countries [11]. For example, autonomous vehicles were perceived as scary among 42% of respondents in a study in Japan, while this rate was 66% among U.S. respondents [12]. Therefore, Americans seem to have more safety concerns than other nationalities, such as the Japanese.

Desirability and willingness to buy is another approach in the current few studies that are related to public perceptions of autonomous vehicles. Casley et al. [13] found safety, legal issues, and cost important for autonomous vehicles' desirability. Jiang et al. [14] also found household size, age, and trip purposes as effective factors for willingness to buy autonomous vehicles and Shabanpour et al. [15] added price, incentives, and policies to these factors.

Autonomous vehicles make eating, working, sleeping, and doing possible during daily travel time [16]. They can increase safety by reducing distractions and human errors [17,18]. Moreover, current and future autonomous vehicles propose more safety benefits, such as intelligent speed assistance and advanced emergency braking. Public perception in addition to the technology and road infrastructure are important factors to find the effects of autonomous vehicles on travel behavior. However, most of the current studies related to autonomous vehicles mainly focus on motor vehicles and connectivity between vehicles and infrastructure [19], and only a few studies focus on the effects of these technologies on public perception.

Schoettle and Sivak [12] found that most respondents are not familiar with autonomous vehicles, but they believe in less distractions and fewer accidents for autonomous vehicles. Some studies focus on safety as one of the most significant factors for public perceptions of autonomous vehicles that can change travel behavior and mode, e.g., [20–24]. However, based on a literature review by Gkartzonikas et al. [25], only Hulse et al. [5] focused on the perceptions of pedestrians and Penmetsa et al. [26] focused on the perceptions of pedestrians and bicyclists. This is another important gap since autonomous vehicles need to operate in a mixed traffic condition that includes a wide range of road users. It is critical to consider the perceptions of vulnerable road users who often feel that they have less priority. If vulnerable road users, such as pedestrians and cyclists, do not feel comfortable sharing the roads with autonomous vehicles, using this new technology can negatively affect active travel options. This paper explores the perceptions of bicyclists and pedestrians to fill the gap of previous studies that did not include a wide range of road users, especially vulnerable road users.

Furthermore, most of the studies that examined perceptions and attitudes use descriptive analysis and prediction models, which can relate the perceptions of sharing the road with autonomous vehicles not been developed to date. This paper explores road users' perceptions, including vulnerable road users, towards autonomous vehicles and develops

prediction models using machine learning techniques to explore feelings about sharing the road with autonomous vehicles as a bicyclist or as a pedestrian.

Using machine learning and non-parametric techniques provides some advantages for this study. For example, these techniques do not need special assumptions or predefined functions that traditional parametric techniques need. In addition, non-parametric techniques can handle multicollinearity issues better than traditional parametric techniques. Because of high potential correlations between variables in this study, non-parametric techniques can be better options. Finally, these non-parametric models can be presented graphically, making them easy to interpret.

## 2. Materials and Methods

Bike Pittsburgh (BikePGH), a registered non-profit company, works to make the city safe and accessible for bicyclists. BikePGH launched two surveys in 2017 and 2019 to explore the feeling of pedestrians and bicyclists about sharing the road with autonomous vehicles, and this paper used the collected data from the latest one. In total, the data were collected from 795 people using the BikePGH related blog, website, and email list. The feeling about sharing the road with autonomous vehicles was the dependent variable in this paper. The independent variables included paying attention to the autonomous vehicles, familiarity with the technology behind autonomous vehicles, the experience of sharing the road with autonomous vehicles while riding a bicycle or walking, feeling safe while sharing the road with autonomous vehicles and human-driven cars, the effects of autonomous vehicles on traffic injuries and fatalities, the maximum speed when operating in autonomous mode, having full-time employees (pilot and co-pilot) at all times, operating in manual mode while in an active school zone, sharing some non-personal data, reporting all safety-related incidents, and previous accidents effects. In addition, some socio-demographic factors, such as postal address, being an active member of BikePGH, car ownership, having a smartphone and age, were also considered. Table 1 shows the description of dependent and independent variables in this paper.

In the first step, the most effective variables among the independent variables to predict the feelings about the use of public streets as a proving ground for autonomous vehicles was identified. In the next step, the identified effective variables were used as selected independent variables to explore the effects of these selected variables on the dependent variable. Extreme gradient boosting (XGBoost) and random forest were used to select the most effective independent variables. This is in line with recent related studies that deal with a high number of independent variables [27–31]. The random forest aggregates many binary decision trees. These trees are the result of a random choice of explanatory variables and bootstrap samples at each node. XGBoost [32] also generates multiple trees to improve accuracy. XGBoost and random forest are better options in comparison with other feature selection techniques. In other techniques, the importance ranking can be affected negatively by other associated inputs [33].

Cross-validation (10-fold cross-validation) is a resampling method that is applied to estimate the accuracy for this limited number of data. Cross-validation generally results in a less biased model than other methods, such as train and test split. After applying random forest and XGBoost, the SHAP (SHapley Additive exPlanations) values [34] were used to select the most effective variables. The SHAP is a value that can explain the contribution of each observation to the dependent variable. Therefore, it is possible to have local interpretability while the traditional importance values are related to each predictor and are based on the entire population. In addition, SHAP values can be estimated for each class (for nominal data) in the dependent variable.

**Table 1.** Description of the dependent and independent variables.

| Variables | Description | Codes |
|---|---|---|
| DV | What do you think about using public streets as a proving ground for autonomous vehicles (AVs)? | Approve (1), others (0) * |
| IV1 | To what extent have you been paying attention to the subject of AVs in the news? | 1–5 |
| IV2 | How familiar are you with the technology behind AVs? | 1–4 |
| IV3 | Have you shared the road with an AV while riding your bicycle? | Yes (1), no (0), not sure (2) |
| IV4 | Have you been near an AV while walking or using a mobility device (wheelchair, etc.)? | Yes (1), no (0), not sure (2) |
| IV5 | On a typical day, how safe do you feel sharing the road with autonomous vehicles? | 1–5 |
| IV6 | On a typical day, how safe do you feel sharing the road with human-driven cars? | 1–5 |
| IV7 | What effect do you think that AVs will have on traffic injuries and fatalities? | 1–5 |
| IV8 | Should AV speeds be capped at 25 mph when operating in autonomous mode? | Yes (1), no (0), not sure (2) |
| IV9 | Should AVs have two full-time employees (pilot and co-pilot) at all times? | Yes (1), no (0), not sure (2) |
| IV10 | Do you think that AVs should operate in manual mode while in an active school zone? | Yes (1), no (0), not sure (2) |
| IV11 | Should AV companies be required to share some non-personal data with the proper authorities? | Yes (1), no (0), not sure (2) |
| IV12 | Should AV companies be required to disclose information and data as to the limitations, capabilities, and real-world performance of their cars with the proper authorities? | Yes (1), no (0), not sure (2) |
| IV13 | Should AV companies be required to report all safety-related incidents with the proper authorities, even if a police report is not required? | Yes (1), no (0), not sure (2) |
| IV14 | In March of 2018, an AV struck and killed Elaine Herzberg, a pedestrian, in Tempe, AZ, U.S.A. As a pedestrian and/or bicyclist, how did this event and its outcome change your opinion about sharing the road with AVs? | 1–5 |
| IV15 | Zip Code | zip code |
| IV16 | Are you currently an active member of BikePGH? | Yes (1), no (0), not sure (2) |
| IV17 | Do you (or someone in your household) own an automobile? | Yes (1), no (0), not sure (2) |
| IV18 | Do you own a smartphone? | Yes (1), no (0) |
| IV19 | What is your age? | 1–7 ** |

Note: DV: dependent variable. IV: independent variable. * Somewhat approve, neutral, somewhat disapprove, disapprove. ** Under 18 (1), 18–24 (2), 25–34 (3), 35–44 (4), 45–54 (5), 55–64 (6), 65+ (7).

All independent variables were included in the random forest and XGBoost models, and then the not important variables were excluded one by one based on the SHAP values. The accuracy rate and the number of input variables were used to find the threshold for SHAP values. This threshold was used with the selected XGBoost or random forest in addition to finding the most effective variables.

A C5.0 model was used in this study to explore the effects of the selected most effective variables on the dependent variable. C5.0 is an improved version of C4.5 that is an extension of the ID3 algorithm [35–38]. In this C5.0 model, 2 and 75 are used as the minimum number for records per child branch and the pruning severity. To collapse weak subtrees, local and global pruning are used. The winnow attributes technique excludes irrelevant predictors and, before modelling, evaluates the relevancy of the predictors.

### 3. Results

Table 2 shows that more than 47% of respondents approve the use of public streets as a proving ground for autonomous vehicles. As was mentioned, the SHAP values can be estimated for each class in the dependent variable. Therefore, these values were used to find the most effective variables for respondents who approve the use of public streets as a proving ground for autonomous vehicles. Table 3 shows that both total and breakdown accuracy values are higher for the XGBoost model in comparison with the random forest model. In addition, in the XGBoost, 80% accuracy is achievable after including only four effective variables based on SHAP values and including more variables cannot significantly enhance the accuracy. In the random forest model, the accuracy after including four effective variables based on SHAP values is 76%. Therefore, the XGBoost model was chosen to find the effective variables.

**Table 2.** Frequency of different classes in the dependent variable.

| What Do You Think about Using Public Streets as a Proving Ground for AVs? | Frequency | Percent |
|---|---|---|
| Disapprove | 66 | 8.3 |
| Somewhat disapprove | 85 | 10.7 |
| Neutral | 92 | 11.6 |
| Somewhat approve | 168 | 21.1 |
| Approve | 381 | 47.9 |
| Missing | 3 | 0.4 |
| Total | 795 | 100.0 |

**Table 3.** Accuracy and confusion matrix for random forest and XGBoost models.

| Random Forest (Accuracy: 79%) *n* = 4 (Accuracy: 76%) | | |
|---|---|---|
| **What Do You Think about Using Public Streets as a Proving Ground for AVs?** | | |
| | 0 | 1 |
| 0 | 78% | 22% |
| 1 | 20% | 80% |
| **XGBoost (Accuracy: 80%)** **_n_ = 4 (Accuracy: 80%)** | | |
| | 0 | 1 |
| 0 | 78% | 22% |
| 1 | 18% | 82% |

Note: *n* = the number of including variables based on removing the not effective variables one by one considering the SHAP values. 1: approve. 0: somewhat approve, neutral, somewhat disapprove, and disapprove.

Table 4 shows the selected effective variables based on SHAP values resulting from the XGBoost model for respondents that approve the use of public streets as a proving ground for autonomous vehicles. Table 4 shows that the effect of autonomous vehicles on traffic injuries and fatalities, being safe to share the road with autonomous vehicles, Elaine Herzberg accident and its outcome, and autonomous vehicles speed when operating in autonomous mode are the most effective factors for respondents that approve the use of

public streets as a proving ground for autonomous vehicles. This table also indicates the most effective classes or attributes for these variables.

**Table 4.** The selected effective variables based on the SHAP values.

| Effective Variables | Classes in the Independent Variables | SHAP (Mean) |
|---|---|---|
| IV7 | Significantly better (5) | 0.612238 |
| IV5 | Being very safe (5) | 0.375330 |
| IV14 | No change (3) | 0.321566 |
| IV8 | No (0) | 0.294447 |

Note: IV7: What effect do you think that AVs will have on traffic injuries and fatalities? IV5: On a typical day, how safe do you feel sharing the road with autonomous vehicles? IV14: In March of 2018, an AV struck and killed Elaine Herzberg, a pedestrian, in Tempe, AZ, USA. As a pedestrian and/or bicyclist, how did this event and its outcome change your opinion about sharing the road with AVs? IV8: Should AV speeds be capped at 25 mph when operating in autonomous mode?

In the next step, a C5.0 model was used in this study to explore the effects of the selected most effective variables on the dependent variable. Figure 1 shows the proposed C5.0 decision tree. The frequency and percentage of each classification in the dependent variable are presented for each node. The overall accuracy is more than 79%, and the breakdown prediction accuracies are around 78% and 81% for 0 (somewhat approve, neutral, somewhat disapprove or disapprove) and 1 (approve) classes. There are five terminal nodes (the bottom nodes of the decision tree), and this model has four splitters, i.e., the effect of autonomous vehicles on traffic injuries and fatalities, being safe to share the road with autonomous vehicles, Elaine Herzberg accident and its outcome, and autonomous vehicles speed when operating in autonomous mode.
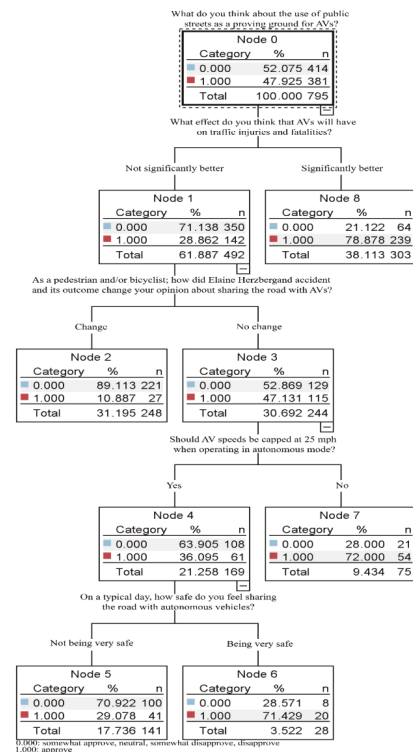


**Figure 1.** The proposed C5.0 model.

The model prediction is 1 for respondents who think that autonomous vehicles make traffic injuries and fatalities situations significantly better (refer to node 8 in Figure 1). The model prediction is 0 for respondents who do not think that autonomous vehicles make traffic injuries and fatalities situations significantly better and the Elaine Herzberg accident changed their opinions about sharing the road with autonomous vehicles (refer to node 2 in Figure 1). For respondents for whom the Herzberg accident did not change their opinions, speed when operating in autonomous mode and being safe to share the road with autonomous vehicles are important factors. For respondents for whom the Herzberg accident did not change their opinions, the model prediction is 1(refer to node 7 in Figure 1) if they do not believe in a maximum 25 mph speed when operating in autonomous mode; if they believe in a maximum 25 mph speed when operating in autonomous mode, the model prediction is 0 (refer node 5 in Figure 1) for respondents who do not think that it is very safe to share the road with autonomous vehicles; and 1 (refer node 6 in Figure 1) for respondents who think that it is very safe to share the road with autonomous vehicles.

## 4. Discussion and Conclusions

This study explores the perceived feelings of sharing roads with autonomous vehicles. The paper expands on the scope of previous studies by exploring the perceptions of bicyclists and pedestrians. Moreover, this paper builds explicitly on the limitation of previous studies that did not include a wide range of road users, especially vulnerable road users who often receive less priority. The findings suggest the XGBoost model finds the most influential variables. In addition, the analysis suggests the effect of autonomous vehicles on traffic injuries and fatalities, being safe to share the road with autonomous vehicles, the Elaine Herzberg accident and its outcome, and a maximum speed when operating in autonomous as effective variables to predict approval for the use of public streets as a proving ground for autonomous vehicles.

There are some other variables included in the model that are not related to safety (e.g., paying attention to the subject of autonomous vehicles in the news, familiarity with the technology behind autonomous vehicles, the experience of sharing roods with autonomous vehicles and human-driven cars, data sharing, related policies and some variables related to socio-demographic data), but the most effective variables are related to safety. However, some of these variables, such as familiarity and awareness, are significant in other studies. For example, Schoettle and Sivak [12], Silberg et al. [1] and Sanbonmatsu et al. [39] found a positive association between level of awareness and the intention to adopt autonomous vehicles. Nordhoff et al. [40] also found a similar association for driverless shuttles.

This is not a surprising result since safety is more important for vulnerable road users in comparison with drivers who are better protected. This point is further confirmed by the effects of the Elaine Herzberg accident, which is among the most effective variables. The findings are in line with previous studies that consider safety as an important factor (e.g., [21–25]). However, most of these studies focus on safety as a significant factor for changing travel behavior and mode. In addition, among these studies, only two considered the perceptions of pedestrians and bicyclists [5,27].

The policy relevance of this paper is underlined by the fact that at the individual level, we found safety as a very important factor, and the authorities need to be sure that autonomous vehicles are safe enough to be shared on the streets. Therefore, autonomous vehicle companies need to consider special procedures and cautions during their testing, and authorities need to provide related policies. Public perception, in this case, can be used both directly and indirectly. In addition, planners and other stakeholders need to provide more information to decrease public confusion about autonomous vehicles.

Non-parametric models, such as the proposed C5.0, have some advantages that make them preferable to the traditional parametric models. Chang and Wang [41] highlighted that non-parametric models (such as the proposed C5.0 model) do not need specific assumptions or a functional form and can handle multicollinearity problems, which are a common issue for independent variables in these data because of potentially high correlations between

these variables. The results are also more useful since these models focus on a reduced set of the most significant factors [41].

Despite these advantages mentioned above, these models have some disadvantages. For example, they do not have formal statistical inference procedures [41]. These models also do not have confidence intervals for the splitters and predictions [41]. Generally, it is not recommended to generalize the results based on the non-parametric techniques since these models are not very stable. Furthermore, the accuracy and structure may change significantly if different partitioning and stratified random sampling are used. Therefore, these models are usually used to find important variables and further techniques are needed to find final models. Since sampling and partitioning are not used in the proposed C5.0 model development, this disadvantage is not a significant concern for this study.

In addition to the mentioned advantages, machine learning has different applications in various engineering fields (e.g., [42–45]). Increasing interest in machine learning is because of various data, better computational tools and processing that make computation cheaper and more powerful. This means that applying machine learning can help us to develop more accurate models to analyze bigger and more complex data faster than the traditional techniques.

Some extensions of this study are essential. For example, consistent data collection for different regions needs to be considered since other areas are very different in terms of regulations and people experience with autonomous vehicles. Frequent additional data collection can be used to evaluate the effects of autonomous vehicles on public perception in addition to the evolution of public perceptions of autonomous vehicles. Additional questions, especially related to socio-demographic data, can be used to have more detailed insights. For example, gender-related data are not included in the BikePGH survey, or there is a very low response rate among the age groups that may have different ideas (just around 4% for 18–24 and around 12% for elderly). Finally, the target population in our study is bicyclists and pedestrians that represent these specific mode users. Adding a general population can be useful to have a baseline and a useful comparison. Future studies can also develop questionnaires following a scientific approach to avoid the gap and potential biases in the questions of the BikePGH survey that an interest group develops.

## References

1. Silberg, G.; Manassa, M.; Everhart, K.; Subramanian, D.; Corley, M.; Fraser, H.; Sinha, V.; Ready, A.W. *Self-Driving Cars: Are We Ready?* Technical Report; KPMG: Amstelveen, The Netherlands, 2013.
2. Begg, D. A 2050 Vision for London: What Are the Implications of Driverless Transport. Transport Times. 2014. Available online: http://www.transporttimes.co.uk/Admin/uploads/64165-transport-times_a-2050-vision-for-london_aw-web-ready.pdf (accessed on 19 July 2020).
3. Pyrialakou, V.D.; Gkartzonikas, C.; Gatlin, J.D.; Konstantina, G. Perceptions of safety on a shared road: Driving, cycling, or walking near an autonomous vehicle. *J. Saf. Res.* **2020**, *72*, 249–258. [CrossRef] [PubMed]
4. Nielsen, T.A.S.; Haustein, S. On sceptics and enthusiasts: What are the expectations towards self-driving cars? *Transp. Policy* **2018**, *66*, 49–55. [CrossRef]

5.  Hulse, L.M.; Xie, H.; Galea, E.R. Perceptions of autonomous vehicles: Relationships with road users, risk, gender and age. *Saf. Sci.* **2018**, *102*, 1–13. [CrossRef]
6.  Smith, A.; Caiazza, T. *Automation in Everyday Life*; Pew Research Center: Washington, DC, USA, 2017.
7.  Kyriakidis, M.; Happee, R.; De Winter, J.C.F. Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. *Transp. Res. Part F Traffic Psychol. Behav.* **2015**, *32*, 127–140. [CrossRef]
8.  Seapine Software. Study Finds 88 Percent of Adults Would Be Worried about Riding in a Driverless Car. 2014. Available online: http://www.seapine.com/about-us/press-release-full?press=217 (accessed on 19 July 2020).
9.  Pakusch, C.; Stevens, G.; Boden, A.; Bossauer, P.; Rosen, M.A. Unintended effects of autonomous driving: A study on mobility preferences in the future. *Sustainability* **2018**, *10*, 2404. [CrossRef]
10.  Haboucha, C.J.; Ishaq, R.; Shiftan, Y. User preferences regarding autonomous vehicles. *Transp. Part C Emerg. Technol.* **2017**, *78*, 37–49. [CrossRef]
11.  Moody, J.; Bailey, N.; Zhao, J. Public perceptions of autonomous vehicle safety: An international comparison. *Saf. Sci.* **2020**, *121*, 634–650. [CrossRef]
12.  Schoettle, B.; Sivak, M. *A Survey of Public Opinion about Autonomous and Self-Driving Vehicles in the U.S., U.K., and Australia*; Report No. UMTRI-2014-21; University of Michigan Transport Research Institute: Ann Arbor, MI, USA, 2014.
13.  Casley, S.V.; Jardim, A.S.; Quartulli, A.M. A Study of Public Acceptance of Autonomous Cars, Interactive Qualifying Project, Worcester Polytechnic Institute. 2013. Available online: https://web.wpi.edu/Pubs/Eproject/Available/Eproject043013155601/unrestricted/A_Study_of_Pblic_Acceptance_of_Autonomous_Cars.pdf (accessed on 19 July 2020).
14.  Jiang, Y.; Zhang, J.; Wang, Y.; Wang, W. Capturing ownership behavior of autonomous vehicles in Japan based on a stated preference survey and a mixed logit model with repeated choices. *Int. J. Sustain. Transp.* **2019**, *13*, 788–801. [CrossRef]
15.  Shabanpour, R.; Golshani, N.; Shamshiripour, A.; Mohammadian, A.K. Eliciting preferences for adoption of fully automated vehicles using best-worst analysis. *Transp. Res. Part C Emerg. Technol.* **2018**, *93*, 463–478. [CrossRef]
16.  Wall Street Journal. Why Your Next Car May Look Like a Living Room. 2017. Available online: http://on.wsj.com/2tlCvYp (accessed on 19 July 2020).
17.  Kok, I.; Zou, S.Y.; Gordon, J.; Mercer, B. Rethinking Transportation 2020–2030: The Disruption of Transportation and the Collapse of the Internal-Combustion Vehicle and Oil Industries, RethinkX. 2017. Available online: http://bit.ly/2pL0cZV (accessed on 19 July 2020).
18.  McKinsey. Automotive Revolution–Perspective Towards 2030: How the Convergence of Disruptive Technology-Driven Trends Could Transform the Auto Industry. 2016. Available online: www.mckinsey.de (accessed on 19 July 2020).
19.  SAE. Levels of Driving Automation Are Defined in New SAE International Standard J3016, Society of Automotive Engineers. 2014. Available online: www.sae.org/misc/pdfs/automated_driving.pdf (accessed on 19 July 2020).
20.  Howard, D.; Dai, D. Public perceptions of self-driving cars: The case of Berkeley, California. *Transp. Res. Board 93rd Annu. Meet.* **2014**, *14*, 1–16.
21.  Elias, W.; Shiftan, Y. The influence of individual's risk perception and attitudes on travel behavior. *Transport. Res. Part A Pol. Pract.* **2012**, *46*, 1241–1251. [CrossRef]
22.  Xu, Z.; Zhang, K.; Min, H.; Wang, Z.; Zhao, X.; Liu, P. What drives people to accept automated vehicles? Findings from a field experiment. *Transport. Res. Part C Emerg. Technol.* **2018**, *95*, 320–334. [CrossRef]
23.  Payre, W.; Cestac, J.; Delhomme, P. Intention to use a fully automated car: Attitudes and a priori acceptability. *Transp. Res. Part F Traffic Psychol. Behav.* **2014**, *27*, 252–263. [CrossRef]
24.  Zhang, T.; Tao, D.; Qu, X.; Zhang, X.; Lin, R.; Zhang, W. The roles of initial trust and perceived risk in public's acceptance of automated vehicles. *Transport. Res. Part C Emerg. Technol.* **2019**, *98*, 207–220. [CrossRef]
25.  Gkartzonikas, C.; Gkritza, K. What have we learned? A review of stated preference and choice studies on autonomous vehicles. *Transport. Res. Part C Emerg. Technol.* **2019**, *98*, 323–337. [CrossRef]
26.  Penmetsa, P.; Adanu, E.K.; Wood, D.; Wang, T.; Jones, S.L. Perceptions and expectations of autonomous vehicles–A snapshot of vulnerable road user opinion. *Technol. Forecast. Soc. Change* **2019**, *143*, 9–13. [CrossRef]
27.  Jahangiri, A.; Rakha, H.; Dingus, T.A. Red-light running violation prediction using observational and simulator data. *Accid. Anal. Prev.* **2016**, *96*, 316–328. [CrossRef]
28.  Kitali, A.E.; Alluri, P.; Sando, T.; Haule, H.; Kidando, E.; Lentz, R. Likelihood estimation of secondary crashes using Bayesian complementary log-log model. *Accid. Anal. Prev.* **2018**, *119*, 58–67. [CrossRef]
29.  Zhu, M.; Li, Y.; Wang, Y. Design and experiment verification of a novel analysis framework for recognition of driver injury patterns: From a multi-class classification perspective. *Accid. Anal. Prev.* **2018**, *120*, 152–164. [CrossRef]
30.  Aghaabbasi, M.; Asadi-Shekari, Z.; Zaly Shah, M.; Olakunle, O.; Armaghani, D.A.; Moeinaddini, M. Predicting the use frequency of ride-sourcing by off-campus university students through random forest and Bayesian network techniques. *Transport. Res. Part A Pol. Pract.* **2020**, *136*, 262–281. [CrossRef]
31.  Lu, H.; Ma, X. Hybrid decision tree-based machine learning models for short-term water quality prediction. *Chemosphere* **2020**, *249*, 126169. [CrossRef]
32.  Chen, T.; Guestrin, C. Xgboost: A Scalable Tree Boosting System. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785e–794.

33. Harb, R.; Yan, X.; Radwan, E.; Su, X. Exploring pre-crash maneuvers using classification trees and random forests. *Accid. Anal. Prev.* **2009**, *41*, 98–107. [CrossRef]
34. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 4768–4777.
35. Quinlan, J.R. *C4.5: Programs for Machine Learning*; Morgan Kaufmann Publishers: Burlington, MA, USA, 1993.
36. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Morgan Kaufmann: Burlington, MA, USA, 2011.
37. Kotsiantis, S.B. Supervised Machine Learning: A Review of Classification Techniques. *Informatica* **2007**, *31*, 249–268.
38. Quinlan, J.R. Improved use of continuous attributes in c4.5. *J. Artif. Intell. Res.* **1996**, *4*, 77–90. [CrossRef]
39. Sanbonmatsu, D.M.; Strayer, D.L.; Yu, Z.; Biondi, F.; Cooper, J.L. Cognitive underpinnings of beliefs and confidence in beliefs about fully automated vehicles. *Transp. Res. Part F* **2018**, *55*, 114–122. [CrossRef]
40. Nordhoff, S.; De Winter, J.; Kyriakidis, M.; Van Arem, B.; Happee, R. Acceptance of driverless vehicles: Results from a large cross-national questionnaire study. *J. Adv. Transp.* **2018**, *22*, 5382192. [CrossRef]
41. Chang, L.Y.; Wang, H.W. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accid. Anal. Prev.* **2006**, *33*, 1019–1027. [CrossRef] [PubMed]
42. Azizi, A.; Pleimling, M. A cautionary tale for machine learning generated configurations in presence of a conserved quantity. *Sci. Rep.* **2021**, *11*, 1–10. [CrossRef] [PubMed]
43. Roshani, M.; Sattari, M.A.; Ali, P.J.M.; Roshani, G.H.; Nazemi, B.; Corniani, E.; Nazemi, E. Application of GMDH neural network technique to improve measuring precision of a simplified photon attenuation based two-phase flowmeter. *Flow Meas. Instrum.* **2020**, *75*, 101804. [CrossRef]
44. Rafiee, P.; Mirjalily, G. Distributed Network Coding-Aware Routing Protocol Incorporating Fuzzy-Logic-BasedForwarders in Wireless Ad hoc Networks. *J. Netw. Syst. Manag.* **2020**, *28*, 1279–1315. [CrossRef]
45. Sanaat, A.; Zaidi, H. Depth of interaction estimation in a preclinical PET scanner equipped with monolithic crystals coupled to SiPMs using a deep neural network. *Appl. Sci.* **2020**, *10*, 4753. [CrossRef]

*Article*

# An Advanced Machine Learning Approach to Predicting Pedestrian Fatality Caused by Road Crashes: A Step toward Sustainable Pedestrian Safety

**Wenlong Tao [1,\*], Mahdi Aghaabbasi [2,\*], Mujahid Ali [3,\*], Abdulrazak H. Almaliki [4], Rosilawati Zainol [2], Abdulrhman A. Almaliki [5] and Enas E. Hussein [6,\*]**

[1] Department of Automotive Technology, Zhejiang Agricultural Business College, Shaoxing 312000, China
[2] Centre for Sustainable Urban Planning and Real Estate (SUPRE), Department of Urban and Regional Planning, Faculty of Built Environment, University of Malaya, Kuala Lumpur 50603, Wilayah Persekutuan Kuala Lumpur, Malaysia; rosilawatizai@um.edu.my
[3] Department of Civil and Environmental Engineering, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Perak, Malaysia
[4] Civil Engineering Department, College of Engineering, Taif University, Taif 21944, Saudi Arabia; a.almaliki@tu.edu.sa
[5] Independent Researcher, Jeddah 12462, Saudi Arabia; a.almaliki1222@gmail.com
[6] National Water Research Center, Shubra El-Kheima 13411, Egypt
[\*] Correspondence: twl87910@163.com (W.T.); mahdi@um.edu.my (M.A.); mujahid_19001704@utp.edu.my (M.A.); enas_el-sayed@nwrc.gov.eg (E.E.H.)

**Abstract:** More than 8000 pedestrians were killed due to road crashes in Australia over the last 30 years. Pedestrians are assumed to be the most vulnerable users of roads. This susceptibility of pedestrians to road crashes conflicts with sustainable transportation objectives. It is critical to know the causes of pedestrian injuries in order to enhance the safety of these vulnerable road users. To achieve this, traditional statistical models are used frequently. However, they have been criticized for their inflexibility in handling outliers and missing or noisy data, and their strict pre-assumptions. This study applied an advanced machine learning algorithm, a Bayesian neural network, which has the characters of both Bayesian theory and neural networks. Several structures of this model were built, and the best structure was selected, which included three hidden neuron layers—sixteen hidden nodes in the first layer and eight hidden nodes in the second and third layers. The performance of this model was compared with the performances of some other machine learning techniques, including standard Bayesian networks, a standard neural network, and a random forest model. The Bayesian neural network model outperformed the other models. In addition, a study on the importance of the features showed that the individuals' characteristics, time, and circumstantial factors were essential. They greatly increased model performance if the model used them. This research lays the groundwork for using machine learning approaches to alleviate pedestrian deaths caused by road accidents.

**Keywords:** pedestrian fatality; road accident; Bayesian neural network; Bayesian theorem; sustainable road network development; machine learning

## 1. Introduction

Pedestrians are the most susceptible road users. Pedestrians also are an important component of the sustainable development of road networks. However, their vulnerability to road crashes conflicts with sustainable transportation objectives. Pedestrian deaths and injuries in road crashes have major socio-economic consequences. This is particularly important in view of developed countries' ongoing efforts to enhance road safety. Since practically anyone can be a pedestrian, pedestrians make up the biggest single road user category. People walk for a variety of reasons, including recreation; traveling to work,

study, or small retailers; and linking up with other means of transportation. In the National Road Safety Strategy, pedestrians are designated as a susceptible road user category. When compared to other road users, they have very limited defence in collisions [1]. Over 50,000 people have died on Australian roads in the last 30 years. Pedestrians accounted for 15.6% of all road accident deaths, even though pedestrians cover fewer miles than other road users. [2]. However, the pedestrian death toll has decreased by almost 57% over the past 30 years. Pedestrians account for a significant portion of fatalities in Australian collisions involving large vehicles and buses. Pedestrians, for example, account for around 30% of those killed in bus collisions [3]. Pedestrians, motorcyclists, and bicyclists make up around a quarter of all deaths in truck crashes [3].

Despite the decrease in pedestrian deaths due to road crashes in Australia, scholars have continued to look for opportunities to acquire a deeper understanding of the factors that impact crash probability in the hopes of effectively estimating the probability of pedestrian-involved crashes and guiding policy initiatives and prevention methods to decrease the incidence of pedestrian-involving crashes [4–7].

There have been several significant data flaws in the literature on pedestrian-related crashes. These problems could lead to erroneous pedestrian crash forecasts and inaccurate conclusions about the factors that cause crashes if analytical models are poorly specified. Imprecision in crash locations and time, challenges in data linkages (for instance, with traffic data) because of database discrepancies, intensity misclassification, errors and incompleteness of affected users' demographics, and wrong identification of accident contributory determinants are just a few of these issues [8]. Furthermore, it is challenging to identify and assess factors influencing pedestrian crash deaths because of the heterogeneity intrinsic in pedestrian crash data, which results from unobservable characteristics that are not recorded by police and cannot be collected from crash reports. As a result of this heterogeneity, parameter estimation may be skewed, leading to possibly inaccurate findings [9–11].

To study the crash data, traditional broadly utilised discrete choice modelling approaches, including mixed logit models, multinomial logit models, ordered logit/probit models, and partial proportional odds logit models have been utilised. Most of the solutions mentioned above, however, rely heavily on pre-existing assumptions. Machine learning (ML) techniques have more flexibility than traditional statistical models in that they can analyse noisy data, outliers, and missing data, without or with minimal previous assumptions about inputs [12–18]. In addition, ML methods are notable instances of data-driven techniques that strive to improve the efficiency and precision of accident data processing and forecasting. Early research employed multiple ML methods, including support vector machines, decision trees, artificial neural network, and ensemble learning, to forecast the severity and frequency of pedestrian-involved crashes, and their findings show that these techniques are very flexible and can outperform conventional methods. Hence, this study selected the ML-based approach in a Bayesian neural network (BNN) to analyse data associated with pedestrian deaths due to road crashes (PDRC).

Due to advancements in computer methods, Bayesian computing approaches are becoming more prominent. Bayesian models offer the privilege of dealing with extremely complicated models, particularly those with difficult-to-calculate probability functions. On the other hand, standard NN models have been criticized for their inability to fit training data accurately, and they may generate forecasted results with undesirable variances [19–22]. Overfitting is among the main causes of this issue. Even if the standard NN model has stronger linear and nonlinear estimation capabilities than traditional statistical approaches, this technique, being vulnerable to the overfitting issue, has poor generalization, which restricts its utility for crash severity and frequency forecasts [19]. In various domains, several earlier studies have shown that applying the Bayesian algorithm in NN models can significantly lessen overfitting while maintaining the NN's excellent nonlinear approximation ability (e.g., [23–25]). However, the combination mentioned above has rarely been used in the domain of crash prediction (e.g., [19]), especially for predicting PDRC.

The main aim of this study was to see how effective the BNN model is for the prediction of PDRC. Furthermore, our study contributes to the area of pedestrian road crash fatality modelling in the following ways: (1) building a combination of architectures to assess the model's performance; (2) evaluating a variety of characteristics that might help with pedestrian fatality classification and forecasting; (3) evaluating BNN in comparison to other machine learning models.

Utilizing data obtained on road transport crash fatalities in Australia, different BNN structures were evaluated to achieve the study's goal. The authors estimated 16 BNN structures and compared their performances utilizing several performance criteria. The authors compared the performance of the best model with other ML models. In addition, the influences of predictors were evaluated using a different approach in which various types of factors were combined to determine the best variable set for predicting PDRC.

The rest of this paper is layed out as follows. A literature review on the methods used for analysing pedestrian crash data is presented in Section 2. Moreover, the necessity of using advanced ML techniques for analysing pedestrian crash data is highlighted in this section. Section 3 presents brief explanations of the methods, performance criteria, and dataset. The feature selection, model development process, selection of the best BNN structure, comparison of the selected BNN model with other ML models, significance of the influential factors, and study limitations are presented in Section 4. The last section presents a summary of the paper and some recommendations for future studies.

## 2. Literature Review

Traditional statistical methods have been employed in the majority of pedestrian-involved crash forecast studies. These models included the ordered probit model [26–30], binary logit model [31], and multinomial logit model (MNL) [29,32–36]. MNL was widely used to study pedestrian crashes; nevertheless, it was criticised since it relies on the assumption that independent variables have the same impacts across instances, which could be contradicted if there are unobserved data heterogeneities. This is a concern because of the incompleteness of the data on road crashes, which means that the impacts may change in different circumstances. Therefore, the mixed logit model was utilised to circumvent the restriction imposed by the independence of irrelevant alternatives (IIA) property by randomly distributing the parameters among individual observations [32,36–39]. Along with the mixed logit model that overcomes the drawbacks of MNL, other models, including partial proportional odds (PPO), also were applied to examine the pedestrian-involved crashes [40–44]. The PPO allows some of the parameter estimates to have different effects on a dependent variable, which is suitable for modelling the pedestrian crash injury severities.

Traditional statistical methods for predicting pedestrian-related crashes are widely used; however, they may become out-of-date if efficacy and accuracy are taken into account. Furthermore, the majority of traditional approaches are regression-based, which include drawbacks such as assuming linear or nonlinear correlations between exploratory factors and the target variable. When such requirements are not satisfied, the models may inadvertently lead to incorrect conclusions [45]. Abreast with the fast evolution of ML techniques and the growing amount of data available, it is becoming increasingly popular to use ML to solve transportation-related issues. In comparison to traditional statistical methods, ML techniques, as non-parametric approaches, have fewer restrictions on pre-existing assumptions regarding the correlations between road accident fatality outcomes and major contributors [46].

Neural networks (NN), random forest, support vector machines (SVMs), decision trees (DTs), and gradient boosting (GB) are among the most frequently used ML techniques for crash data analysis. A list of some studies that have employed the ML techniques for analysing pedestrian crash data is provided in Table 1. It should be noted identifying contributing elements in road crashes is basically a multiclass or binary class problem. Among all ML techniques utilized for the pedestrian crash data, DT-based models, including classification and regression trees (CART), XGBoost, and random forest (RF), were the

most frequently used techniques. Instead, powerful models such as NN were rarely used for analysis of pedestrian-related crash data (e.g., [47]). The standard NN models have been criticized for their vulnerability to overfitting and poor generalization [19]. Consequently, some solutions have been proposed, including combining the Bayesian inference method with the NN algorithm [24,48–50]. This combination allows the neural network to choose hidden neurons and input variables with greater freedom. The BNN model has been successfully applied in many fields of research. However, to date, a very limited number of studies in traffic safety have adopted this model (e.g., [19,51]). These studies mostly attempted to predict motor vehicle collisions and estimate the energy equivalent speed. Given the benefits of NN models over traditional statistical models, along with the gains made by consolidating Bayesian inference into NN, it is worth looking into whether the BNN model can be utilized to model PDRC effectively and whether it outperforms other ML techniques.

**Table 1.** Some studies on the prediction of crashes related to pedestrians using ML techniques.

| Study | Study Aim | ML Technique Employed |
|---|---|---|
| Pour et al. [52] | To determine the impact of temporal, geographical, and personal variables on the severity of vehicle-pedestrian collisions. | DT, KDE |
| Ding et al. [53] | To provide a different perspective on the effects of pedestrian collisions. | MAPRT |
| Mokhtarimousavi [54] | To predict the severity of injuries in pedestrian collisions. | SVM, MNL |
| Das et al. [55] | To create a framework for classifying crash kinds from unstructured textual input using ML algorithms. | RF, SVM, XGBoost |
| Rahimi et al. [56] | To identify death patterns in heavy truck-related pedestrian/bike collisions. | RF, DT |
| Guo et al. [57] | To simulate the issue of categorizing three levels of severity in older pedestrian traffic crashes. | XGBoost |
| Saha and Dumbaugh [58] | To assess the characteristics of the relationships between built environment variables and pedestrian crash frequency at the census block group level. | GB, DT, GAM |
| Zhu [47] | To look into the elements that contribute to the intensity of vehicle-pedestrian collisions at crossings. | CART, GB, RF, ANN, SVM |

Support vector machines = SVM; artificial neural network = ANN; random forest = RF; decision tree = DT; classification and regression trees = CART; kernel density estimation = KDE; multiple additive Poisson regression trees = MAPRT; multinomial logit model = MNL; extreme gradient boosting = XGBoost; generalized additive model = GAM; gradient boosting = GB.

## 3. Methodology

This study primarily aimed at predicting and classifying PDRC using a dataset from Australia. This study employed the BNN algorithm to achieve the objective mentioned above. The flowchart of the investigation is shown in Figure 1. The following sections provide more in-depth descriptions of the stages.
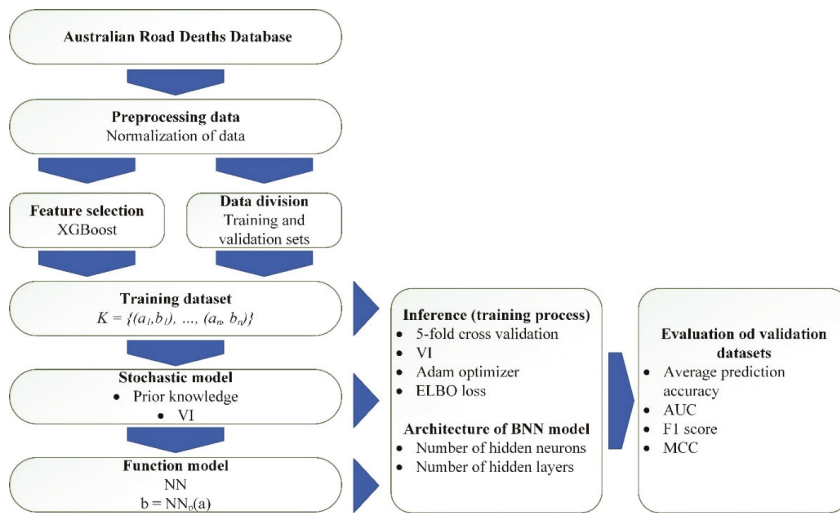
**Figure 1.** A flowchart of this study.

*3.1. A Basic Understanding of the Bayesian Neural Network and Bayesian Inference*

This study utilized a Bayesian method to forecast and classify PDRC. Employing Bayes' theory, Bayesian models attempt to derive and determine characteristics regarding a likelihood distribution from collected data (Equation (1)).

$$P(\alpha|K) = \frac{P(K|\alpha)}{P(K)} \qquad (1)$$

where $\alpha$ is a collection of uncalibrated model parameters, which must be calibrated with dataset $K$. Posterior distribution on $\alpha$ is indicated by $P(\alpha)$, and it reflects our understanding of how data are produced prior to observing them. The posterior distribution, abbreviated as $P(\alpha \,|\, K)$, represents the uncertainty levels of attribute values that accurately describe observed data. The probability function $P(K \,|\, \alpha)$ denotes how likely distinct values of $\alpha$ are to produce the observed dataset $K$. $P(K)$ uses a proper probability density to normalize the posterior distribution.

The use of Bayesian inference in NN has gotten a great deal of interest. This study focuses on expanding the BNN's usage for forecasting and classifying PDRC. A BNN is a NN that has been trained to fit measured values utilizing Bayesian inference, with the assumption that the network's parameters are arbitrary based on a prior probability distribution [49]. In the training stage, various sorts of NN use different approaches to learn from the data and adjust network weights [59]. The weights of a standard NN are regarded deterministic, and then when the model is trained, a single data point approximation is achieved. Contrastingly, instead of assuming a singular point estimation following training, the BNN's weights are expressed as likelihood distributions across feasible data points. The variance of the weights' network distribution reveals the BNN's performance uncertainty. The distinction between a BNN and a deterministic NN is shown in Figure 2.
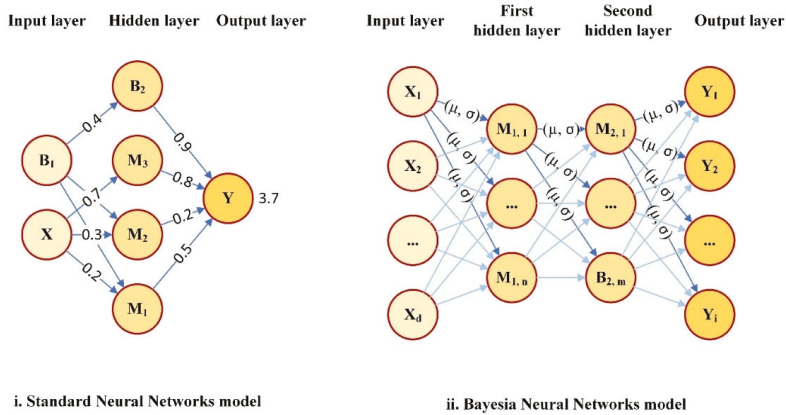
**Figure 2.** Typical structures of NN and BNN.

### 3.2. Bayesian Neural Network

The authors employed a BNN in this research to perform a binary classification between the two tasks—0 = non-pedestrian death and 1 = pedestrian death—while considering data uncertainty. The authors utilize variational inference (VI) to train the BNN, an optimization algorithm for approximating likelihood densities. VI is different from other traditional approaches, such as Markov chain and Monte Carlo, as it determines the parameters of these distributions rather than the weights directly.

The BNN used in this study can be regarded as a probabilistic model $P(b|a, \gamma)$. Here, $b$ is a collection of our categories—$b = 0$ or 1; $a$ is a collection of attributes; $\gamma$ is the weight parameter; $P(b|a, \gamma)$ is a categorical probability. The likelihood function (LF) that is a function of the parameter Y could be generated using the training dataset $K$. The following is the LF:

$$P(K|\gamma) = \prod P(b|a, \gamma) \qquad (2)$$

The maximum likelihood estimate (MLE) of $\gamma$ can be obtained via maximisation of the LF, with the objective function being negative log-likelihood. Based on the Bayes theory, the posterior distribution is proportionate to the outcome of the prior distribution, $P(\gamma)$ and the probability $P(K|\gamma)$. MLE, on the other hand, uses point calculations for parameters; therefore, the uncertainty in the weights is not represented. As a result, a BNN averages forecasts from a number of NN that are weighted according to the posterior distribution of the $\gamma$. The following is the mathematical equation for the posterior predictive distribution:

$$P(b|a, K) = \int P(b|a, \gamma) P(\gamma|K) d\gamma \qquad (3)$$

A BNN can employ a variational distribution $S(\gamma|\vartheta)$ of established functional form to estimate the correct posterior distribution because determining the posterior distribution, $P(\gamma|K)$, is complicated. To accomplish this, the Kullback–Leibler (KL) divergence between the correct posterior $P(\gamma|K)$ and $S(\gamma|\vartheta)$ concerning $\vartheta$ is reduced [60]. The following is the relevant objective function:

$$KL(S(\gamma|\vartheta)||P(\gamma|K)) = E[logS(\gamma|\vartheta] - E[logP(\gamma)] - E[logP(K|\gamma)] + logP(K) \qquad (4)$$

Since the *KL* cannot be determined, this study employs the evidence lower bound (ELBO) that does not comprise the component $logP(K)$ and is the inverse of the *KL* divergence function. Since log $p(K)$ is a constant, it may be ignored, making maximization of the ELBO function equal to minimization of the *KL* divergence. The adaptive moment

estimation (Adam) optimizer is employed to calibrate the variational parameters γ, which can be modified adaptively. The ELBO function's mathematical form is given below.

$$ELBO(S) = E[logP(\gamma)] + E[logP(K|\gamma)] - E[logS(K|\gamma)] \tag{5}$$

### 3.3. Evaluation of Various Models' Performances

This work used the k-fold cross-validation method to arbitrarily divide a whole dataset into five distinct subdivisions with nearly equivalent numbers of data points to avoid biases and overfitting throughout model training. The performances of BNN models in classifying and forecasting pedestrian fatalities due to traffic crashes were assessed using the set of criteria:

- Average training accuracy (ATA): Prediction accuracy in this study's binary class case is defined as the total number of correct forecasts over two classes divided by the total number of forecasts.
- Average F1-score: In the binary-class forecasting study, the average F1-score was employed to approximate criteria for each classification, and the average was calculated by estimating the number of correctly predicted occurrences.
- The area under the ROC curve (AUC): the area under the receiver operating characteristic curve (AUC) was utilised to estimate a scoring classifier at multiple cutoffs in this investigation. The AUC measures a model's ability to distinguish between positive and negative classifications.
- Matthew's correlation coefficient (MCC): The MCC was employed to assess the quality of binary classifications in this investigation. The MCC is a balanced measure that can be utilized even if the categories are of significantly distinct sizes, since it considers true and inaccurate positives and negatives. This criterion is a correlation coefficient that produces a number between −1 and +1 for actual and forecasted binary classes.

This study also used some other common criteria to assess the performance of various BNN architectures. However, the final evaluations and comparisons were based on the four metrics mentioned above. These additional criteria included false discovery rate, false negative rate, false positive rate, negative predictive value, precision, sensitivity, and specificity.

### 3.4. Dataset

The Australian Road Deaths Database (ARDD) provided the data for this research [2]. This database contains information on deaths in road transport crashes in Australia, as provided by the police to state/local road safety bodies monthly. The ARDD collects demographic and crashes information for individuals who died in car accidents in Australia. A road death, often known as a fatality, occurs when an individual dies because of injuries sustained in a car accident within 30 days of the accident. In this dataset, a pedestrian crash is defined as any collision in which a pedestrian is killed, regardless of the number of cars involved. The ARDD includes 24 columns/variables, and 13 of these variables are suitable for predicting pedestrian crashes. It is worth noting that the data utilized in this study were the most up to date, having been collected between 1989 and 2021. This dataset has a sample size of 52,843, and it was used in its entirety to forecast pedestrian fatalities. Table 2 provides a summary of the variables used in this research. This dataset includes basic information about the PDRC. These variables allowed us to achieve the objective of this study, which was applying the combination of Bayesian theory and neural network to pedestrian crash data. Future studies can extend this study by employing datasets with a higher number of variables.

**Table 2.** Summary of the variables employed in this present research.

| Variable and Sub-variable | Description | Value | Mean/Mode |
|---|---|---|---|
| *Individual characteristics (IC)* | | | |
| Age | Age of the individual who was killed (years) | 1–101 | 39.662 |
| Gender | Person's sex | Female, male | Male |
| *Time and occasions (TO)* | | | |
| Month | Month of crash | 1–12 | 12 |
| Day of week | Specifies whether the crash happened on a weekday or on a weekend. | Weekend; weekday | Weekday |
| Time of day | Specifies whether the crash happened during the day or at night. | Day, Night | Day |
| Christmas Period | Specifies whether the crash happened in the 12 days starting on 23 December. | Yes, no | No |
| Easter Period | Specifies whether the crash happened within the five days leading up to Good Friday. | Yes, no | No |
| *Road characteristics (RC)* | | | |
| Speed limit | The designated speed limit at the location of the crash. | 10–130 km | 82.10 |
| National Road Type | | Access Road, Arterial Road, Collector Road, Local Road, National or State Highway, Pedestrian Thoroughfare, Sub-arterial Road | National or State Highway |
| *Crash attributes (CA)* | | | |
| Crash Type | If a pedestrian was died in a collision, it is marked as a pedestrian crash; else, the vehicles engaged is recorded. | Multiple, single | Single |
| Bus involvement | Shows that a bus was involved in the accident. | Yes, no | No |
| Heavy Rigid Truck Involvement | Shows that a heavy rigid truck was involved in the collision. | Yes, no | No |
| Road User (target variable) | Road user type of killed person. | Non-pedestrian, pedestrian | Non-pedestrian |

It is worth mentioning that input variables were normalized and transformed as follows:

- The order of nominal variables was rearranged, with the smallest category appearing first and the largest category appearing last.
- In continuous variables, missing values were substituted with the mean.
- The mode was used to substitute missing values in nominal variables.
- The median was utilized to substitute missing values in ordinal variables.
- The target variable (road user) was initially nominal, and its values included driver, motorcycle pillion passenger, motorcycle rider, passenger, pedal cyclist, pedestrian. The road user was transformed into a binary variable. This new variable included two classes: non-pedestrian death and pedestrian death.

## 4. Results and Discussions

### 4.1. Determination of Significant Variables

This study applied the advanced XGBoost technique to refine irrelevant inputs for a Bayesian-inferred pedestrian death model. It has been proven that the XGBoost method is superior to other non-linear classification methods; however, few studies have applied this technique for feature selection in pedestrian crash prediction and classification (e.g., [57,61]). XGBoost adopts the F-score to determine the significance score (weight) of each variable. A greater F-score is assigned to a variable that embodies more information for classification.

The F-score is calculation using the number of occasions an input is employed for dividing, weighted through the squared enhancement of the model as a consequence of every division, and averaged over all probabilities [62]. This criterion is capable of treating both categorical and continuous inputs fairly to evaluate and rank the inputs. The authors applied the XGBoost technique on 12 variables. Figure 3 illustrates the input rank outcomes organised by their influence. This algorithm selected the ten most important inputs, including speed limit, crash type, age, time of day, bus involvement, gender, day of the week, month, Christmas period, and national road type.
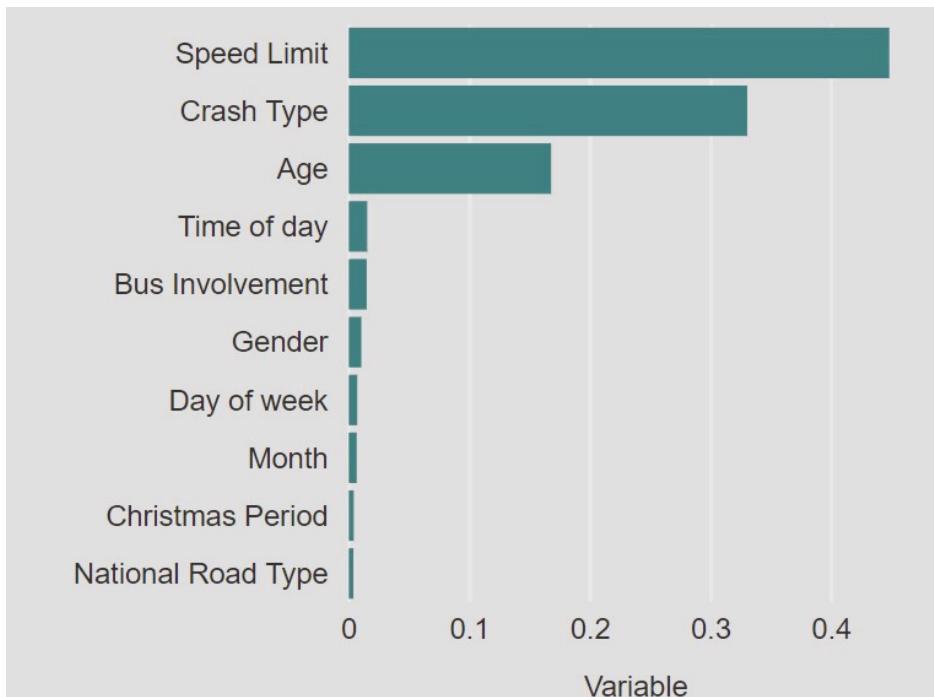


**Figure 3.** Importance of variables analysed by XGBoost technique.

*4.2. Development and Performance Assessment of the BNN Model*

Creating a proper neural network structure is reliant on problems and data. Initially, the authors used a rectified linear unit (ReLU) as the activation function between the consecutive hidden layers to induce non-linearity in the neuron's output. To calculate the error gradient, a batch size of 64 samples from the training dataset was employed. In order to detect the error gradient of the model optimization during the learning stage, various learning rates (LRs) for the Adam optimizer operation were evaluated (10-3, 10-2, 10-1). Then, ELBO loss was observed on validation and training sets. In the prediction of PDRC, Figure 4 shows in what way LRs affected model convergence utilising a BNN model with a single hidden layer (hidden units = 16). Figure 4a illustrates a desirable match, as the validation and training losses rapidly climb to the established position, with little divergence between the two ultimate loss rates. Figure 4b,c shows noisy fluctuations around the training and validation loss, with every iteration moving ahead at an excessively large step size thanks to the high LR. The authors tuned the BNN model utilizing the Adam optimizer's LR of 0.001 to determine the best number of hidden layers and neurons.
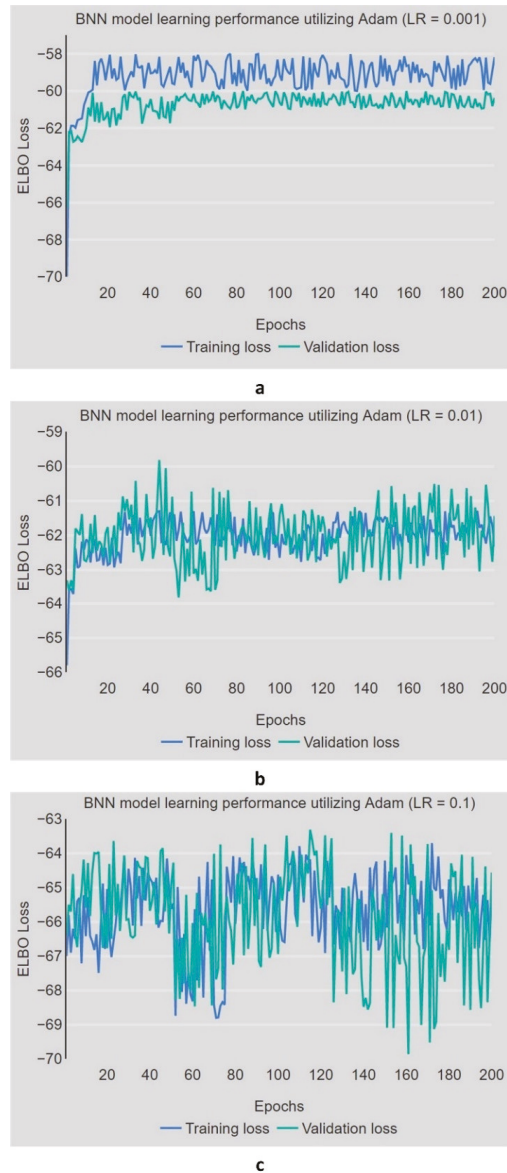
**Figure 4.** Convergence of the BNN model with varying LR. (**a**): LR = 0.001, (**b**): LR = 0.01, (**c**): LR = 0.1.

Various structures of BNN were trained 200 times. Table 3 presents the Bayesian-inferred PDRC model's forecasting performance. The authors evaluated the forecasting performances of several model structures with eleven performance criteria.

**Table 3.** Performances of several BNN designs.

| NS | HL | ATA | Sensitivity | Specificity | Precision | NPV | FPR | FDR | FNR | F1 Score | AUC | MCC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $NS_1$ | 1 | 0.802 | 0.940 | 0.421 | 0.817 | 0.721 | 0.578 | 0.182 | 0.059 | 0.874 | 0.808 | 0.442 |
| $NS_2$ | 1 | 0.834 | 0.844 | 0.171 | 0.986 | 0.015 | 0.828 | 0.013 | 0.155 | 0.909 | 0.748 | 0.005 |
| $NS_3$ | 1 | 0.879 | 0.890 | 0.747 | 0.974 | 0.345 | 0.252 | 0.021 | 0.110 | 0.932 | 0.776 | 0.454 |
| $NS_4$ | 1 | 0.887 | 0.900 | 0.752 | 0.974 | 0.413 | 0.247 | 0.025 | 0.100 | 0.939 | 0.780 | 0.503 |
| $NS_5$ | 2 | 0.607 | 0.967 | 0.271 | 0.553 | 0.900 | 0.728 | 0.446 | 0.032 | 0.704 | 0.811 | 0.329 |
| $NS_6$ | 2 | 0.570 | 0.972 | 0.256 | 0.504 | 0.922 | 0.743 | 0.4951 | 0.027 | 0.664 | 0.811 | 0.312 |
| $NS_7$ | 2 | 0.670 | 0.9605 | 0.303 | 0.636 | 0.858 | 0.696 | 0.363 | 0.039 | 0.765 | 0.804 | 0.361 |
| $NS_8$ | 2 | 0.767 | 0.947 | 0.379 | 0.766 | 0.771 | 0.620 | 0.233 | 0.052 | 0.847 | 0.796 | 0.419 |
| $NS_9$ | 2 | 0.784 | 0.945 | 0.399 | 0.791 | 0.750 | 0.600 | 0.208 | 0.055 | 0.861 | 0.822 | 0.432 |
| $NS_{10}$ | 3 | 0.867 | 0.879 | 0.687 | 0.976 | 0.274 | 0.313 | 0.023 | 0.120 | 0.925 | 0.824 | 0.377 |
| **$NS_{11}$** | **3** | **0.894** | **0.906** | **0.774** | **0.975** | **0.453** | **0.225** | **0.0245** | **0.093** | **0.939** | **0.844** | **0.540** |
| $NS_{12}$ | 3 | 0.792 | 0.928 | 0.399 | 0.816 | 0.661 | 0.600 | 0.183 | 0.071 | 0.868 | 0.788 | 0.396 |
| $NS_{13}$ | 3 | 0.781 | 0.944 | 0.394 | 0.786 | 0.750 | 0.605 | 0.213 | 0.055 | 0.858 | 0.734 | 0.426 |
| $NS_{14}$ | 3 | 0.717 | 0.936 | 0.334 | 0.711 | 0.750 | 0.665 | 0.289 | 0.063 | 0.808 | 0.719 | 0.353 |
| $NS_{15}$ | 3 | 0.707 | 0.949 | 0.334 | 0.687 | 0.811 | 0.665 | 0.312 | 0.050 | 0.797 | 0.717 | 0.376 |
| $NS_{16}$ | 3 | 0.668 | 0.946 | 0.304 | 0.640 | 0.811 | 0.696 | 0.359 | 0.053 | 0.764 | 0.719 | 0.336 |

NS = network structure; NS1 = 16; NS2 = 32; NS3 = 64; NS4 = 128; NS5 = (16, 8); NS6 = (16, 16); NS7 = (32, 8); NS8 = (32, 16); NS9 = (32, 32); NS10 = (8, 8, 8); NS11 = (16, 8, 8); NS12 = (16, 16, 8); NS13 = (32, 8, 8); NS14 = (32, 16, 8); NS15 = (32, 32, 16); NS16 = (64, 32, 16). HL = hidden layers; AAT = average training accuracy; NPV = negative predictive value; FPR = false positive rate; FDR = false discovery rate; FNR = false negative rate; AUC = area under curve; MCC = Matthews's correlation coefficient.

Concerning ATA, the BNN with three hidden neuron layers (NS11) had the best results (ATA = 0.894). The second best ATA belonged to a BNN architecture including a hidden layer of 128 elements (namely, NS4). NS5 and NS6 were the two poorest network architectures. Regarding AUC, F1 score, and MCC, NS11 also outperformed other BNN structures, which indicated the model's success in classifying PDRC.

The BNN design with three hidden layers (NS11) performed reasonably well, with sixteen hidden neurons in the first layer and eight hidden neurons each in the second and third layers. As a result, this research focuses on this BNN model in the subsequent sections to see how the model's classification uncertainties affect the forecasts of PDRC.

*4.3. Quantification of Ambiguity in the Forecast and Classifying Probability*

A Sankey plot was built to depict the relationship between actual and forecasted labels to understand the classification errors of the BNN model (Figure 5). The actual classes are represented by the left nodes on the Sankey plot, whereas the forecasted classes are displayed by the right-hand points. The thicknesses of the color connections and streams are proportional to the amounts of data. As seen in Figure 5, non-pedestrian deaths (class 0) were mainly predicted to be non-pedestrian deaths, with only a few being misclassified as pedestrian deaths (class 1). However, more than half of pedestrian deaths (54.6%) were incorrectly predicted as non-pedestrian deaths. The proposed BNN's classification of the "non-pedestrian death" class is superior to that of the "pedestrian death" class with forecast rates of 97.5% and 45.37% accuracy, according to the comparison of forecast performance across each category.
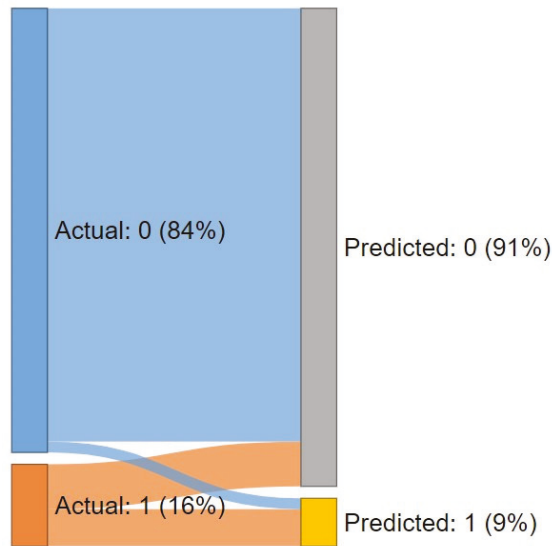
**Figure 5.** Correct and wrong forecasts classified by the BNN model.

The Bayesian technique has two notable features: (1) it yields predictive class probabilities rather than deterministic class label forecasts, and (2) it produces the standard deviation of the posterior prediction to indicate the level of uncertainty. The findings are shown as a raincloud graphic, which mixes a data distribution depiction and box plots overlaid on jittered raw data. For two death categories, Figure 6 depicts the range of the predictive probabilities and the forecast uncertainty. As can be seen in thick regions, the probability values for both classes are predominantly concentrated in the great probability zones that are in the range of 0.8–1.0. Both classes' prediction uncertainties are highly aggregated in the range of 0.0 and 0.1, indicating a low level of ambiguity. Overall, the BNN had a great level of confidence in classifying both death classes.
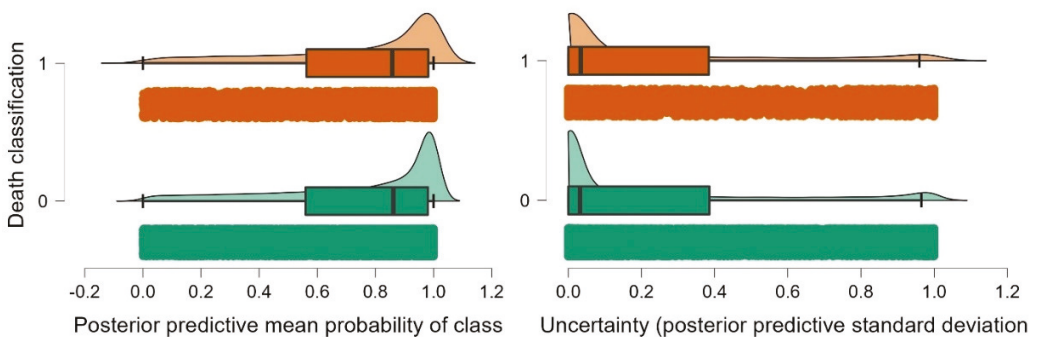


**Figure 6.** All class labels' posterior predictive mean probabilities and uncertainties.

*4.4. Variable Significance*

When performing field research, knowing the impacts of variables on a model's predictive ability can lower the cost of gathering data on PDRC. Assessing the significance of all specified traits and their conceivable combinations, on the other hand, is time-intensive and computationally costly. In this investigation, ten XGBoost-selected variables were categorised according to the kinds to which they related. This study built eleven combinations

in which different types of factors were combined to identify the best variable combination. Simultaneously, the model's performance was analysed in order to determine the smallest number of variables that must be collected while maintaining reliable prediction performance. Table 4 presents the outcomes of the models' executions. The outcomes of this analysis showed that ARR8 (TO + RC) was the weakest combination. In contrast, ARR7 (IC + TO + RC + CA) was the best combination, followed by ARR6 (IC + RC + CA). These findings imply that the combination of factors related to the time, occasions, and road characteristics is not able to predict the PDRC accurately alone. The predictions based on these two types of data should be improved using other factors, such as individual characteristics and crash attributes. The findings of this study are in line with those of Onieva-García et al. [63], Toran Pour et al. [64], Park and Ko [65], Li and Fan [44], and Kim et al. [66], who confirmed the significant roles of age and gender in pedestrian-related crashes and deaths. Several studies also confirmed the effects of bus involvement on the risks of injury and death of pedestrians (e.g., [67–69]), which indicates the significant role of crash attributes in the prediction of PDRC. Overall, when personal characteristics and crash features are factored in, this model appears to be successful and accurate.

**Table 4.** The performance of the BNN model with various variable arrangements.

| Arrangement | Combination of Variables | ATA | F1 Score | AUC | MMC |
|---|---|---|---|---|---|
| ARR$_1$ | IC + TO | 0.845 | 0.915 | 0.591 | 0.1375 |
| ARR$_2$ | IC + RC | 0.855 | 0.918 | 0.698 | 0.324 |
| ARR$_3$ | IC + CA | 0.858 | 0.920 | 0.761 | 0.317 |
| ARR$_4$ | IC + TO + RC | 0.860 | 0.921 | 0.754 | 0.3527 |
| ARR$_5$ | IC + TO + CA | 0.863 | 0.923 | 0.762 | 0.361 |
| ARR$_6$ | IC + RC + CA | 0.890 | 0.937 | 0.788 | 0.526 |
| ARR$_7$ | IC + TO + RC + CA | 0.894 | 0.939 | 0.844 | 0.540 |
| ARR$_8$ | TO + RC | 0.844 | 0.915 | 0.500 | 0 |
| ARR$_9$ | TO + CA | 0.846 | 0.916 | 0.695 | 0.113 |
| ARR$_{10}$ | TO + RC + CA | 0.866 | 0.924 | 0.774 | 0.387 |
| ARR$_{11}$ | RC + CA | 0.847 | 0.916 | 0.762 | 0.127 |

ARR = arrangement; AAT = average training accuracy; AUC = area under curve; MCC = Matthews's correlation coefficient; IC = individuals' characteristics; TO = time and occasions; RC = road characteristics; CA = crash attributes.

### 4.5. Comparison of BNN Modes with Other ML Models

The authors of this study compared the BNN model with various ML models, including a random forest (RF), a standard Bayesian network (BN), and a standard neural network (NN). This comparison helps with determining which machine learning algorithm has the highest prediction accuracy. It will help to reduce future work spent on selecting acceptable methods for PDRC data analysis. The advantages of using BNN for PDRC prediction are further confirmed by this comparison. The outcomes of this comparison are presented in Table 5. This comparison shows that the BNN model outperformed the other models, especially the standard NN model. Additionally, the standard BN model showed a poor prediction performance compared with the other models. The RF model showed a desirable performance that can be rooted in its capabilities for ensembling weak learners [70]. This study's Bayesian-inferred pedestrian fatality model performs well in prediction and classification based on the presented results.

**Table 5.** Comparisons of the BNN model with other ML models used to predict the PDRC.

| Model | ATA | F1 Score | AUC | MMC |
|---|---|---|---|---|
| BNN | 0.89 | 0.94 | 0.84 | 0.54 |
| RF | 0.87 | 0.92 | 0.79 | 0.60 |
| BN | 0.85 | 0.91 | 0.80 | 0.36 |
| NN | 0.84 | 0.91 | 0.53 | 0.07 |

*4.6. Limitations and Future Enhancements*

The Australian Road Deaths Database (ARDD) was employed to create and test Bayesian inference with NN for forecasting and classifying road-related pedestrian deaths. However, there are a few drawbacks to be aware of, and potential enhancements for the future. Even though the Bayesian-inferred pedestrian fatality model outperformed traditional ML models, BNN, like many other ML techniques, is a data-driven modelling approach, and the ARDD contains little variety and skewed distributions. This suggests that in certain severe circumstances, the model would be unstable. Future investigations are required to improve this model by consolidating a more varying set of environmental factors, built environment factors, and road characteristics (e.g., weather conditions, use patterns, and road widths), as past studies have confirmed their usefulness (e.g., [71–73]).

While this study was effective at using a BNN model to predict PDRC, it is important to remember that the performances of ML models vary depending on the data. If the data are within the range of the current study's data, the results of this study can be replicated. Future research could use this technique, possibly with some tweaks, to analyse other datasets and present their findings. It enables a valid assessment of the BNN's ability to forecast PDRC.

Several prior studies also have found that walking behaviors can have a role in pedestrian fatalities as a result of road crashes. When it comes to pedestrian-involved collisions, the pedestrian crossing pattern is one of the most essential features of walking behaviour [74]. Pedestrians who were tragically wounded or admitted to hospital were typically crossing unlawfully and/or at fault, according to prior research (e.g., [33,75]). However, ARDD does not capture pedestrian activities at the moment of a collision, such as crossing and use of a mobile phone. The ARDD must include a wide variety of characteristics of both sides, vehicles and pedestrians, to gain a deep understanding of the reasons behind pedestrian-involved crashes.

**5. Conclusions**

The Australian Road Deaths Database was employed to train the BNN model to generate sound pedestrian death forecasts based on individuals' characteristics, time, occasions, road characteristics, and crash attributes in this study. For every road crash fatality class, this study created BNN models, including various structures, to assess their performances and to examine their corresponding predictive ambiguities. Below is a summary of this study's findings:

- The BNN model, which consists of three hidden neuron layers with sixteen hidden nodes in the first layer, provided the best training accuracy of 0.894. Its posterior predictive probabilities are in the great probability range, and the predictive ambiguity is tightly concentrated in the 0–0.1 range.
- BNN model outperformed RF, BN, and NN models.
- Personal characteristics and time and occasion factor groups are clearly essential, greatly boosting the performance of the model if they are used as inputs.
- Individually, the most important parameters in PDRC prediction were the speed limit, collision type, and age.

The following are some practical implications of the major results that may be of interest to both academics and practitioners in the domain.

For pedestrian safety on special occasions, such as Christmas and Easter, specific effective pedestrian safety strategies should be implemented. These policies may assist pedestrians in using roads safely and developing sustainable commute habits.

The speed limit has emerged as the most important factor for predicting pedestrian fatalities due to road crashes. It is obvious that increasing vehicle speed raises the collision risk exponentially [76]. According to Australian and worldwide case reports, lowering the posted maximum speed on rural roads by 10 km/h reduces the chance of an accident by 20–25%. Furthermore, after the removal of unrestricted speeds in some highways, the Australian road mortality database reveals that there was a 3.4 per year decline in fatalities

on highways with speed restrictions of 110 km/h and above. In Australia, for every person killed on the road, another 23 persons are injured as a result of an accident, highlighting the social benefit of any speed restriction lowering [77].

Another key factor in predicting pedestrian fatalities due to traffic collisions was age. Several prior studies have found that senior pedestrians are more prone to pedestrian crossing collisions. Elder pedestrian crashes are more likely to occur in congested metropolitan locations, and older pedestrians are more likely to be at fault because of their incapacity to manage complicated traffic scenarios, such as crossing roads [78]. These problems can be avoided if government agencies and licensing departments enhance crossing safety by reducing intersection ambiguity, increasing visibility, increasing conspicuity, and eliminating right-hand turns that require gap acceptance decisions. In addition, they can install or retrofit systems that defend pedestrians in locations where there is a significant risk of pedestrian fatality, such as high-pedestrian-activity places.

Our BNN model is capable of predicting future PDRC accurately, and it has a low level of predictive uncertainty. Although further research is needed in this area, the methods utilised in this study could be employed as a starting point for finding pedestrian risk determinants and developing appropriate legislation.

## References

1.  Australian Transport Council (ATC). *National Road Safety Strategy 2011–2020*; Australian Transport Council: Canberra, Australia, 2011.
2.  Department of Infrastructure Regional Development and Cities. Australian Road Deaths Database. 2021. Available online: https://www.bitre.gov.au/statistics/safety/fatal_road_crash_database (accessed on 13 November 2021).
3.  Bureau of Infrastructure, Transport and Regional Economics. *Road Trauma Involving Heavy Vehicles 2018 Crash Statistical Summary*; BITRE: Canberra, Australia, 2020.
4.  Zegeer, C.V.; Bushell, M. Pedestrian crash trends and potential countermeasures from around the world. *Accid. Anal. Prev.* **2012**, *44*, 3–11. [CrossRef] [PubMed]
5.  Anderson, R.; Ponte, G.; Doecke, S. *A Survey of Bullbar Prevalence at Pedestrian Crash Sites in Adelaide, South Australia*; Centre for Automotive Safety Research: Adelaide, Australia, 2008.
6.  Samerei, S.A.; Aghabayk, K.; Shiwakoti, N.; Karimi, S. Modelling bus-pedestrian crash severity in the state of Victoria, Australia. *Int. J. Inj. Control Saf. Promot.* **2021**, *28*, 233–242. [CrossRef] [PubMed]
7.  Arnold, P.; Rosman, D.; Thornett, M. Pedestrian crash risk in Western Australia for both pedestrians and drivers. *Road Transp. Res.* **1992**, *1*, 60–75.
8.  Imprialou, M.; Quddus, M. Crash data quality for road safety research: Current state and future directions. *Accid. Anal. Prev.* **2019**, *130*, 84–90. [CrossRef]
9.  Mannering, F.L.; Bhat, C.R. Analytic methods in accident research: Methodological frontier and future directions. *Anal. Methods Accid. Res.* **2014**, *1*, 1–22. [CrossRef]
10. Shaheed, M.S.; Gkritza, K. A latent class analysis of single-vehicle motorcycle crash severity outcomes. *Anal. Methods Accid. Res.* **2014**, *2*, 30–38. [CrossRef]
11. Sun, M.; Sun, X.; Shan, D. Pedestrian crash analysis with latent class clustering method. *Accid. Anal. Prev.* **2019**, *124*, 50–57. [CrossRef]

12. Aghaabbasi, M.; Shekari, Z.A.; Shah, M.Z.; Olakunle, O.; Armaghani, D.J.; Moeinaddini, M. Predicting the use frequency of ride-sourcing by off-campus university students through random forest and Bayesian network techniques. *Transp. Res. Part A Policy Pract.* **2020**, *136*, 262–281. [CrossRef]

13. Qian, Y.; Aghaabbasi, M.; Ali, M.; Alqurashi, M.; Salah, B.; Zainol, R.; Moeinaddini, M.; Hussein, E.E. Classification of Imbalanced Travel Mode Choice to Work Data Using Adjustable SVM Model. *Appl. Sci.* **2021**, *11*, 11916. [CrossRef]

14. Aghaabbasi, M.; Shah, M.Z.; Zainol, R. Investigating the Use of Active Transportation Modes Among University Employees Through an Advanced Decision Tree Algorithm. *Civ. Sustain. Urban Eng.* **2021**, *1*, 26–49. [CrossRef]

15. Ali, M.; de Azevedo, A.R.G.; Marvila, M.T.; Khan, M.I.; Memon, A.M.; Masood, F.; Almahbashi, N.M.Y.; Shad, M.K.; Khan, M.A.; Fediuk, R.; et al. The Influence of COVID-19-Induced Daily Activities on Health Parameters—A Case Study in Malaysia. *Sustainability* **2021**, *13*, 7465. [CrossRef]

16. Ali, M.; Dharmowijoyo, D.B.; Harahap, I.S.; Puri, A.; Tanjung, L.E. Travel behaviour and health: Interaction of Activity-Travel Pattern, Travel Parameter and Physical Intensity. *Solid State Technol.* **2020**, *63*, 4026–4039.

17. Ali, M.; Dharmowijoyo, D.B.E.; de Azevedo, A.R.G.; Fediuk, R.; Ahmad, H.; Salah, B. Time-Use and Spatio-Temporal Variables Influence on Physical Activity Intensity, Physical and Social Health of Travelers. *Sustainability* **2021**, *13*, 12226. [CrossRef]

18. Chen, Y.; Aghaabbasi, M.; Ali, M.; Anciferov, S.; Sabitov, L.; Chebotarev, S.; Nabiullina, K.; Sychev, E.; Fediuk, R.; Zainol, R. Hybrid Bayesian Network Models to Investigate the Impact of Built Environment Experience before Adulthood on Students' Tolerable Travel Time to Campus: Towards Sustainable Commute Behavior. *Sustainability* **2022**, *14*, 325. [CrossRef]

19. Xie, Y.; Lord, D.; Zhang, Y. Predicting motor vehicle collisions using Bayesian neural network models: An empirical analysis. *Accid. Anal. Prev.* **2007**, *39*, 922–933. [CrossRef]

20. Ali, M.; Abbas, S.; Salah, B.; Akhter, J.; Saleem, W.; Haruna, S.; Room, S.; Abdulkadir, I. Investigating Optimal Confinement Behaviour of Low-Strength Concrete through Quantitative and Analytical Approaches. *Materials* **2021**, *14*, 4675. [CrossRef] [PubMed]

21. Ali, M.; Room, S.; Khan, M.I.; Masood, F.; Ali Memon, R.; Khan, R.; Memon, A.M. Assessment of local earthen bricks in perspective of physical and mechanical properties using Geographical Information System in Peshawar, Pakistan. *Structures* **2020**, *28*, 2549–2561. [CrossRef]

22. De Azevedo, A.R.G.; Marvila, M.T.; Ali, M.; Khan, M.I.; Masood, F.; Vieira, C.M.F. Effect of the addition and processing of glass polishing waste on the durability of geopolymeric mortars. *Case Stud. Constr. Mater.* **2021**, *15*, e00662. [CrossRef]

23. Liu, T.; Liu, Y.; Liu, J.; Wang, L.; Xu, L.; Qiu, G.; Gao, H. A Bayesian learning based scheme for online dynamic security assessment and preventive control. *IEEE Trans. Power Syst.* **2020**, *35*, 4088–4099. [CrossRef]

24. Marzban, C.; Witt, A. A Bayesian neural network for severe-hail size prediction. *Weather Forecast.* **2001**, *16*, 600–610. [CrossRef]

25. Yan, D.; Zhou, Q.; Wang, J.; Zhang, N. Bayesian regularisation neural network based on artificial intelligence optimisation. *Int. J. Prod. Res.* **2017**, *55*, 2266–2287. [CrossRef]

26. Zajac, S.S.; Ivan, J.N. Factors influencing injury severity of motor vehicle–crossing pedestrian crashes in rural Connecticut. *Accid. Anal. Prev.* **2003**, *35*, 369–379. [CrossRef]

27. Rifaat, S.M.; Chin, H.C. Accident severity analysis using ordered probit model. *J. Adv. Transp.* **2007**, *41*, 91–114. [CrossRef]

28. Obeng, K.; Rokonuzzaman, M. Pedestrian injury severity in automobile crashes. *Open J. Saf. Sci. Technol.* **2013**, *3*, 33341. [CrossRef]

29. Kwigizile, V.; Sando, T.; Chimba, D. Inconsistencies of ordered and unordered probability models for pedestrian injury severity. *Transp. Res. Rec.* **2011**, *2264*, 110–118. [CrossRef]

30. Yasmin, S.; Eluru, N. Evaluating alternate discrete outcome frameworks for modeling crash injury severity. *Accid. Anal. Prev.* **2013**, *59*, 506–521. [CrossRef]

31. Sze, N.-N.; Wong, S. Diagnostic analysis of the logistic model for pedestrian injury severity in traffic crashes. *Accid. Anal. Prev.* **2007**, *39*, 1267–1278. [CrossRef]

32. Kim, S.; Ulfarsson, G.F. Traffic safety in an aging society: Analysis of older pedestrian crashes. *J. Transp. Saf. Secur.* **2019**, *11*, 323–332. [CrossRef]

33. Ulfarsson, G.F.; Kim, S.; Booth, K.M. Analyzing fault in pedestrian–motor vehicle crashes in North Carolina. *Accid. Anal. Prev.* **2010**, *42*, 1805–1813. [CrossRef]

34. Tay, R.; Choi, J.; Kattan, L.; Khan, A. A multinomial logit model of pedestrian–vehicle crash severity. *Int. J. Sustain. Transp.* **2011**, *5*, 233–249. [CrossRef]

35. Zhou, Z.-P.; Liu, Y.-S.; Wang, W.; Zhang, Y. Multinomial logit model of pedestrian crossing behaviors at signalized intersections. *Discret. Dyn. Nat. Soc.* **2013**, *2013*, 172726. [CrossRef]

36. Chen, Z.; Fan, W. Modeling pedestrian injury severity in pedestrian-vehicle crashes in rural and urban areas: Mixed logit model approach. *Transp. Res. Rec.* **2019**, *2673*, 1023–1034. [CrossRef]

37. Kim, J.-K.; Ulfarsson, G.F.; Shankar, V.N.; Mannering, F.L. A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accid. Anal. Prev.* **2010**, *42*, 1751–1758. [CrossRef]

38. Haleem, K.; Alluri, P.; Gan, A. Analyzing pedestrian crash injury severity at signalized and non-signalized locations. *Accid. Anal. Prev.* **2015**, *81*, 14–23. [CrossRef]

39. Tulu, G.S.; Washington, S.; Haque, M.M.; King, M.J. Injury severity of pedestrians involved in road traffic crashes in Addis Ababa, Ethiopia. *J. Transp. Saf. Secur.* **2017**, *9*, 47–66. [CrossRef]

40. Rifaat, S.M.; Tay, R.; de Barros, A. Urban street pattern and pedestrian traffic safety. *J. Urban Des.* **2012**, *17*, 337–352. [CrossRef]

41. Sasidharan, L.; Menéndez, M. Partial proportional odds model—An alternate choice for analyzing pedestrian crash injury severities. *Accid. Anal. Prev.* **2014**, *72*, 330–340. [CrossRef]
42. Pour, A.T.; Moridpour, S.; Tay, R.; Rajabifard, A. A partial proportional odds model for pedestrian crashes at mid-blocks in Melbourne metropolitan area. *MATEC Web Conf.* **2016**, *81*, 02020. [CrossRef]
43. Li, Y.; Fan, W.D. Modelling severity of pedestrian-injury in pedestrian-vehicle crashes with latent class clustering and partial proportional odds model: A case study of North Carolina. *Accid. Anal. Prev.* **2019**, *131*, 284–296. [CrossRef]
44. Li, Y.; Fan, W. Pedestrian injury severities in pedestrian-vehicle crashes and the partial proportional odds logit model: Accounting for age difference. *Transp. Res. Rec.* **2019**, *2673*, 731–746. [CrossRef]
45. Chang, L.-Y.; Chen, W.-C. Data mining of tree-based models to analyze freeway accident frequency. *J. Saf. Res.* **2005**, *36*, 365–375. [CrossRef]
46. Gong, Y.; Abdel-Aty, M.; Cai, Q.; Rahman, M.S. A decentralized network level adaptive signal control algorithm by deep reinforcement learning. In Proceedings of the Transportation Research Board 98th Annual Meeting, Washington, DC, USA, 13–17 January 2019.
47. Zhu, S.Y. Analyse vehicle-pedestrian crash severity at intersection with data mining techniques. *Int. J. Crashworthiness* **2021**, 1–9. [CrossRef]
48. Mackay, D.J.C. Bayesian Methods for Adaptive Models. Ph.D. Thesis, California Institute of Technology, Pasadena, CA, USA, 1992.
49. Neal, R.M. *Bayesian Learning for Neural Networks*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012; Volume 118.
50. Liang, F. Bayesian neural networks for nonlinear time series forecasting. *Stat. Comput.* **2005**, *15*, 13–29. [CrossRef]
51. Riviere, C.; Lauret, P.; Ramsamy, J.M.; Page, Y. A Bayesian neural network approach to estimating the energy equivalent speed. *Accid. Anal. Prev.* **2006**, *38*, 248–259. [CrossRef]
52. Pour, A.T.; Moridpour, S.; Rajabifard, A.; Tay, R. Spatial and temporal distribution of pedestrian crashes in Melbourne metropolitan area. *Road Transp. Res.* **2017**, *26*, 4–20.
53. Ding, C.; Chen, P.; Jiao, J.F. Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: A machine learning approach. *Accid. Anal. Prev.* **2018**, *112*, 116–126. [CrossRef]
54. Mokhtarimousavi, S. A Time of Day Analysis of Pedestrian-Involved Crashes in California: Investigation of Injury Severity, a Logistic Regression and Machine Learning Approach Using HSIS Data. *ITE J.-Inst. Transp. Eng.* **2019**, *89*, 25–33.
55. Das, S.; Le, M.; Dai, B.Y. Application of machine learning tools in classifying pedestrian crash types: A case study. *Transp. Saf. Environ.* **2020**, *2*, 106–119. [CrossRef]
56. Rahimi, A.; Azimi, G.; Asgari, H.; Jin, X. Injury Severity of Pedestrian and Bicyclist Crashes Involving Large Trucks. In Proceedings of the ASCE International Conference on Transportation and Development (ASCE ICTD), Seattle, WA, USA, 26–29 May 2020; pp. 110–122.
57. Guo, M.; Yuan, Z.; Janson, B.; Peng, Y.; Yang, Y.; Wang, W. Older pedestrian traffic crashes severity analysis based on an emerging machine learning XGBoost. *Sustainability* **2021**, *13*, 926. [CrossRef]
58. Saha, D.; Dumbaugh, E. Use of a model-based gradient boosting framework to assess spatial and non-linear effects of variables on pedestrian crash frequency at macro-level. *J. Transp. Saf. Secur.* **2021**, *13*, 32. [CrossRef]
59. Wang, H.; Yeung, D.-Y. Towards Bayesian deep learning: A framework and some existing methods. *IEEE Trans. Knowl. Data Eng.* **2016**, *28*, 3395–3408. [CrossRef]
60. Blei, D.M.; Kucukelbir, A.; McAuliffe, J.D. Variational inference: A review for statisticians. *J. Am. Stat. Assoc.* **2017**, *112*, 859–877. [CrossRef]
61. Wu, S.; Yuan, Q.; Yan, Z.; Xu, Q. Analyzing Accident Injury Severity via an Extreme Gradient Boosting (XGBoost) Model. *J. Adv. Transp.* **2021**, *2021*, 3771640. [CrossRef]
62. Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
63. Onieva-García, M.Á.; Martínez-Ruiz, V.; Lardelli-Claret, P.; Jiménez-Moleón, J.J.; Amezcua-Prieto, C.; de Dios Luna-del-Castillo, J.; Jiménez-Mejías, E. Gender and age differences in components of traffic-related pedestrian death rates: Exposure, risk of crash and fatality rate. *Inj. Epidemiol.* **2016**, *3*, 14. [CrossRef]
64. Toran Pour, A.; Moridpour, S.; Tay, R.; Rajabifard, A. Influence of pedestrian age and gender on spatial and temporal distribution of pedestrian crashes. *Traffic Inj. Prev.* **2018**, *19*, 81–87. [CrossRef]
65. Park, S.; Ko, D. Investigating the Factors Influencing Pedestrian–Vehicle Crashes by Age Group in Seoul, South Korea: A Hierarchical Model. *Sustainability* **2020**, *12*, 4239. [CrossRef]
66. Kim, J.-K.; Ulfarsson, G.F.; Shankar, V.N.; Kim, S. Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Accid. Anal. Prev.* **2008**, *40*, 1695–1702. [CrossRef]
67. Park, H.-C.; Joo, Y.-J.; Kho, S.-Y.; Kim, D.-K.; Park, B.-J. Injury severity of bus–pedestrian crashes in South Korea considering the effects of regional and company factors. *Sustainability* **2019**, *11*, 3169. [CrossRef]
68. Li, P.; Abdel-Aty, M.; Yuan, J. Using bus critical driving events as surrogate safety measures for pedestrian and bicycle crashes based on GPS trajectory data. *Accid. Anal. Prev.* **2021**, *150*, 105924. [CrossRef]

69. Sivasankaran, S.K.; Balasubramanian, V. Severity of Pedestrians in Pedestrian-Bus Crashes: An Investigation of Pedestrian, Driver and Environmental Characteristics Using Random Forest Approach. In Proceedings of the Congress of the International Ergonomics Association, Online, 13–18 June 2021; pp. 825–833.
70. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
71. Shawky, M.; Garib, A.; Al-Harthei, H. The impact of road and site characteristics on the crash-injury severity of pedestrian crashes. *Adv. Transp. Stud.* **2014**, *1*, 27–36.
72. Zhai, X.; Huang, H.; Sze, N.; Song, Z.; Hon, K.K. Diagnostic analysis of the effects of weather condition on pedestrian crash severity. *Accid. Anal. Prev.* **2019**, *122*, 318–324. [CrossRef] [PubMed]
73. Ukkusuri, S.; Miranda-Moreno, L.F.; Ramadurai, G.; Isa-Tavarez, J. The role of built environment on pedestrian crash frequency. *Saf. Sci.* **2012**, *50*, 1141–1151. [CrossRef]
74. Jain, A.; Gupta, A.; Rastogi, R. Pedestrian crossing behaviour analysis at intersections. *Int. J. Traffic Transp. Eng.* **2014**, *4*, 103–116. [CrossRef]
75. Kucerová, I.; Sulíková, B.; Poláčková, K. Pedestrian accidents-actual trend in the Czech Republic. *Trans. Transp. Sci.* **2013**, *6*, 145. [CrossRef]
76. Aarts, L.; Van Schagen, I. Driving speed and the risk of road crashes: A review. *Accid. Anal. Prev.* **2006**, *38*, 215–224. [CrossRef]
77. Read, D.J. Open speeds on Northern Territory roads: Not so fast. *Med. J. Aust.* **2015**, *203*, 14–15. [CrossRef]
78. Oxley, J.; Whelan, M. It cannot be all about safety: The benefits of prolonged mobility. *Traffic Inj. Prev.* **2008**, *9*, 367–378. [CrossRef]

*Article*

# Nonlinear Relationships between Vehicle Ownership and Household Travel Characteristics and Built Environment Attributes in the US Using the XGBT Algorithm

**Te Ma [1],\*, Mahdi Aghaabbasi [2],\*, Mujahid Ali [3],\*, Rosilawati Zainol [2], Amin Jan [4],\*, Abdeliazim Mustafa Mohamed [5,6] and Abdullah Mohamed [7]**

1   School of Tourism, Dalian University, Dalian 116000, China
2   Centre for Sustainable Urban Planning and Real Estate (SUPRE), Department of Urban and Regional Planning, Faculty of Built Environment, University of Malaya, Kuala Lumpur 50603, Malaysia; rosilawatizai@um.edu.my
3   Department of Civil and Environmental Engineering, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia
4   Faculty of Hospitality, Tourism and Wellness. Universiti Malaysia Kelantan, City Campus, Pengkalan Chepa 16100, Malaysia
5   Department of Civil Engineering, College of Engineering, Prince Sattam Bin Abdulaziz University, Alkharj 16273, Saudi Arabia; a.bilal@psau.edu.sa
6   Building & Construction Technology Department, Bayan College of Science and Technology, Khartoum 210, Sudan
7   Research Centre, Future University in Egypt, New Cairo 11745, Egypt; mohamed.a@fue.edu.eg
\*   Correspondence: mate987654321@163.com (T.M.); mahdi@um.edu.my (M.A.); mujahid_19001704@utp.edu.my (M.A.); aminjan@umk.edu.my (A.J.)

**Abstract:** In the United States, several studies have looked at the association between automobile ownership and sociodemographic factors and built environment qualities, but few have looked at household travel characteristics. Their interactions and nonlinear linkages are frequently overlooked in existing studies. Utilizing the 2017 US National Household Travel Survey, the authors employed an extreme gradient boosting tree model to evaluate the nonlinear and interaction impacts of household travel characteristics and built environment factors on vehicle ownership in three states of the United States (California, Missouri, and Kansas) that are different in population size. To develop these models, three main XGBT parameters, including the number of trees, maximal depth, and minimum rows, were optimized using a grid search technique. In California, the predictability of vehicle ownership was driven by household travel characteristics (cumulative importance: 0.62). Predictions for vehicle ownership in Missouri and Kansas were dominantly influenced by sociodemographic factors (cumulative importance: 0.53 and 0.55, respectively). In all states, the authors found that the number of drivers in a household plays a vital role in the vehicle ownership decisions of households. Regarding the built environment attributes, deficiencies in cycling infrastructure were the most prominent attribute in predicting household vehicle ownership in California. This variable, however, has threshold connections with vehicle ownership, but the magnitude of these relationships is small. The outcomes imply that improving the condition of cycling infrastructure will help reduce the number of vehicles. In addition, incentives that encourage the households' drivers not to buy new vehicles are helpful. The outcomes of this study might aid policymakers in developing policies that encourage sustainable vehicle ownership in the United States.

**Keywords:** sustainable vehicle ownership; nonlinear relationships; built environment; XGBT

## 1. Introduction

In the United States, each household has an average of 1.88 vehicles [1]. In 2017, the rate of households with no vehicle was roughly 9%, implying that over 90% of families had

access to at least one light vehicle [1]. The growing use of automobiles has resulted in a slew of severe consequences, including traffic jams, pollution, and poor health outcomes [2,3]. In the United States, figuring out how to slow the rise of vehicle ownership has now become a pressing concern.

Planners from all over the globe have offered measures to improve the built environment (e.g., [4–8]). Past research has shown that certain aspects of the built environment are connected to vehicle ownership, supporting the proposal. Some of these aspects include the condition of the cycling and walking environment [9], population density [10–13], urban area size [14,15], and type of living area [14,16]. Vehicle ownership is typically viewed as a result of a household's demographic and socioeconomic profile [17]. Several investigations utilized monthly or average income to predict vehicle ownership. Home ownership, size of the household, number of children, adults, and employees in the household have all been identified as crucial determinants of vehicle ownership [17–23].

While sociodemographic and built-environment attributes have been widely utilized to predict vehicle ownership globally, studies have rarely employed household travel characteristics indicators, such as household drivers' count, household members' count on the trip, and household vehicle used on the trip. These variables are important because they can be assumed as indicators of independent trips. Independent trips mean each household member may have a different life responsibility and, in turn, different travel needs. Thus, they might be encouraged to buy more vehicles.

According to past studies, the majority of built environment variables exhibit nonlinear relationships with vehicle ownership [24–30]. Some recent studies reveal that a considerable number of built environment variables have threshold relationships with vehicle ownership, and the nonlinear trends are inconsistent (e.g., [28]). Nonlinear relationships may help policymakers comprehend the influence of a variety of built-environment characteristics on vehicle ownership, and it will be interesting to see if this discovery holds true in various urban and rural settings. This aids policymakers and planners in fine-tuning their plans. Despite the fact that the nonlinear relationships between the built environment variable and vehicle ownership have been assessed by several studies, only a very limited number of studies has evaluated the relationships between household travel characteristics and vehicle ownership. As a result of this information, policymakers may be able to give households incentives to drive less.

Several advanced machine learning techniques and mathematical formulations have been used to solve different engineering and planning problems [31–40]. To keep abreast with the advancement of machine learning techniques and their vast applications across the world, the authors utilize extreme gradient boosting trees (XGBT) to examine the main determinants of vehicle ownership and highlight their nonlinear interactions, employing data from the 2017 US National Household Travel Survey (NHTS). The following are the questions that this research aims to answer: (1) How important are built environment attributes, household travel characteristics, and sociodemographic characteristics in influencing household vehicle ownership decisions in the United States? (2) Does vehicle ownership have a nonlinear relationship with household travel characteristics and built environment factors? (3) To what extent do key household travel characteristics mediate the links between key built-environment variables and vehicle ownership?

This research adds to the literature in three main ways. Initially, it adds to the research of vehicle ownership in several US states with diverse populations. Furthermore, this research evaluates the significance of several factors in determining car ownership, as well as the relevance of policy and incentive implementation in various US states. It also demonstrates that the majority of household travel characteristics and built environment variables have inconsistently nonlinear connections, bolstering the scant body of evidence and providing recommendations for planning approaches in US states. Lastly, this research shows how important household travel characteristics, as well as their interactions with major built-environment variables, play a significant role in limiting vehicle expansion in each state.

To the best of the authors' knowledge, to date, no study has employed XGBT to reveal the complex relationships between built environment attributes, household travel characteristics, and sociodemographic characteristics in predicting household vehicle ownership. This research assists policymakers in providing families with motivation to reduce their vehicle ownership. In addition, this study can show the capabilities of the XGBT algorithm to reveal the complex relationships between various variables in transportation science.

The following sections make up the remainder of this paper. A literature overview of research that used NHTS data for various purposes is included in Section 2. The modeling method, data, and variables are introduced in Section 3. The results of the models used in this investigation are described in Section 4. The findings, implications, and limitations of the study are discussed in Section 5. The final section outlines the most important findings.

## 2. Background: Employment of NHTS Dataset

The National Household Travel Survey (NHTS) is the official source on the travel behaviour of the American public, which is carried out by the Federal Highway Administration (FHWA). These data are the singular national dataset that allows the study of personal and household travel patterns. It encompasses non-commercial travel on a daily basis in all commute modes and the features of the travellers, their households, and their vehicles. Several researchers employed these data for different purposes, including investigation of trends in taxi use and ride hailing [41–43], determining the occurrence of rural and urban cycling [44,45], ownership and usage assessment of unconventional fuel vehicles [46], preferences of public transportation users [47], and so on. A summary of some studies that used 2017 NHTS data is shown in Table 1.

**Table 1.** Some recent studies that employed 2017 NHTS data.

| Study | Study Aim | Variables Used | Analysis Technique(s) |
|---|---|---|---|
| Conway, Salon, and King [41] | To report on taxi usage patterns and the rise of ride-hailing services. | Sociodemographic, personal trips | Descriptive analysis, logistic regression |
| Godfrey et al. [48] | To address some of the most pressing concerns affecting public transit. | Sociodemographic | Descriptive analysis |
| Li, Liu, and Jia [46] | To look at the current state of conventional car ownership and usage, as well as renewable fuels vehicle ownership and consumption. | Sociodemographic | Descriptive analysis |
| Tribby and Tharp [44] | To determine the prevalence of cycling patterns by city, as well as the features that best distinguish cyclists from non-cyclists. | Sociodemographic | Logistic regression |
| Das [43] | To determine the impact of ride hailing service uptake on sustainable mobility options. | Sociodemographic, built environment attributes | Logistic regression |
| Jiao, Bischak, and Hyden [42] | To determine the effect of shared mobility on trip production. | Sociodemographic, built environment attributes | Negative binomial (NB) model |
| Porter, Kontou, McDonald, and Evenson [45] | To describe the overall impediments to riding as self-reported. | Sociodemographic | Descriptive analysis |
| Sadeghvaziri and Tawfik [49] | To learn more about how the elderly travel. | Sociodemographic | Descriptive analysis |
| Jin and Yu [47] | To gain a better understanding of the fundamental reasons why people avoid taking public transportation by looking at the viewpoints of various users. | Sociodemographic, descriptive analysis | |
| Sabouri, Tian, Ewing, Park, and Greene [5] | Using regional household travel data and constructed environmental characteristics from 32 regions across the United States, vehicle ownership models were assessed. | Sociodemographic, built environment attributes | Logistic regression |

The information presented in Table 1 reveals some shortcomings in the employment of NHTS data. First, a few studies predicted vehicle ownership using these data (e.g., [5]). Second, a very limited number of variables was used to perform the analysis by different studies. For example, no indicator of household travel characteristics (e.g., household drivers' count, household members' count on the trip, and household vehicle used on the trip) was used by these studies. In addition, a very limited number of built environment attributes were employed (e.g., the condition of walking and cycling infrastructure). Third, a narrow range of statistical analyses were employed by different studies. Most studies used traditional statistical analysis techniques or simple descriptive analysis. Traditional statistical methods such as regression models have strict assumptions regarding the quality of the data. In addition, these methods do not reveal the nonlinear relationships between the target variable and inputs effectively. Lastly, most studies did not differentiate between the US states or cities in terms of population size or other characteristics, which may cause some serious differences in the prediction results.

### 3. Methodology

#### 3.1. Extreme Gradient Boosting (XGBT)

The XGBT model was used to determine the primary correlates of vehicle ownership and their complex relationships. XGBT was originally developed for data science [50], but it has also been used increasingly in urban planning and transportation science (e.g., [51–53]). The XGBT algorithm is a more regularized variant of the gradient boosting tree (GBT). In comparison to the GBT, the XGBT is better at generalization and takes less time to train [54]. Additionally, GBT and XGBT are better than traditional statistical methods (e.g., linear regression) in a number of ways. Firstly, they outperform conventional techniques in terms of data fitting. Secondly, they are capable of dealing with a variety of different sorts of data, such as categorical and continuous. Thirdly, they are insensitive to outliers and can deal with incomplete data in a flexible manner. Fourth, they help solve the problem of multicollinearity [28,55]. Furthermore, GBT and XGBT may fit any irregular connection between variables, and modelers are not required to estimate their correlations in advance. According to previous research, vehicle ownership has a nonlinear connection with the factors that are associated with it, and the complex patterns vary according to the factor [28]. While traditional statistical approaches may describe nonlinear interactions via variable transformation, the transformation is ineffective due to the irregular nonlinearity.

Owing to its advantages of high reliability and considerable flexibility, XGBT, an advanced supervised method presented by Chen and Guestrin [50] under the Gradient Boosting architecture, has been well acknowledged in Kaggle machine learning contests. XGBT's loss function provides an extra regularization term to the objective function that attempts to smooth the ultimate learning weights and prevent over-fitting [50]. To optimize the loss function, it furthermore employs 1st or 2nd order gradient statistics. Additionally, XGBT enables row and column sample selection to address this problem, in addition to providing regular terms to avoid over-fitting. Because parallel and distributed computation allow for rapid learning, faster model exploration is conceivable. The XGBT architecture will be simply described in the subsequent paragraphs.

The aggregate of the prediction scores, $f_m(a_i)$ of all trees can be represented as the predicted output $\hat{b}_i$ of the XGBT model:

$$\hat{b}_i = \sum_{m=1}^{M} f_m(a_i), f_m \in \gamma \tag{1}$$

where $\gamma$ represents the regression trees' space, $M$ shows the regression trees' number, and the attributes associated with sample $i$ are denoted by $a_i$. Every leaf node $j$ in a particular dataset has a forecast score $f_m(a_i)$, commonly referred to as leaf weight. $s_j$ is the leaf weight and regression values of entire samples at this leaf node $j$, where $j \in \{1, 2, \dots Q\}$. In the tree, the leaves' number is shown by $Q$.

In machine learning issues, objective functions become the most fundamental expression, and the boosting process repeats until the objective function minimization is limited in order to estimate the number of functions used in the model, which establishes the regularized objective function as follows:

$$\theta = \sum_{i=1}^{h} z\left(b_i, \hat{b}_i\right) + \alpha Q + \frac{1}{2}\beta \sum_{j=1}^{Q} s_j^2 \tag{2}$$

where, $h$ is the number of data samples provided, and $\sum_{i=1}^{h} z(b_i, \hat{b}_i)$ is the training loss function that describes how well the model fits the training sets. For punishing the model's complexity, $\alpha Q + \frac{1}{2}\beta \sum_{j=1}^{Q} s_j^2$ is a regularization term. The complexity cost of adding an extra leaf is $\alpha$, the regularization hyper-parameter is $\beta$, and the L2 norm of leaf node $j$ weights is $s_j^2$ in the regularization term.

Every recently introduced tree learns from its previous trees and adjusts the residuals in the estimated values in the incremental learning procedure. As a result, all of the trees' iteration outcomes have already been included in $\hat{b}_i^{(m-1)}$. Consequently, $\hat{b}_i^{(m)}$ can denote $\hat{b}_i^{(m-1)} + f_m(a_i)$ for the $m$th repetition, and the objective function "C" is represented as:

$$\theta = \sum_{i=1}^{h} z(b_i, \hat{b}_i^{(m-1)} + f_m(a_i)) + \alpha Q + \frac{1}{2}\beta \sum_{j=1}^{Q} s_j^2 \tag{3}$$

The 2nd order Taylor expansion is employed to optimize the objective effectively in the general situation for the first term loss training function.

$$\theta_m \simeq \sum_{i=1}^{h} \left[ z(b_i, \hat{b}_i^{(m-1)} + d_i f_m(a_i)) + \frac{1}{2} e_i f_m^2(a_i) \right] + \alpha Q + \frac{1}{2}\beta \sum_{j=1}^{Q} s_j^2 \tag{4}$$

where $d_i = \sigma_{\hat{b}^{(m-1)}} z(a_i, \hat{b}_i^{(m-1)})$ and $e_i = \sigma_{\hat{b}^{(m-1)}}^2 z(a_i, \hat{b}_i^{(m-1)})$ are the loss function's first and second-order gradient statistics. In step m, the constant terms can be subtracted to obtain the approximate objective:

$$\theta_m \simeq \sum_{i=1}^{h} \left[ d_i f_m(a_i) + \frac{1}{2} e_i f_m^2(a_i) \right] + \alpha Q + \frac{1}{2}\beta \sum_{j=1}^{Q} s_j^2 \tag{5}$$

A tree is characterized as a vector of scores in branches and a leaf index mapping function that transfers an instance to a leaf $j$, and this procedure is written as $\sum_{i=1}^{h} f_m(a) = \sum_{j=1}^{Q} s_j$ and Equation (5) can be rephrased as:

$$\theta_{(m)} = \sum_{j=1}^{Q} \left[ \left( \sum_{i \in I_j} g_i \right) s_j + \frac{1}{2} \left( \sum_{i \in I_j} e_i + \alpha \right) s_j^2 \right] + \beta Q \tag{6}$$

With a fixed tree structure, quadratic function programming is used to select the perfect branch weight scores on every leaf node $s_j^*$ as well as the extreme value of $\theta_{(m)}^*$:

$$s_j^* = -\frac{\sum_{i \in I_j} d_i}{\sum_{i \in I_j} e_i + \beta} \tag{7}$$

$$\theta_{(m)}^* = -\frac{1}{2} \sum_{j=1}^{Q} \frac{\left( \sum_{i \in I_j} d_i \right)^2}{\sum_{i \in I_j} e_i + \beta} + \alpha Q \tag{8}$$

Equation (8) is a framework scoring function that determines the suitability of a specified vector of leaf scores. A lower value is preferable since it fits the data more effectively. In practical uses, a greedy method has been used to discover an ideal tree structure to prevent an endless number of alternative tree architectures. To develop an XGBT model, it is important to fine-tune three main XGBT parameters, including the number of trees, maximal depth, and minimum rows. Once we have trained the XGBT model, it is possible to evaluate the significance of every predictor in forecasting the response. In addition, XGBT can assess the partial dependence and association between predictors and target variables after controlling for other variables in the model. Chen and Guestrin [50] provide more thorough descriptions of the XGBT algorithm.

### 3.2. Data

The data come from the 2017 National Household Travel Survey (NHTS), which is conducted by the US Department of Transportation [56]. The 2017 NHTS is the 8th in a series of nationally representative cross-sectional surveys on the daily commute conducted at random times. Data were gathered from stratified random samples of households in the United States. The 2017 NHTS consists of two main processes: (1) a mail-based household recruiting survey that gathered data on the household, transport, and travel behavior; and (2) a predominantly web-based person-level retrieval survey that asked about travel on a study-assigned day.

There were 458 variables in this dataset. As previously mentioned, the main goal of this present study is to reveal the nonlinear relationship between the count of household vehicles (vehicle ownership) and sociodemographic, household travel characteristics, and built environment attributes. Consequently, based on literature, only variables that were related to household vehicle ownership were employed. These variables and their descriptions are shown in Table 2. It is worth stating that there are a limited number of built-environment variables in the NHTS. For example, only two variables, namely "reasons for not walking more = infrastructure" and "reasons for not biking more = infrastructure," assessed the condition of walking and cycling environments. Thus, the authors considered these variables as two indicators of the condition of the walking and cycling environments. Finally, 14 variables were used as inputs in this study's analysis, and one variable, household vehicle counts, was used as the target variable.

In this study, the authors evaluated different states with different populations. To this end, three categories of the population were considered: (1) high-population states, (2) medium-population states, and (3) low-population states. Regarding the population of US states, the authors used the United States Census Bureau [57] as the principal source. As previously stated, a list of US states that was provided by the United States Census Bureau was used. The states in this list were sorted by population. Then, this list was simply divided into the three categories. In each category, the state that had the highest population was selected. For the first category, California (CA) was selected. Missouri (MO) was selected for the second category. For the low-population states, Kansas (KS) was selected. The authors then selected 5000 samples in each state. This sampling approach prevents any bias resulting from over- or under-sampling. A flowchart of this study is presented in Figure 1.

**Table 2.** Variables' description.

| Variable | Description | Value |
|---|---|---|
| *Independent variable* | | |
| HHVEHCNT | Household vehicles' count | [0–12] |
| *Sociodemographic (SD)* | | |
| HHFAMINC | Household income ($) | (1) <10,000; (2) 10,000–14,999; (3) 15,000–24,999; (4) 25,000–34,999; (5) 35,000–49,999; (6) 50,000–74,999; (7) 75,000–99,999; (8) 100,000–124,999; (9) 125,000–149,999; (10) 150,000–199,999; (11) >200,000 |
| HHSIZE | Household members' count | [1–13] |
| HOMEOWN | Home ownership | (1) own; (2) rent |
| NUMADLT | Count of adults in the household over the age of 18 | [1–10] |
| WRKCOUNT | Household workers' count | [1–7] |
| YOUNGCHILD | Count of children aged 0 to 4 in the household | [1–5] |
| *Household travel characteristics (HTC)* | | |
| DRVRCNT | Household drivers' count | [0–9] |
| TRPHHACC | Household members' count on the trip | [0–10] |
| TRPHHVEH | Household vehicle used on trip | (1) yes; (2) no |
| *Built environment attributes (BEA)* | | |
| BIKEINFRA | Deficiencies in cycling infrastructure * | (1) no adjacent paths or trails; (2) no sidewalks or sidewalks are in poor condition; (3) no adjacent parks; (4) 1 and 2; (5) 1 and 3; (6) 2 and 3; (7) 1, 2, and 3 |
| HBPPOPDN | Category of population density (persons per sqmi) in the household's home census block group | 50 = 0–99; 300 = 100–499; 750 = 500–999; 1500 = 1000–1999; 3000 = 2000–3999; 7000 = 4000–9999; 17,000 = 10,000–24,999; 30,000 = 25,000–999,999 |
| URBANSIZE | Size of the urban area in which the residence is located | (1) 50,000–199,999; (2) 200,000–499,999; (3) 500,000–999,999; (4) 1 million or more without heavy rail; (5) 1 million or more with heavy rail; (6) not in urbanized area |
| URBRUR | Household in urban/rural area | (1) urban; (2) rural |
| WALKIFRA | Deficiencies in walking infrastructure * | (1) no adjacent paths or trails; (2) no sidewalks or sidewalks are in poor condition; (3) no adjacent parks; (4) 1 and 2; (5) 1 and 3; (6) 2 and 3; (7) 1, 2, and 3 |

* These variables were originally employed in the NHTS to evaluate reasons for not walking or cycling, and their acronyms are WALK_DEF and BIKE_DFR, respectively.
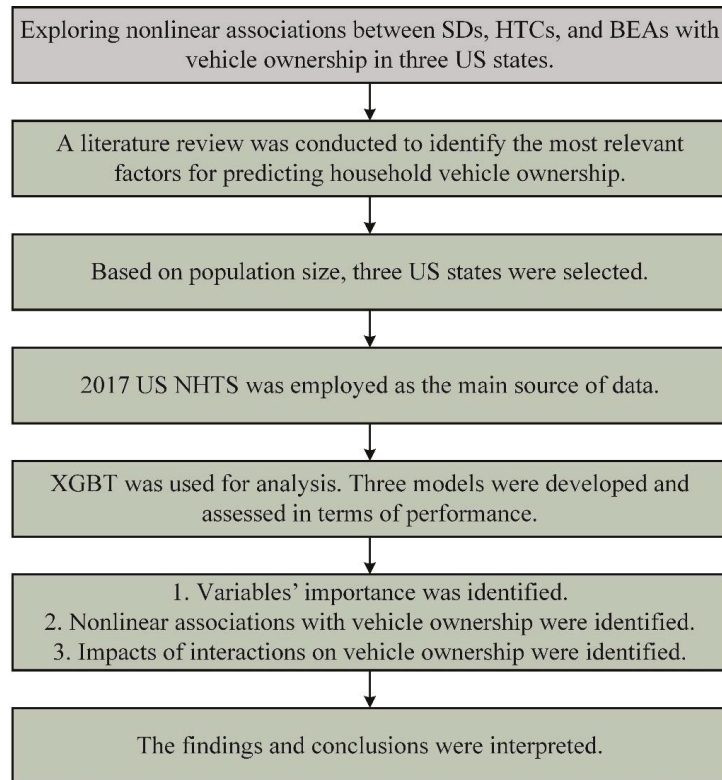
**Figure 1.** Flowchart of this study.

## 4. Results

*4.1. Nonlinear Models Development and Performance Assessment*

One XGBT model was constructed for each of the three US states based on population differences in this study. These three models were developed using a set of parameters, and each of these parameters has its own value. This study employed the grid search technique to discover the optimized value of these parameters. Table 3 shows the optimum values of the XGBT models' parameters.

**Table 3.** Values of key parameters of XGBT models in three US states.

| Parameter | CA | MO | KS |
|---|---|---|---|
| Number of trees | 1 | 70 | 80 |
| Maximal depth | 10 | 80 | 60 |
| Minimum rows | $4.9 \times 10^{-324}$ | $4.9 \times 10^{-324}$ | $4.9 \times 10^{-324}$ |

To develop the XGBT models, the data were divided into training and testing sets with a ratio of 80:20. In addition, to avoid overfitting and reduce the generalization error, this study employed a 10-fold cross validation approach. The performance of these three models was evaluated using two famous performance criteria, including linear correlation

(*R*) and mean absolute error (*MAE*). Equations (9) and (10) illustrate the mathematical forms of these criteria.

$$R = \frac{\sum_{i=1}^{h}\left(k_i - \bar{k}_i\right)(s_i - \bar{s}_i)}{\sqrt{\sum_{i=1}^{n}\left(k_i - \bar{k}_i\right)^2 (s_i - \bar{s}_i)^2}} \tag{9}$$

$$MAE = \frac{\sum_{i=1}^{n}|k_i - s_i|}{h} \tag{10}$$

where $k_i$ and $s_i$ signify nth actual and predicted values, respectively; $\bar{k}_i$ and $\bar{s}_i$ indicate the average values of actual and predicted values, respectively; $h$ shows the number of samples in the dataset. Table 4 shows the outcomes of the models' evaluations. As can be seen, the highest training performance belonged to Kansas.

**Table 4.** XGBT models' performance.

| Criterion | | CA | MO | KS |
|---|---|---|---|---|
| R | Train | 0.814 | 0.934 | 0.995 |
| | Test | 0.817 | 0.935 | 0.965 |
| MAE | Train | 0.664 | 0.303 | 0.246 |
| | Test | 0.662 | 0.308 | 0.244 |

*4.2. Variables' Importance*

Table 5 shows the cumulative importance (CI) of all variables in forecasting vehicle ownership. In California, household travel characteristics were the most influential factors in predicting vehicle ownership (CI: 0.62). In Missouri and Kansas, sociodemographic factors were the most important predictors of household vehicle ownership (CI: 0.53 and 0.55, respectively).

**Table 5.** Cumulative importance of variables for predicting vehicle ownership in three states of the US.

| State | Cumulative Importance | | |
|---|---|---|---|
| | **Sociodemographic (SD)** | **Built Environment Attributes (BEA)** | **Household Travel Characteristics (HTC)** |
| CA | 0.08 | 0.30 | 0.62 |
| MO | 0.53 | 0.12 | 0.35 |
| KS | 0.55 | 0.20 | 0.25 |

Figure 2 shows the variables' importance in three different states of the US with different population sizes for vehicle ownership prediction. The number of drivers in a household (B) was the most important variable in California and Missouri. The importance of the number of drivers in a household was slightly lower in Kansas than that of home ownership (F).

In California, the second most important variable for the prediction of vehicle ownership was deficiencies in cycling infrastructure, followed by deficiencies in walking infrastructure. Several variables, including the count of adults in a household over the age of 18, household vehicle used on the trip, household members' count on the trip, count of person trips on travel day, household living area (urban or rural), and count of children aged 0 to 4 in the household, had no contribution to vehicle ownership prediction in California.
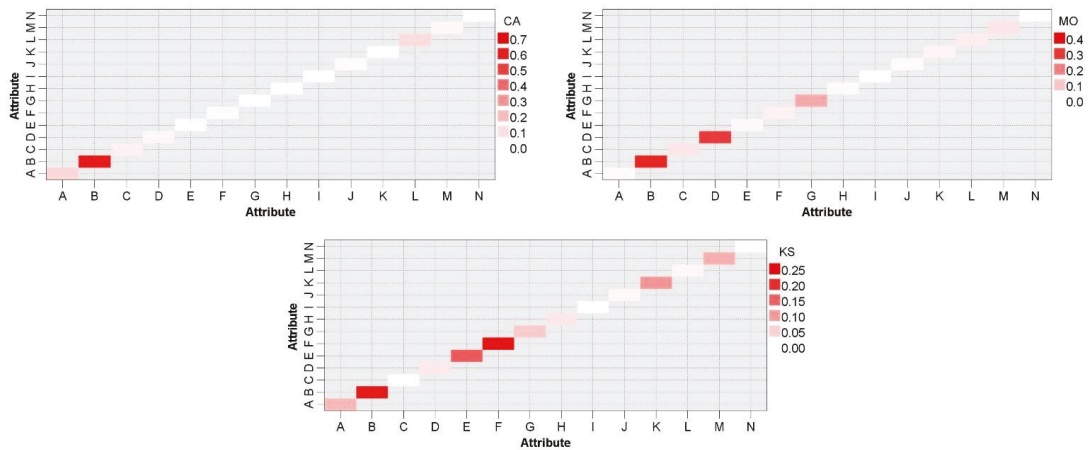
**Figure 2.** Importance of variables by different US states: (A = BIKEINFRA; B = DRVRCNT; C = HBPPOPDN; D = HHFAMINC; E = HHSIZE; F = HOMEOWN; G = NUMADLT; H = TRPHHACC; I = TRPHHVEH; J = URBANSIZE; K = URBRUR; L = WALK_DEF; M = WRKCOUNT; N = YOUNGCHILD).

Household income, which was followed by the number of adults in the household over the age of 18, was the second most influential variable in Missouri for predicting vehicle ownership. The number of children aged 0 to 4 in the household and the household vehicle used on the trip had no effect on the car ownership prediction in Missouri.

As mentioned above, in Kansas, home ownership was the most important variable, and the second most important variable for vehicle ownership forecasting was household drivers' count, followed by household members' count. In Kansas, the zero-contributed variables included population density, the number of children aged 0 to 4 in the household, the number of person's trips on the travel day, and the household vehicle used on the trip.

### 4.3. Nonlinear Associations with Car Ownership

The nonlinear associations between the predicted number of household vehicles and each state's two most important variables are provided in this section. Figure 3 shows associations between predicted household vehicle counts and various variables in three different US states.

In California, there is a cubic relationship between the number of drivers in the household (DRVRCNT) and the household vehicle count. When the number of household drivers is within the range of two, it has a negligible effect on vehicle ownership. Beyond the threshold, it has a positive relationship with vehicle ownership. However, when the DRVRCNT exceeds six, the impact of the DRVRCNT is saturated. The cubic relationships for Missouri and Kansas are different. In Missouri, when the DRVRCNT is in the range of one to four drivers, it has a strong positive relationship with vehicle ownership. However, when the DRVRCNT is beyond four drivers, vehicle ownership starts to decrease. In Kansas, the cubic relationship between vehicle ownership and DRVRCNT is predominantly concave between 1 and 3 drivers. It seems that when the DRVRCNT exceeds four, the impact of the DRVRCNT is saturated in Kansas. Overall, the best range of DRVRCNT for cutting down on car ownership in California is between zero and two drivers. This range for Missouri and Kansas is between four and five. These findings corroborate prior research indicating that the number of drivers in a household has a considerable impact on vehicle ownership (e.g., [58–61]). No study, however, has examined the nonlinear relationship between the number of drivers in a household and vehicle ownership. As a result, the findings from this research are unique.
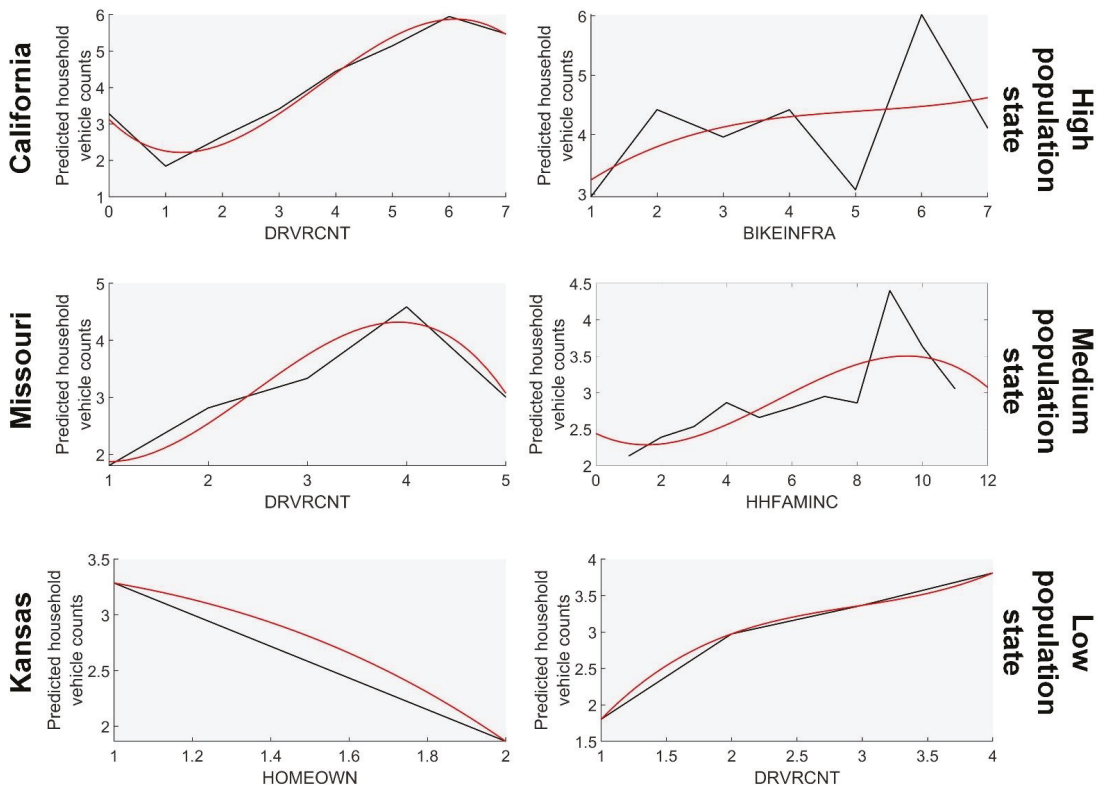
**Figure 3.** Relationships between predicted household vehicle counts and various variables in three different US states.

In the cubic relationship between vehicle ownership and deficits in cycling infrastructure (BIKEINFRA), it seems that when the BIKEINFRA is within the range of three, its impact on vehicle ownership is greater than when it is within the range of four to seven. This means that when Californian households are disappointed to find adjacent paths, trails, sidewalks, or parks, they lose their inclination to bike and switch to buying new vehicles. Several studies confirmed that providing adequate infrastructure for biking may encourage people to substitute this mode for private vehicles, but to the authors' best knowledge, very few studies have assessed the influence of this factor on vehicle ownership. In addition, no study has specifically examined the nonlinear relationship between these factors and vehicle ownership.

In Missouri, the cubic connection between household income (HHFAMINC) and vehicle ownership indicates that when household income is between 10,000 and 14,999 USD, it has a minor influence on vehicle ownership. It has a positive correlation with vehicle ownership after the threshold is exceeded. When the HHFAMINC crosses nine (125,000–149,999 USD), the HHFAMINC's effect becomes saturated. Several previous studies reported the positive and linear relationship between household income and vehicle ownership (e.g., [19]), but very few studies have assessed the nonlinear relationship between household income and vehicle ownership (e.g., [62]).

In Kansas, there is a strong link between home ownership (HOMEOWN) and the number of vehicles in a household, so possessing a home increases the likelihood of owning more vehicles. Since home ownership can be assumed as an indicator of family wealth, the

positive relationship between home ownership and vehicle ownership is not surprising and has been reported in several previous studies (e.g., [19,62]).

### 4.4. Impacts of Interactions on Vehicle Ownership

A strong positive relationship between the number of drivers in the household and vehicle ownership in all states was observed. This association implies that if the number of drivers in the household was lowered, vehicle ownership would decline substantially. This section looks at how household travel characteristics (HTCs) in each state moderate the effects of the most relevant BEA factors on vehicle ownership. BIKEINFRA was the most significant BEA variable in California, whereas HBPPOPDN and URBRUR were the most significant BEA variables in Missouri and Kansas, respectively. In all states, DRVRCNT was the most influential HTC variable. Figure 4 shows the change in predicted household vehicle counts when biking infrastructure conditions change from one to seven, a household living area changes from urban to rural, and population density increases from a category of 50 to a category of 30,000.
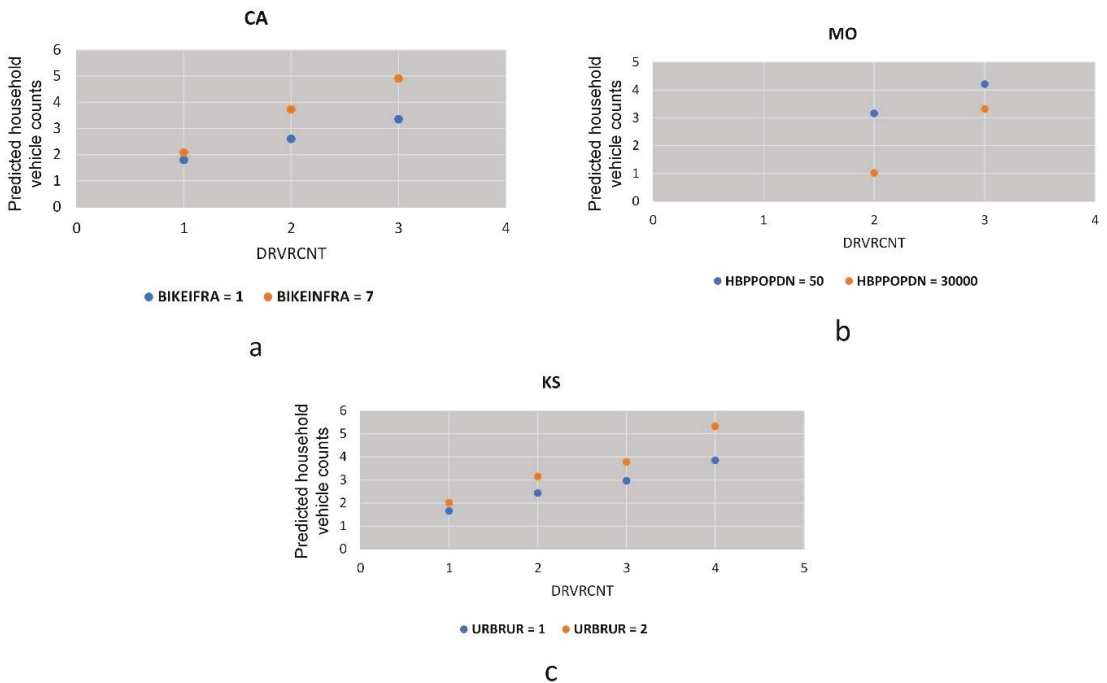


**Figure 4.** Relationships between essential BEA variables in each state and vehicle ownership variations moderated by key HTC variables. (**a**): Increase of vehicle ownership when condition of walking infrastructure changes from 1 to 7; (**b**): decrease in vehicle ownership as population density increases from 50 to 50,000 people; (**c**): increase in vehicle ownership as people's living environments shift from urban to rural.

DRVRCT has a complex moderating influence on the relationship between the built environment and household vehicle count. For example, when biking infrastructure conditions change from one to seven, predicted household vehicle counts for all the number of drivers in a household increase, but the predicted household vehicle count growth varies by the number of drivers in a household (Figure 4a). When the number of drivers in a household is one, the smallest increment (0.28) in the number of household vehicles occurs. A medium increase (1.13) in the number of household vehicles occurs when the number of household drivers is two. Finally, when there are three people who drive in a household,

the number of vehicles in the household increases the most (1.56). This suggests that the number of drivers in a household strengthens the positive influence of the deficiencies in biking infrastructure on vehicle ownership in California. The interaction effect of household living area (urban or rural) and DRVRCT on predicted household vehicle counts has a similar pattern in Kansas (Figure 4c). As living areas change from urban (1) to rural (2), vehicle ownership increases and the growth varies by the number of drivers in a household. When the household has four drivers, the largest increase in vehicle ownership occurs, suggesting that the number of drivers in a household amplifies the population density's positive effect on vehicle ownership.

As shown in Figure 4b, in Missouri, DRVRCT also moderates the impact of population density on predicted household vehicle counts. When the population density rises from 50 to 30,000, the predicted number of household vehicles decreases as well. When a household has two drivers, the number of household vehicles decreases the most ($-2.14$), whereas when a household has three drivers, the number of household vehicles decreases the least ($-0.89$). These findings show that having more drivers in a household can lessen the negative effects of high population density on vehicle ownership.

## 5. Discussions

It was expected that the number of drivers in the households plays a dominant role in predicting the count of the households' vehicles. However, very few studies have investigated the direct effects of the count of households' drivers on vehicle ownership. Some studies [61,63] found positive associations between the total number of household vehicles, vehicle usage, and energy consumption, which can be interpreted as indirect indicators of vehicle ownership trends. A possible reason that the number of drivers in the household became the most important household travel characteristics variable in predicting vehicle ownership in the three US states could be the direct and positive relationship between this variable and the number of adults in the households. Having more adults in a household means that people have different responsibilities and can travel independently. Thus, each adult household member may require their own vehicle, which cannot be shared with others due to time constraints. The importance of the number of drivers in the household in all three states shows that this variable is a determinant of households' vehicle ownership regardless of the state's population size.

Many previous studies have confirmed that providing adequate cycling and walking facilities encourages people to use these modes more frequently (e.g., [64–66]). At least for recreational or short trips, this may also encourage people to replace vehicles with walking and cycling [67,68]. These may be the causes of emerging deficiencies in cycling facilities as an important predictor of vehicle ownership in California. Having poor cycling facilities may increase the tendency of adult household members to buy more vehicles. According to The League of American Bicyclists [69], among all the US states, California, Missouri, and Kansas are ranked 4, 35, and 37, respectively, in terms of their suitability for cycling. Thus, the emergence of biking infrastructure conditions in California as an important factor is sensible. California has better conditions in terms of infrastructure and funding, education and encouragement, legislation and enforcement, policies and programs, and evaluation and planning than the other two states [69]. In addition, other factors such as biking culture, topography, and integration of walking and cycling facilities with public transport services can make a difference among the US states in terms of adoption of walking and biking instead of using private vehicles.

### 5.1. Findings' Implications

The practical examinations in the earlier sections accomplished the investigation objectives by revealing the characteristics of households that belonged to different US states and different populations. The results have significant implications for households' vehicle ownership. This paper's analysis clearly showed that the number of drivers in the household and deficiencies in cycling infrastructure heavily impacted the household

vehicle numbers. Moreover, the results revealed that these variables are determinants of household vehicle ownership regardless of the state's population. Thus, to discourage households from possessing multiple vehicles, any policy that reduces the impact of these variables is desired.

The members of the household have varying life commitments and travel requirements. As a result, considering all family members' needs and encouraging them to share their vehicles with other family members rather than purchasing more vehicles is a daunting task. However, some solutions, including using a minivan, flexible working time, using micro-mobility for first and last connections, and sending children to schools near the house, can be used to reduce the number of drivers in the household.

Improvements to the cycling infrastructure in all states (especially Kansas) should be at the center of attention. Some measures include the construction of paths, trails, or parks near housing units; the construction of sidewalks along all local and arterial streets; and the consistent assessment of sidewalks to ensure that they can serve all people, regardless of physical ability [64].

In most states, regardless of their population, the BEAs could not have the highest cumulative contribution to household vehicle ownership. Most BEAs had a minor impact on reducing vehicle ownership growth in the short term, but a BEA that made alternate modes of transportation competitive with a vehicle may have created a positive circle between the BEA and vehicle ownership in the long term. However, since transportation infrastructure and construction persist for years, a motorized-oriented urban layout is difficult to reverse once it has been established. Moreover, the motorized-oriented metropolitan structure will foster people's intention to purchase vehicles, which will be harmful to sustainable mobility.

### 5.2. Limitations

The study has significant limitations. First, the NHTS dataset is one of the largest household travel survey datasets in the world. However, its built environment indicators are limited. Some of these overlooked factors are location and transit accessibility (e.g., distance to the central business district and distance to the nearest metro/bus stop). Thus, future studies can employ other datasets containing more built-environment attributes and apply the XGBT method to perform their analysis. Additionally, it is suggested that the NHTS consider the factors mentioned above since these factors allow researchers to conduct a more comprehensive study regarding the issue of vehicle ownership in the US. Second, the NHTS includes items regarding reasons for not walking and biking. However, there are no items regarding the deficiencies in public transportation, particularly public buses. Future studies may complement the NHTS dataset with field observations on public transport infrastructure conditions. Finally, the authors believe that a sample of 5000 per state were enough to analyze the nonlinear relationship between vehicle ownership and other variables. However, future studies can use a larger sample to perform their analysis.

### 6. Conclusions

By means of data from the US National Household Travel Survey, this research utilized an extreme gradient boosting tree (XGBT) model to investigate the importance of sociodemographic factors, the HTCs, and the BEAs to vehicle ownership and their nonlinear associations with vehicle ownership. It is one of the few studies that look at how key HTCs moderate the effects of important BEAs on vehicle ownership in three different states in the United States with different populations. However, this study could not find a substantial difference in the results based on the states' populations. The main findings of this study for each state are as follows:

- In California, the predictability of vehicle ownership was driven by household travel characteristics (CI: 0.62). In this state, the number of drivers in a household and the deficiencies in cycling infrastructure were the two most important factors in predicting vehicle ownership.
- In Missouri, sociodemographic factors were dominant factors in predicting vehicle ownership (CI: 0.53). The number of drivers in a household and household income were the two most important predictors of vehicle ownership in Missouri.
- In Kansas, sociodemographic factors were the most influential factors in predicting vehicle ownership (CI: 0.55). Home ownership and the number of drivers in a household were the most influential factors in vehicle ownership in Kansas.

The outcomes demonstrate that the number of drivers in a household plays a dominant role in households' choice of vehicle ownership in the three US states. Crowded families with many drivers are more likely to possess more vehicles. In addition, deficiencies in the cycling infrastructure are another vital determinant of vehicle ownership in California. These two variables in California are the most significant predictors, accounting for 0.74 of the predictive capabilities. Identifying effective strategies to discourage households' drivers from buying new vehicles and improving the cycling infrastructure is key to sustainable transport in these states.

Policymakers could utilize land use and transport strategies to transform the built environment. The BEAs have a modest impact on vehicle ownership, and several BEAs may be used as proxies for the number of drivers in a household. Because practically all BEAs have a minor impact on their own, policymakers will need a combination of tactics if they intend to restrict vehicle ownership using land use and transport policy.

Some of the findings of this study are unique. For example, the nonlinear relationship between vehicle ownership and the number of drivers in a household has not been assessed by the previous studies. Thus, policymakers can use the findings of this study (thresholds, relationships, and interaction effects) to propose strategies to cope with the growth of vehicle ownership in the US.

Several factors are only connected with vehicle ownership when they fall within a specified range. It can result in a subjective interpretation of the associations between variables if the nonlinear associations are overlooked. This can lead planners and researchers to misjudge the significance of these variables and inaccurately signify their associations with vehicle ownership. More significantly, these ranges provide policymakers with recommendations about how to efficiently reduce the increase in vehicle ownership.

The findings of this research also showed that the XGBT can be successfully applied to reveal the complex relationships between the input variables and the target variables. Future studies can use this method to solve other issues in transportation science. To get more accurate results, they can combine the XGBT with other machine learning techniques, such as those that were proposed in Kumar et al. [70], Golilarz et al. [71], Golilarz et al. [72], Najafi Moghaddam Gilani et al. [73], Gilani et al. [74], and Tao et al. [75].

## References

1. Bureau of Transportation Statistics. National Household Travel Survey Daily Travel Quick Facts. Available online: https://www.bts.gov/statistical-products/surveys/national-household-travel-survey-daily-travel-quick-facts (accessed on 1 February 2022).
2. Handy, S.L.; Boarnet, M.G.; Ewing, R.; Killingsworth, R.E. How the built environment affects physical activity: Views from urban planning. *Am. J. Prev. Med.* **2002**, *23*, 64–73. [CrossRef]
3. Zhao, P.; Zhang, Y. Travel behaviour and life course: Examining changes in car use after residential relocation in Beijing. *J. Transp. Geogr.* **2018**, *73*, 41–53. [CrossRef]
4. Zegras, C. The built environment and motor vehicle ownership and use: Evidence from Santiago de Chile. *Urban Stud.* **2010**, *47*, 1793–1817. [CrossRef]
5. Sabouri, S.; Tian, G.; Ewing, R.; Park, K.; Greene, W. The built environment and vehicle ownership modeling: Evidence from 32 diverse regions in the US. *J. Transp. Geogr.* **2021**, *93*, 103073. [CrossRef]
6. Ao, Y.; Chen, C.; Yang, D.; Wang, Y. Relationship between rural built environment and household vehicle ownership: An empirical analysis in rural Sichuan, China. *Sustainability* **2018**, *10*, 1566. [CrossRef]
7. Kim, S.H.; Mokhtarian, P.L. Taste heterogeneity as an alternative form of endogeneity bias: Investigating the attitude-moderated effects of built environment and socio-demographics on vehicle ownership using latent class modeling. *Transp. Res. Part A Policy Pract.* **2018**, *116*, 130–150. [CrossRef]
8. Rahul, T.; Verma, A. The influence of stratification by motor-vehicle ownership on the impact of built environment factors in Indian cities. *J. Transp. Geogr.* **2017**, *58*, 40–51. [CrossRef]
9. Ruas, E.B. The Influence of Shared Mobility and Transportation Policies on Vehicle Ownership: Analysis of Multifamily Residents in Portland, Oregon. Ph.D. Thesis, Portland State University, Portland, OR, USA, 2019.
10. Li, J.; Walker, J.L.; Srinivasan, S.; Anderson, W.P. Modeling private car ownership in China: Investigation of urban form impact across megacities. *Transp. Res. Rec.* **2010**, *2193*, 76–84. [CrossRef]
11. Song, S.; Diao, M.; Feng, C.-C. Effects of pricing and infrastructure on car ownership: A pseudo-panel-based dynamic model. *Transp. Res. Part A Policy Pract.* **2021**, *152*, 115–126. [CrossRef]
12. Dargay, J.M.; Madre, J.-L.; Berri, A. Car ownership dynamics seen through the follow-up of cohorts: Comparison of France and the United Kingdom. *Transp. Res. Rec.* **2000**, *1733*, 31–38. [CrossRef]
13. Yang, Z.; Jia, P.; Liu, W.; Yin, H. Car ownership and urban development in Chinese cities: A panel data analysis. *J. Transp. Geogr.* **2017**, *58*, 127–134. [CrossRef]
14. Ma, Z. Multi-level Probabilistic Model for Population Synthesis and Vehicle Ownership Modeling Based on Samples with Missing Values. Master's Thesis, McGill University, Montreal, QC, Canada, 2021.
15. Cirillo, C.; Liu, Y. Vehicle ownership modeling framework for the state of Maryland: Analysis and trends from 2001 and 2009 NHTS data. *J. Urban Plan. Dev.* **2013**, *139*, 1–11. [CrossRef]
16. Chu, M.Y.; Law, T.H.; Hamid, H.; Law, S.H.; Lee, J.C. Examining the effects of urbanization and purchasing power on the relationship between motorcycle ownership and economic development: A panel data. *Int. J. Transp. Sci. Technol.* 2020; *in press*. [CrossRef]
17. Dargay, J.; Hanly, M. Volatility of car ownership, commuting mode and time in the UK. *Transp. Res. Part A: Policy Pract.* **2007**, *41*, 934–948. [CrossRef]
18. Bhat, C.R.; Paleti, R.; Pendyala, R.M.; Lorenzini, K.; Konduri, K.C. Accommodating Immigration Status and Self-Selection Effects in a Joint Model of Household Auto Ownership and Residential Location Choice. *Transp. Res. Rec.* **2013**, *2382*, 142–150. [CrossRef]
19. Li, S.; Zhao, P. Exploring car ownership and car use in neighborhoods near metro stations in Beijing: Does the neighborhood built environment matter? *Transp. Res. Part D Transp. Environ.* **2017**, *56*, 1–17. [CrossRef]
20. Huang, X.; Cao, X.J.; Yin, J.; Cao, X. Effects of metro transit on the ownership of mobility instruments in Xi'an, China. *Transp. Res. Part D: Transp. Environ.* **2017**, *52*, 495–505. [CrossRef]
21. Matas, A.; Raymond, J.-L.; Roig, J.-L. Car ownership and access to jobs in Spain. *Transp. Res. Part A Policy Pract.* **2009**, *43*, 607–617. [CrossRef]
22. Tyrinopoulos, Y.; Antoniou, C. Factors affecting modal choice in urban mobility. *Eur. Transp. Res. Rev.* **2013**, *5*, 27–39. [CrossRef]
23. Sabouri, S.; Brewer, S.; Ewing, R. Exploring the relationship between ride-sourcing services and vehicle ownership, using both inferential and machine learning approaches. *Landsc. Urban Plan.* **2020**, *198*, 103797. [CrossRef]
24. Holtzclaw, J.; Clear, R.; Dittmar, H.; Goldstein, D.; Haas, P. Location efficiency: Neighborhood and socio-economic characteristics determine auto ownership and use-studies in Chicago, Los Angeles and San Francisco. *Transp. Plan. Technol.* **2002**, *25*, 1–27. [CrossRef]
25. Shay, E.; Khattak, A.J. Household travel decision chains: Residential environment, automobile ownership, trips and mode choice. *Int. J. Sustain. Transp.* **2012**, *6*, 88–110. [CrossRef]
26. Ding, C.; Cao, X.; Dong, M.; Zhang, Y.; Yang, J. Non-linear relationships between built environment characteristics and electric-bike ownership in Zhongshan, China. *Transp. Res. Part D Transp. Environ.* **2019**, *75*, 286–296. [CrossRef]

27. Wang, X.; Yin, C.; Zhang, J.; Shao, C.; Wang, S. Nonlinear effects of residential and workplace built environment on car dependence. *J. Transp. Geogr.* **2021**, *96*, 103207. [CrossRef]
28. Zhang, W.; Zhao, Y.; Cao, X.J.; Lu, D.; Chai, Y. Nonlinear effect of accessibility on car ownership in Beijing: Pedestrian-scale neighborhood planning. *Transp. Res. Part D Transp. Environ.* **2020**, *86*, 102445. [CrossRef]
29. Bargegol, I.; Najafi Moghaddam Gilani, V.; Jamshidpour, F. Relationship between Pedestrians' Speed, Density and Flow Rate of Crossings through Urban Intersections (Case Study: Rasht Metropolis) (RESEARCH NOTE). *Int. J. Eng.* **2017**, *30*, 1814–1821.
30. Bargegol, I.; Amlashi, A.T.; Gilani, V.N.M. Estimation the Saturation Flow Rate at Far-side and Nearside Legs of Signalized Intersections—Case Study: Rasht City. *Procedia Eng.* **2016**, *161*, 226–234. [CrossRef]
31. Zhao, T.H.; Khan, M.I.; Chu, Y.M. Artificial neural networking (ANN) analysis for heat and entropy generation in flow of non-Newtonian fluid between two rotating disks. *Math. Methods Appl. Sci.* **2021**. Available online: https://onlinelibrary.wiley.com/doi/10.1002/mma.7310 (accessed on 1 January 2022). [CrossRef]
32. Zhao, T.-H.; Shi, L.; Chu, Y.-M. Convexity and concavity of the modified Bessel functions of the first kind with respect to Hölder means. *Rev. Real Acad. Cienc. Exactas Fís. Nat. Ser. A Matemáticas* **2020**, *114*, 96. [CrossRef]
33. Wang, M.-K.; Hong, M.-Y.; Xu, Y.-F.; Shen, Z.-H.; Chu, Y.-M. Inequalities for generalized trigonometric and hyperbolic functions with one parameter. *J. Math. Inequal* **2020**, *14*, 1–21. [CrossRef]
34. Zha, T.-H.; Castillo, O.; Jahanshahi, H.; Yusuf, A.; Alassafi, M.O.; Alsaadi, F.E.; Chu, Y.-M. A fuzzy-based strategy to suppress the novel coronavirus (2019-NCOV) massive outbreak. *Appl. Comput. Math.* **2021**, *20*, 160–176.
35. Zhao, T.-H.; Wang, M.-K.; Hai, G.-J.; Chu, Y.-M. Landen inequalities for Gaussian hypergeometric function. *Rev. Real Acad. Cienc. Exactas Físicas Nat. Ser. A Matemáticas* **2022**, *116*, 53. [CrossRef]
36. Iqbal, M.A.; Wang, Y.; Miah, M.M.; Osman, M.S. Study on Date–Jimbo–Kashiwara–Miwa Equation with Conformable Derivative Dependent on Time Parameter to Find the Exact Dynamic Wave Solutions. *Fractal Fract.* **2022**, *6*, 4. [CrossRef]
37. Zhao, T.-H.; Wang, M.-K.; Chu, Y.-M. Monotonicity and convexity involving generalized elliptic integral of the first kind. *Rev. Real Acad. Cienc. Exactas Fís. Nat. Ser. A Matemáticas* **2021**, *115*, 46. [CrossRef]
38. Zhao, T.-H.; He, Z.-Y.; Chu, Y.-M. On some refinements for inequalities involving zero-balanced hypergeometric function. *AIMS Math* **2020**, *5*, 6479–6495. [CrossRef]
39. Zhao, T.-H.; Wang, M.-K.; Chu, Y.-M. A sharp double inequality involving generalized complete elliptic integral of the first kind. *AIMS Math* **2020**, *5*, 4512–4528. [CrossRef]
40. Song, Y.-Q.; Zhao, T.-H.; Chu, Y.-M.; Zhang, X.-H. Optimal evaluation of a Toader-type mean by power mean. *J. Inequalities Appl.* **2015**, *2015*, 408. [CrossRef]
41. Conway, M.W.; Salon, D.; King, D.A. Trends in taxi use and the advent of ridehailing, 1995–2017: Evidence from the US National Household Travel Survey. *Urban Sci.* **2018**, *2*, 79. [CrossRef]
42. Jiao, J.; Bischak, C.; Hyden, S. The impact of shared mobility on trip generation behavior in the US: Findings from the 2017 National Household Travel Survey. *Travel Behav. Soc.* **2020**, *19*, 1–7. [CrossRef]
43. Das, V. Does Adoption of Ridehailing Result in More Frequent Sustainable Mobility Choices? An Investigation Based on the National Household Travel Survey (NHTS) 2017 Data. *Smart Cities* **2020**, *3*, 385–400. [CrossRef]
44. Tribby, C.P.; Tharp, D.S. Examining urban and rural bicycling in the United States: Early findings from the 2017 National Household Travel Survey. *J. Transp. Health* **2019**, *13*, 143–149. [CrossRef]
45. Porter, A.K.; Kontou, E.; McDonald, N.C.; Evenson, K.R. Perceived barriers to commuter and exercise bicycling in US adults: The 2017 National Household Travel Survey. *J. Transp. Health* **2020**, *16*, 100820. [CrossRef]
46. Li, X.; Liu, C.; Jia, J. Ownership and usage analysis of alternative fuel vehicles in the United States with the 2017 national household travel survey data. *Sustainability* **2019**, *11*, 2262. [CrossRef]
47. Jin, H.; Yu, J. Gender Responsiveness in Public Transit: Evidence from the 2017 US National Household Travel Survey. *J. Urban Plan. Dev.* **2021**, *147*, 04021021. [CrossRef]
48. Godfrey, J.; Polzin, S.E.; Roessler, T. *Public Transit in America: Observations from the 2017 National Household Travel Survey*; Center for Urban Transportation Research: Tampa, FL, USA, 2019.
49. Sadeghvaziri, E.; Tawfik, A. Using the 2017 National Household Travel Survey Data to Explore the Elderly's Travel Patterns. In Proceedings of the International Conference on Transportation and Development, Seattle, WA, USA, 26–29 May 2020; pp. 86–94.
50. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; Volume 10, pp. 785–794.
51. Hak Lee, E.; Kim, K.; Kho, S.-Y.; Kim, D.-K.; Cho, S.-H. Estimating Express Train Preference of Urban Railway Passengers Based on Extreme Gradient Boosting (XGBoost) using Smart Card Data. *Transp. Res. Rec.* **2021**, *2675*, 64–76. [CrossRef]
52. Mahmoud, N.; Abdel-Aty, M.; Cai, Q.; Yuan, J. Predicting cycle-level traffic movements at signalized intersections using machine learning models. *Transp. Res. Part C Emerg. Technol.* **2021**, *124*, 102930. [CrossRef]
53. Cai, Q.; Abdel-Aty, M.; Zheng, O.; Wu, Y. Applying machine learning and google street view to explore effects of drivers' visual environment on traffic safety. *Transp. Res. Part C Emerg. Technol.* **2022**, *135*, 103541. [CrossRef]
54. Kaur, H.; Kumar, M. Offline handwritten Gurumukhi word recognition using eXtreme Gradient Boosting methodology. *Soft Comput.* **2021**, *25*, 4451–4464. [CrossRef]

55. Ding, C.; Cao, X.J.; Næss, P. Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. *Transp. Res. Part A Policy Pract.* **2018**, *110*, 107–117. [CrossRef]
56. Transportation Secure Data Center. *2017 National Household Travel Survey—California*; National Household Travel Survey: Sacramento, CA, USA, 2019.
57. United States Census Bureau. Most Populous. 2017. Available online: https://www.census.gov/topics/population.html/ (accessed on 2 January 2022).
58. Potoglou, D.; Kanaroglou, P.S. Modelling car ownership in urban areas: A case study of Hamilton, Canada. *J. Transp. Geogr.* **2008**, *16*, 42–54. [CrossRef]
59. Kim, H.S.; Kim, E. Effects of Public Transit on Automobile Ownership and Use in Households of the USA. In *Review of Urban & Regional Development Studies: Journal of the Applied Regional Science Conference*; Blackwell Publishing: Oxford, UK; Boston, MA, USA, 2004.
60. Chu, Y.-L. Automobile Ownership Analysis Using Ordered Probit Models. *Transp. Res. Rec.* **2002**, *1805*, 60–67. [CrossRef]
61. Brownstone, D.; Golob, T.F. The impact of residential density on vehicle usage and energy consumption. *J. Urban Econ.* **2009**, *65*, 91–98. [CrossRef]
62. Shao, Q.; Zhang, W.; Cao, X.J.; Yang, J. Nonlinear and interaction effects of land use and motorcycles/E-bikes on car ownership. *Transp. Res. Part D Transp. Environ.* **2022**, *102*, 103115. [CrossRef]
63. Liu, F.; Zhao, F.; Liu, Z.; Hao, H. The Impact of Purchase Restriction Policy on Car Ownership in China's Four Major Cities. *J. Adv. Transp.* **2020**, *2020*, 7454307. [CrossRef]
64. Aghaabbasi, M.; Moeinaddini, M.; Shah, M.Z.; Asadi-Shekari, Z. A new assessment model to evaluate the microscale sidewalk design factors at the neighbourhood level. *J. Transp. Health* **2017**, *5*, 97–112. [CrossRef]
65. Buehler, R. Determinants of bicycle commuting in the Washington, DC region: The role of bicycle parking, cyclist showers, and free car parking at work. *Transp. Res. Part D Transp. Environ.* **2012**, *17*, 525–531. [CrossRef]
66. Piatkowski, D.P.; Marshall, W.E. Not all prospective bicyclists are created equal: The role of attitudes, socio-demographics, and the built environment in bicycle commuting. *Travel Behav. Soc.* **2015**, *2*, 166–173. [CrossRef]
67. Verma, A.; Vajjarapu, H.; Thuluthiyil Manoj, M. Planning and Usage Analysis of Bike Sharing System in a University Campus. In Proceedings of the Recent Advances in Traffic Engineering, Singapore, 29 August 2020; pp. 339–349.
68. Torcat, A.; McCray, T.; Durden, T. *Changing Perceptions of Cycling in the African American Community to Encourage Participation in a Sport that Promotes Health in Adults*; Southwest Region University Transportation Center (US): College Station, TX, USA, 2015.
69. The League of American Bicyclists. Bicycle-Friendly. Available online: https://rosap.ntl.bts.gov/view/dot/29321 (accessed on 3 January 2022).
70. Kumar, R.; Khan, A.A.; Kumar, J.; Zakria; Golilarz, N.A.; Zhang, S.; Ting, Y.; Zheng, C.; Wang, W. Blockchain-Federated-Learning and Deep Learning Models for COVID-19 Detection Using CT Imaging. *IEEE Sens. J.* **2021**, *21*, 16301–16314. [CrossRef]
71. Golilarz, N.A.; Mirmozaffari, M.; Gashteroodkhani, T.A.; Ali, L.; Dolatsara, H.A.; Boskabadi, A.; Yazdi, M. Optimized Wavelet-Based Satellite Image De-Noising With Multi-Population Differential Evolution-Assisted Harris Hawks Optimization Algorithm. *IEEE Access* **2020**, *8*, 133076–133085. [CrossRef]
72. Golilarz, N.A.; Addeh, A.; Gao, H.; Ali, L.; Roshandeh, A.M.; Munir, H.M.; Khan, R.U. A New Automatic Method for Control Chart Patterns Recognition Based on ConvNet and Harris Hawks Meta Heuristic Optimization Algorithm. *IEEE Access* **2019**, *7*, 149398–149405. [CrossRef]
73. Najafi Moghaddam Gilani, V.; Hosseinian, S.M.; Ghasedi, M.; Nikookar, M. Data-Driven Urban Traffic Accident Analysis and Prediction Using Logit and Machine Learning-Based Pattern Recognition Models. *Math. Probl. Eng.* **2021**, *2021*, 9974219. [CrossRef]
74. Gilani, V.N.M.; Hosseinian, S.M.; Hamedi, G.H.; Safari, D. Presentation of predictive models for two-objective optimization of moisture and fatigue damages caused by deicers in asphalt mixtures. *J. Test. Eval.* **2021**, *49*, 1–22.
75. Tao, W.; Aghaabbasi, M.; Ali, M.; Almaliki, A.H.; Zainol, R.; Almaliki, A.A.; Hussein, E.E. An Advanced Machine Learning Approach to Predicting Pedestrian Fatality Caused by Road Crashes: A Step toward Sustainable Pedestrian Safety. *Sustainability* **2022**, *14*, 2436. [CrossRef]

*Article*

# How Sustainable Is People's Travel to Reach Public Transit Stations to Go to Work? A Machine Learning Approach to Reveal Complex Relationships

**Panyu Tang [1,\*], Mahdi Aghaabbasi [2,\*], Mujahid Ali [3,\*], Amin Jan [4,\*], Abdeliazim Mustafa Mohamed [5,6] and Abdullah Mohamed [7]**

[1]  ChengDu Jincheng College, Chengdu 611731, China
[2]  Centre for Sustainable Urban Planning and Real Estate (SUPRE), Department of Urban and Regional Planning, Faculty of Built Environment, University of Malaya, Kuala Lumpur 50603, Wilayah Persekutuan Kuala Lumpur, Malaysia
[3]  Department of Civil and Environmental Engineering, Universiti Teknologi Petronas, Seri Iskandar 32610, Perak, Malaysia
[4]  Faculty of Hospitality, Tourism and Wellness, Universiti Malaysia Kelantan, City Campus, Kota Bharu 16100, Kelantan, Malaysia
[5]  Department of Civil Engineering, College of Engineering, Prince Sattam Bin Abdulaziz University, Alkharj 16273, Saudi Arabia; a.bilal@psau.edu.sa
[6]  Building & Construction Technology Department, Bayan College of Science and Technology, Khartoum 210, Sudan
[7]  Research Centre, Future University in Egypt, New Cairo 11745, Egypt; mohamed.a@fue.edu.eg
\*  Correspondence: tangmi127@126.com (P.T.); mahdi@um.edu.my (M.A.); mujahid_19001704@utp.edu.my (M.A.); aminjan@umk.edu.my (A.J.)

**Abstract:** Several previous studies examined the variables of public-transit-related walking and privately owned vehicles (POVs) to go to work. However, most studies neglect the possible non-linear relationships between these variables and other potential variables. Using the 2017 U.S. National Household Travel Survey, we employ the Bayesian Network algorithm to evaluate the non-linear and interaction impacts of health condition attributes, work trip attributes, work attributes, and individual and household attributes on walking and privately owned vehicles to reach public transit stations to go to work in California. The authors found that the trip time to public transit stations is the most important factor in individuals' walking decision to reach public transit stations. Additionally, it was found that this factor was mediated by population density. For the POV model, the population density was identified as the most important factor and was mediated by travel time to work. These findings suggest that encouraging individuals to walk to public transit stations to go to work in California may be accomplished by adopting planning practices that support dense urban growth and, as a result, reduce trip times to transit stations.

**Keywords:** sustainable travel to public transit stations; complex relationship; Bayesian network algorithm; work trip

## 1. Introduction

A transition away from privately owned vehicles (POVs) toward active transport can have major health advantages [1]. Despite the vast benefits of active transport modes, particularly walking, many individuals still prefer POVs. For example, just 36% of all journeys in the United States were below one mile, and only 27% of such journeys were conducted by walking or biking [2]. According to statistics from the American Community Survey, the percentage of people who walk to work in the United States has declined from 5.6 per cent in 1980 to 2.8 per cent in 2012 [3].

The "park and ride" concept, which promotes the use of POVs to reach public transit (PT) stations and combines the use of private cars and PT stations to reduce the negative

consequences of private vehicle use, has been the subject of several studies in the past [4–7]. Typically, this system is found at rail transportation terminals and transportation hubs, which allow for both rail and public bus access. However, this system may not be available at local bus stops. POVs are feasible options to reach PT stations where the stations are not within walking distance or in low-density areas [8,9]. Although there is hope that this approach will reduce the negative consequences of private vehicle use (e.g., traffic congestion, pollution, and physical inactivity), it is more desirable for planners to minimize the role of POVs in people's daily travels, particularly those related to work.

PT stations may supplement and extend the variety of active modes significantly [10,11]. Because of this, as well as the fact that POVs are associated with a slew of other well-known issues, there may be room for a modal shift away from POVs toward walking and PT that might reduce the usage of POVs, while also contributing to increased physical activity [12–15]. Because health conditions, work trip qualities, work attributes, and sociodemographic factors may all impact travel patterns [16–21], it is important to know how walking connects to PT travel to work to reap the most advantages.

California has always had a problem with traffic congestion. The cause for this ongoing issue is that the region's population and POV usage have exceeded the transportation facilities. If California's transportation system cannot keep pace with the state's fast urban growth, and if Californians' priority for POVs continues, the traffic problem will undoubtedly worsen soon. To deal with long-term urban traffic issues, dwellers in crowded regions are encouraged to replace POVs with active transportation options and PT, especially for work-related journeys. When people combine walking with PT, which is a hot topic among planners, the advantages of this replacement may be maximized. Walking is the most cost-effective mode of transportation and the most basic form of physical activity [22–25]. Walking also needs a low-cost infrastructure. As a result, it makes sense if planners encourage individuals to walk to PT stations over other active forms of transportation.

Many studies have been conducted on the topic of first-mile connection, which addresses how people reach PT stations. Several studies assessed the impact of sociodemographic characteristics on walking to reach PT stations. Factors, such as age [17,18,26–28], gender [29,30], vehicle ownership [31,32], income [33–35], and education [32,36,37], were significantly correlated with the walking to reach PT stations. Although there are a lot of built environment (BE) factors that impact travel behavior, only a very small number of these factors were included in the first-mile connection studies. These factors included density and distance to PT stations [38–42]. Earlier studies have shown that population density is one of the most significant BE variables, and its impact on travel behavior is stronger than other BE attributes [32,36,43,44]. People in low-density areas are more likely to use POVs than those in medium- and high-density areas [45,46]. Similarly, for individuals who live in low-density regions where the station is too far away to walk to and bus service is not accessible, driving to PT stations may be the sole choice for reaching PT stations [8].

While sociodemographic variables have been well covered in earlier studies, health-related factors and their impact on mode choice have rarely been considered in most PT-related walking investigations [1,47,48]. BMI, self-assessed health, self-reported smoker, and yearly frequency of hospital and primary care visits are characteristics addressed in these studies. To the best of the authors' knowledge, no study considered the impact of medical conditions in PT-related walking research. Furthermore, most of these studies neglected job-related issues, as well as those associated with work trips. Flexibility in work arrival time, full-time/part-time worker, the possibility of working from home, the distance between home and work location, trip time to work, time spent transferring on the commute to work if PT is taken, and travel time to PT station are some of the aspects that are overlooked [49–51]. The existence of such data in the U.S. National Household Travel Survey can give an excellent chance to look at the influence of a medical condition and work-related trips on travel mode selection.

There are non-linear and complex interactions between variables in transportation systems (e.g., the relationship between the built environment and car ownership) that

are difficult to study using typical statistical approaches and linear programing methods [52–58]. Non-linear relationships may be inconsistent, and factors may have threshold correlations with a variable of interest. Because non-linear relationships can help planners to understand the effective influence range of important factors on the target variable, it is interesting to see if this result can be applied to other fields [59]. This supports planners in fine-tuning their strategies [60]. Most PT-related walking studies employed traditional statistical methods (Table 1). However, these methods are unable to reveal complex relationships. In addition, these models have strict linearity assumptions, which limits the ability of these models to be effectively generalized [61–66]. Finally, these models are vulnerable to missing and incomplete data. Machine learning (ML) approaches can be used to solve the problems outlined before [67–70]. The Bayesian Network (BN) model is one of these powerful tools, and it has lately been used successfully in various transport-related research [67,71–75]. A BN model can effectively deal with heterogeneous and under-sampling data, as well as missing, erroneous, or ambiguous data. Because it can effectively alter its network depending on the data provided or entered into it, BN is indeed thought to be excellent for learning changeable behaviors (e.g., mode choice) [76–81].

**Table 1.** Some recent studies on PT-related walking.

| Study | Location | Sample Size (Unit) | Analysis Method |
|---|---|---|---|
| Patterson, Webb, Millett and Laverty [19] | USA | 2448 (passengers) | LRM |
| Ratanawaraha et al. [82] | Thailand | 1020 (travelers) | LRM |
| Rodriguez-Gonzalez and Aguero-Valverde [83] | Costa Rica | N/A | PDF and CDF |
| Sun, Zacharias, Ma and Oreskovic [38] | China | 495 (adults) | LRM |
| Townsend and Zacharias [84] | Thailand | 1489 (travelers) | LRM |
| Vandebona and Tsukaguchi [85] | Japan | 3560 (persons) | DE |
| Voss et al. [86] | Canada | 42 (students) | LRM |
| Wang and Cao [87] | USA | 7077 (travelers) | LRM |
| Wasfi, Ross and El-Geneidy [27] | Canada | 6913 (trips) | LRM |
| Xi et al. [88] | Canada | 21,470 (trips) | DE |
| Yu and Lin [39] | USA | 18,180 (trips) | LRM |
| Zacharias and Zhao [89] | China | 2409 (passengers) | LRM |
| Zhao et al. [90] | China | 1544 (travelers) | LRM |
| Zuo et al. [91] | USA | 1330 (trips) | DE |

LRM = logistic regression model; PDF = probability density function; CDF = cumulative distribution function; DE = Descriptive analysis.

The authors of this research utilize the BN model to explore the major indicators of travel mode to reach PT stations and highlight their non-linear interactions using the 2017 U.S. National Household Travel Survey (2017 NHTS). The following are the questions that this research aims to answer: (1) How important are the health condition, work trip, work, and individual and household attributes to individuals who use walking or POV to reach PT stations to go to work in California? (2) Do the most important variables have associations with walking or POV to reach PT stations to go to work?

This paper contributes to the literature in three major ways. To begin with, it adds to the research of mode choice for reaching PT stations to go to work in California and other regions where traffic congestion is a problem. Furthermore, this study evaluates the relative relevance of several elements in walking to work and gives insight into the policy implementation priorities in California and other places with similar conditions. It also demonstrates that important factors have irregularly complex relationships, corroborating the scant data in the literature and providing recommendations for California planning approaches. Finally, this study demonstrates the significant role of trip time to work and its combined effect with population density in POV usage to reach PT stations to go to work, as well as the significant role of population density and its interaction impacts with trip time to PT stations in walking reach PT stations to go to work, thereby bolstering the case for dense urban development.

The following is a breakdown of how the research is structured. The data, variables, and modeling technique are introduced in Section 2. Section 3 discusses the models' results and performance, variable importance, relationships with travel mode to reach PT stations in California, and interaction impacts on mode choice to reach PT stations. The final section highlights the most important findings and explains policy implications.

## 2. Materials and Methods

In this study, two associative and predictive BN models were developed to reveal the complex relationships between various variables and PT-related walking and PT-related POVs in California. As previously mentioned, the 2017 U.S. National Household Travel Survey (2017 NHTS) was employed to conduct this study. These models discover interaction effects of independent factors on the usage of walking and POVs to reach PT stations to go to work and assess the relevance of variables in predicting the choice of walking and POV to reach PT stations to go to work. Figure 1 shows the flowchart of this study.
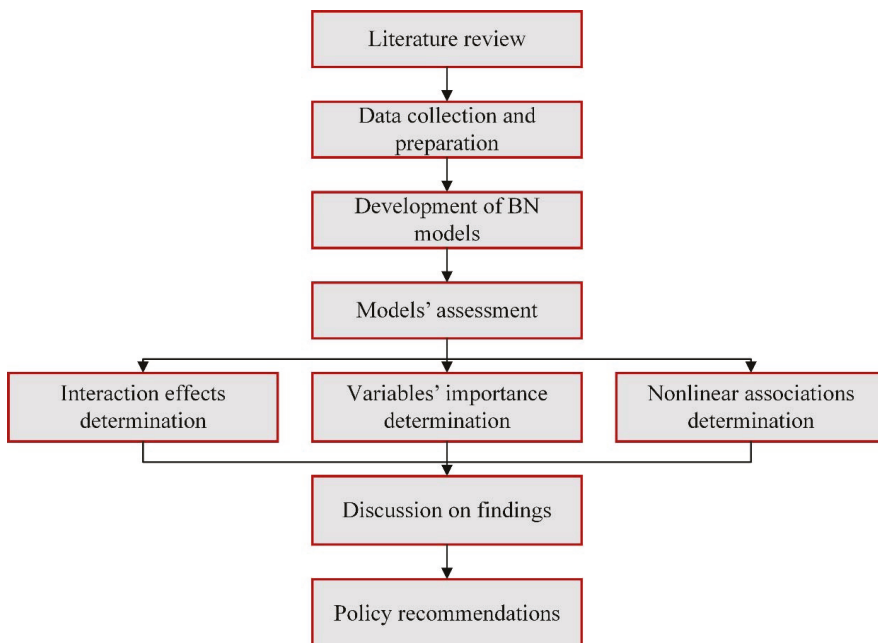


**Figure 1.** This study's flowchart.

*2.1. Data*

This study used data from the 2017 U.S. National Household Travel Survey (2017 NHTS). The NHTS has now become the country's rich source of information on commuting by U.S. citizens throughout all fifty states. This commute behavior database contains journeys taken in a variety of ways and for a variety of reasons. The data for the NHTS are gathered from a randomly selected sample of U.S. households. The NHTS supplies data about individual and household travel behavior patterns. These patterns are related to sociodemographic and geographic factors that impact travel choices and are used to estimate demand. More information can be found at https://nhts.ornl.gov (accessed on 1 January 2020).

This study looked into how Californians utilize walking and privately owned vehicles (POVs) to reach public transportation to go to work. Thus, the study team mined the whole

dataset for relevant data. Public transport in this study refers to public or commuter buses, subways, elevated and light rail, and Amtrak. The following criteria were used to choose the samples: (1) residence in California, (2) the use of public transportation to commute to work, and (3) the use of walking and POVs to reach public transportation to go to work. A total of 796 samples were used to create the final dataset. A total of 19 input variables and 2 target variables were included in the dataset (walk to reach public transit stations and POVs to reach public transit stations). Table 2 lists all of the variables utilized in this investigation.

**Table 2.** Variables employed in this study.

| Variable | Acronym | Category/Data Range |
|---|---|---|
| Health condition attributes (HCA) | | |
| Medical condition results in using bus or subway less frequently | CONDPUB | yes (1) and no (2) |
| Medical condition results in giving up driving | CONDRIVE | yes (1) and no (2) |
| Opinion of health | HEALTH | excellent (1), very good (2), good (3), fair (4), poor (5) |
| Medical condition, length of time | MEDCOND6 | 6 months or less (1), more than 6 months (2), and all life (3) |
| Built environment characteristics (BEC) | | |
| Category of population density (persons per square mile) in the census block group of the household's home location | HBPPOPDN | 0–99 (50), 100–499 (300), 500–999 (750), 1000–1999 (1500), 2000–3999 (3000), 4000–9999 (7000), 10,000–24,999 (17,000), and 25,000–999,999 (30,000) |
| Individual and household attributes (IHA) | | |
| Educational attainment | EDUC | Less than a high school graduate (1), high school graduate or GED (2), some college or associated degree (3), bachelor's degree (4), graduate degree or professional degree (5) |
| Household income | HHFAMINC | less than USD 10,000 (1), 10,000–14,999 (2), 15,000–24,999 (3), 25,000–34,999 (4), 35,000–49,999 (5), 50,000–74,999 (6), 75,000–99,999 (7), 100,000–124,999 (8), 125,000–149,999 (9), 150,000–199,999 (10) and 200,000 or more (11) |
| Age | AGE | 16–75 |
| Gender | SEX | male (1) and female (2) |
| Owned vehicle longer than a year | VEHOWNED | yes (1) and no (2) |
| Number of workers in household | WRKCOUNT | 1–4 |
| Work attributes (WA) | | |
| Flexibility in work arrival time | FLEXTIME | yes (1) and no (2) |
| Count of days working from home in the last month | WKFMHMXX | 0–16 |
| Full-time or part-time worker | WKFTPT | full time (1) and part time (2) |
| Option of working from home | WKRMHM | yes (1) and no (2) |

**Table 2.** *Cont.*

| Variable | Acronym | Category/Data Range |
|---|---|---|
| Work trip attributes (WTA) | | |
| Road network distance, in miles, between respondent's home location and work location | DISTTOWK17 | 0–399.07 |
| Minutes spent commuting to work | PUBTIME | 0–50 |
| Trip time to work in minutes | TIMETOWK | 0–170 |
| Trip time to transit station in minutes | TRACCTM | 0–70 |
| Target variable | | |
| Walk as mode used to reach public transit station | TRACC_WLK | yes (1) and no (2) |
| POV as mode used to reach public transit station | TRACC_POV | yes (1) and no (2) |

In the NHTS dataset, several active modes, including bikes and e-scooters, and passive modes, such as ride-sourcing, are not considered to reach PT stations. This can be regarded as a drawback of the NHTS dataset and a limitation of this study. Furthermore, although the literature suggests that the most critical BE factors for examining the first mile connection to PT stations are population density and distance to the PT station, the NHTS only considers population density. As a result, the only BE input in this investigation was population density, whose direct and significant impacts on travel behavior and mode choice for PT have been widely proven.

### 2.2. Bayesian Network (BN) Model

Bayesian Networks (BNs), commonly referred to as belief networks, are probabilistic network models that combine probability and graph theory. The following are the main two methods for acquiring BN structures. The first method is based on expert judgment and uses subjective causal links to construct a BN structure. The second method, known as structural learning, uses certain learning models to detect and guide the edges on a given dataset. By using the latter method, this investigation creates the BN architecture. There are numerous data-driven techniques, including Nave Bayesian Networks (NBN), Augmented Naive Bayesian Networks (ABN), and Tree Augmented Networks (TAN). TAN learning generates qualitative BN-depicting variables' interacting dependencies, which aids in generating insights into the crucial elements that influence travel mode choice. Friedman et al. [92] have noted that TAN beats naive Bayes, while retaining the calculation efficiency and stability that naive Bayes is known for. Other data-driven configuration algorithms are less effective and reliable than TAN [93]. In this research, the analysis was performed using SPSS Modeller, which is worth noting.

A BN that is a labelled directed acyclic graph (DAG) represents a joint probability distribution over a collection of random inputs $Q$. Let $Q = \{B_1, \ldots B_i, D\}$, where $i$ refers to the number of inputs, the inputs $A_1, \ldots A_i$ are the variables, and $D$ signifies the class variable (mode to public transit station).

Assume a network structure in which the target variable serves as the root, namely $\prod D = \varnothing$, and every variable possesses the target variable as its sole parent, namely $\prod B_j = \{D\}$ for $1 \le j \le i$. Equation (1) characterizes a BN as a single joint probability distribution across $Q$.

$$P(B_1, \ldots B_i, D). \prod_{j=1}^{i} P(B_j > \mid D) \tag{1}$$

When $\prod B_j$ has just one parent for any and all $B_j$ apart from one variable-lacking parent, the DAG over $\{B_1, \ldots A_i\}$ is a tree. When there is only one $j$ so that $\pi(j) = 0$, and therefore there is no series $j_1, \ldots j_s$ so that $\pi(jh) = j_{h+1}$ given $j \le h \le s$ and $\pi(j_s) = i_1$, there is indeed a function $\pi$ that can describe a tree across $B_1, \ldots B_i$. Such a function

describes a tree network where $\prod B_j = \{D, \ldots B_{\pi(j)}\}$ if $\pi(i) > 0$, and $\prod B_j = \{D\}$ if $\pi(j) = 0$.

It is an optimization challenge to learn a TAN structure. Chow and Liu [94], who employed conditional mutual information between characteristics, offered a broad technique for solving this problem. The following is a definition of the function:

$$IM(B_j, \, B_h|D) = \sum_{b_{jj}, b_{hj}, d_j} P\Big(b_{jj}, b_{hj}, d_j\Big) log \frac{P(b_{jj}, \, b_{hj}|d_j)}{P(b_{jj}|d_j)P(b_{hj}|d_j)} \qquad (2)$$

where *IM* denotes the conditional mutual information, $b_{jj}$ is the *j*th state of variable $B_j$, $b_{hj}$ is the *j*th state of variable $B_h$, $d_j$ is the *j*th state of "mode choice to transit station". The optimization challenge of learning a TAN structure is to develop a tree characterizing function across $B_1, \ldots B_i$ that maximises the log-likelihood.

## 3. Results

### 3.1. Models' Results and Performance

Two Bayesian Network (BN) models were developed in this study to predict the choice of walking and POVs to reach PT stations among Californians. To develop these models, the structure type of the BN models was the TAN algorithm and the parameter learning method was Bayes adjustment. It is worth mentioning that the data were split into train and test partitions with a ratio of 80:20 before the models' development. The training partition was used to build the models, whereas the test partition was utilized to evaluate the created model using unseen data. The BN models were used to (1) determine the importance of variables in predicting the choice of walking and POVs to reach PT stations, (2) determine relationships with travel mode to reach PT stations in California, and (3) identify the interaction effects of independent variables on the use of walking and POVs to reach PT stations. The structures of the BN models developed in this study are shown in Figure 2.
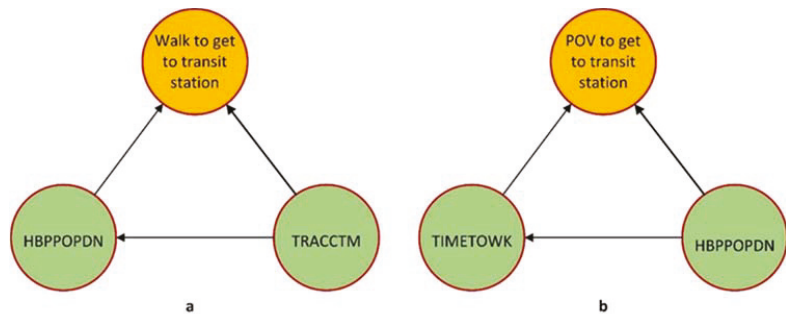


**Figure 2.** BNs' structure: associations between the travel modes to reach public transport and their most important variables and mediators as identified by the BN model. (**a**) BN model for walking to reach public transit station; (**b**) BN model for POV to reach public transit station.

The performance of the two BN models is shown in Table 3. Both models achieved a high accuracy in both the training and testing phases. In addition, the accuracies of the training and testing phases are almost similar, which implies the stability of both models. The models' performance also was assessed using receiver operating characteristic (ROC) diagrams (Figure 3). The ROC curve depicts the sensitivity–specificity trade-off. Models with curves nearer to the top-left corner perform much better. A random model is expected to yield diagonal points (sensitivity = specificity) as a reference point. The nearer the curve is to the ROC space's 45 degree diagonal, the less accurate the test becomes. As can be seen, both models indicated a great performance for both classes (yes and no).

**Table 3.** Models' performance.

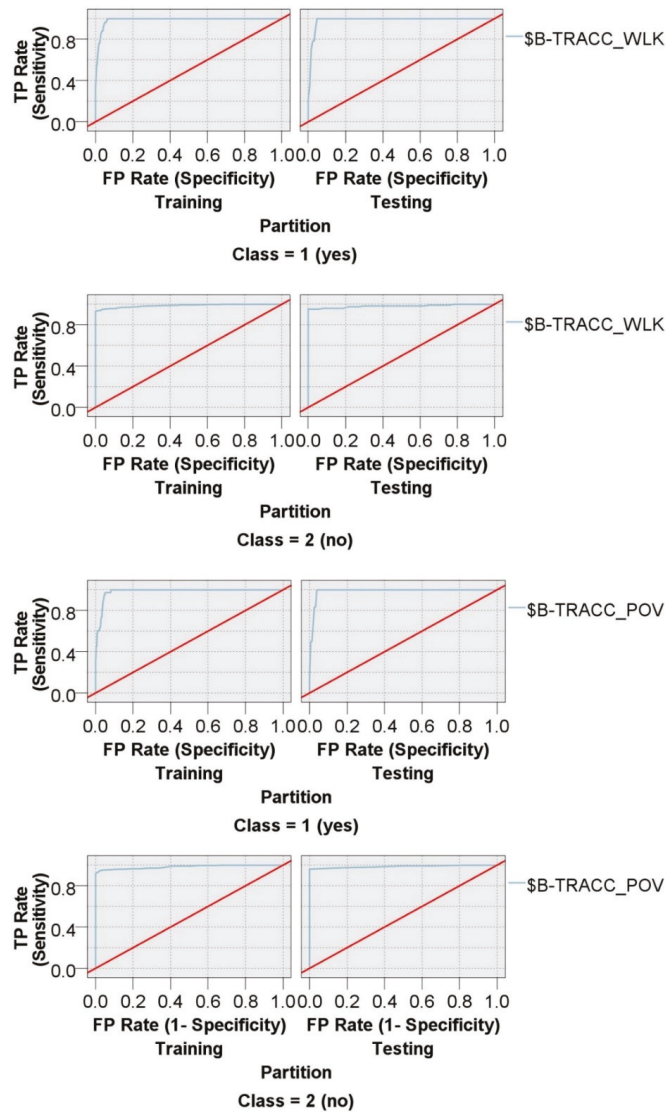| Model | Train | | Test | |
|---|---|---|---|---|
| | **Correct (%)** | **Wrong (%)** | **Correct (%)** | **Wrong (%)** |
| TRACC_WALK | 94.94 | 5.06 | 94.48 | 5.52 |
| TRACC_POV | 96.52 | 3.48 | 96.93 | 3.07 |



**Figure 3.** Receiver operating characteristic graphs for the BN models developed in this study.

*3.2. Variable Importance*

Table 4 shows the significance of all independent variables in forecasting travel mode to reach public transit stations. In addition, the cumulative impact of four types of factors is shown in Table 5. For walking, the results showed that work trip attributes (WTA)

dominated the prediction of mode choice to reach PT stations in California. For POV, individual and household attributes (IHA) largely influenced the forecast of mode choice to use PT. Especially, the predictive power of all the WTAs was 0.58. The combined contribution of IHA variables was 0.33 for POVs.

**Table 4.** Importance of the various types of variables.

| Factor Type | Variable | Walk | POV |
|---|---|---|---|
| Health condition attributes (HCA) | | | |
| HCA | CONDPUB | 0.03 | 0.04 |
| HCA | CONDRIVE | 0.02 | 0.04 |
| HCA | HEALTH | 0.02 | 0.05 |
| HCA | MEDCOND6 | 0.02 | 0.04 |
| Built environment characteristics (BEC) | | | |
| BEC | HBPPOPDN | 0.03 | 0.12 |
| Individual and household attributes (IHA) | | | |
| IHA | EDUC | 0.05 | 0.05 |
| IHA | HHFAMINC | 0.07 | 0.05 |
| IHA | AGE | 0.05 | 0.08 |
| IHA | SEX | 0.02 | 0.05 |
| IHA | VEHOWNED | 0.02 | 0.03 |
| IHA | WRKCOUNT | 0.06 | 0.07 |
| Work attributes (WA) | | | |
| WA | FLEXTIME | 0.02 | 0.07 |
| WA | WKFMHMXX | 0.00 | 0.00 |
| WA | WKFTPT | 0.00 | 0.06 |
| Work trip attributes (WTA) | | | |
| WTA | DISTTOWK17 | 0.03 | 0.03 |
| WTA | PUBTIME | 0.04 | 0.03 |
| WTA | TIMETOWK | 0.05 | 0.09 |
| WTA | TRACCTM | 0.42 | 0.07 |
| WTA | WKRMHM | 0.04 | 0.03 |

**Table 5.** Cumulative importance of factors.

| Factor Type | Cumulative Importance | |
|---|---|---|
| | Walk | POV |
| Health condition attributes (HCA) | 0.10 | 0.17 |
| Built environment characteristics (BEC) | 0.03 | 0.12 |
| Work trip attributes (WTA) | 0.58 | 0.25 |
| Work attributes (WA) | 0.02 | 0.13 |
| Individual and household attributes (IHA) | 0.27 | 0.33 |
| Sum | 1.00 | 1.00 |

In terms of the WTAs' impact on choosing the walking mode to reach PT stations, the trip time to the transit station (TRACCTM) was the most important variable in predicting walking choice to the PT station. Previous research has found a negative association between distance to transit stations and nonmotorized travel behavior [95–97]. As a result, it was expected that the time spent traveling to the transit station would emerge as the most relevant factor in predicting the likelihood of walking to the transit station. Individuals who walk to reach PT stations to go to work may place a different value on their time.

People's gender, income, family responsibilities, and other factors can all contribute to this difference [98]. As a result of these distinctions, different levels of sensitivity to walking time to PT stations may emerge.

In California, the population density (HBPPOPDN) has a 0.12 predictive power for POV usage to reach PT stations. The transport mode to the transit station is heavily influenced by population density [99]. In high-density areas, active transportation modalities are commonly used to reach transit stops. On the other hand, cars are the most prevalent form of transportation to transit stations in low-density areas.

### 3.3. Relationships with Travel Mode to Reach Public Transit Stations in California

In this section, the non-linear associations of the most important variable of walking and POVs to reach PT stations and the prediction of occurrence of these travel modes are discussed. It is vital to determine these complex relationships since it helps to identify the relevant impact ranges of these factors. According to the results of the BN models, the most important factor for predicting walking adoption to reach PT stations was the trip time to the transit station (TRACCTM), while the population density of participants' house location (HBPPOPDN) was chosen as the most important predictor of POV adoption to reach PT stations.

Figure 4 displays the relationships mentioned above. When the average time to reach PT stations is around 10 min, Californians are more inclined to walk to the transit stations. If the typical commute duration to PT stations is around 40 min, Californians are less likely to walk to PT stations. This study's results are consistent with Sun and Yin [100] findings, which revealed that shorter travel times and shorter distances to PT stations might increase the likelihood of walking to them.
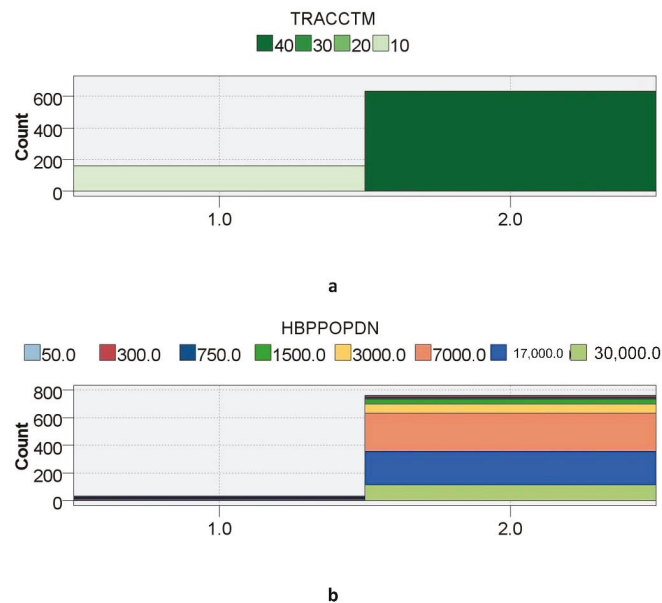


a



b

**Figure 4.** Non-linear relationship between the most important variable in each model and prediction of the travel mode choice to reach public transit stations. (**a**) Prediction of the walking choice to get to public transit; (**b**) prediction of the POV choice to get to public transit.

If the participants' dwelling is in a densely populated area (e.g., 7000–30,000 persons per square mile), it is unlikely that they will utilize a POV. In contrast, if the dwelling units are in a low-density region (e.g., 750 persons per square mile), POV is more likely

to be used. These findings corroborate those of Nigro, Bertolini and Moccia [99], which found that population density influences the mode of transportation used to access PT stations. Combining the results of the time to the transit station (for the walking model) and population density (for the POV model), when the time to the transit station is less than 10 min or the population density in the household's home location is 7000–30,000 persons per square mile, walking to reach PT stations could be increased by densifying land use around PT stations.

### 3.4. Interaction Impacts on Mode Choice to Reach Public Transit Stationsr

The strong negative connections between travel time to transit stations and walking to each PT stations suggest that, if the trip duration to transit stations can be reduced, walking will become more popular. The BN model revealed that another variable, population density (HBPPOPDN), mediates the effect of trip time to reach public transit stations (TRACCTM) on walking to PT stations (Figure 5a). Figure 5a illustrates the combined influence of these two variables on forecasting walking to reach PT stations in California. Walking is more probable when the trip time to the PT stations is less than 10 min. These lower trip times to transit stations occur in high density areas (e.g., 7000–30,000 persons per square mile). This means that the TRACCTM's negative relationship with the walking level to the transit station is amplified by HBPPOPDN. The influence of trip time to transit stations on walking to PT stations is mediated more by a population density of 17,000 persons per square mile than by other HBPPOPDN values.
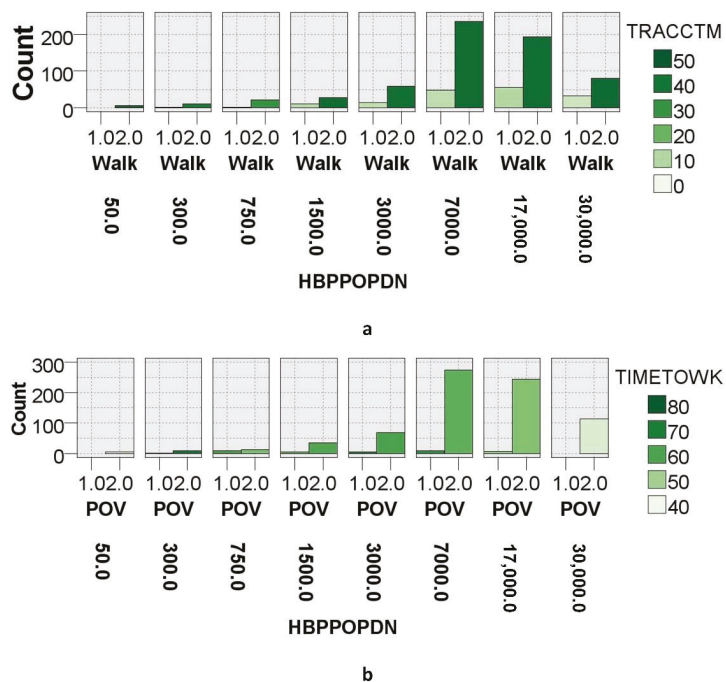


**Figure 5.** Associations between key variables and travel mode choice to reach public transit stations mediated by various variables. (**a**) The combined influence of population density and trip time to reach public transit stations on forecasting walking to reach PT stations; (**b**) The combined influence of population density and trip time to work on forecasting usage of POV to reach PT stations.

The substantial negative correlations between participants' housing population density and their usage of POVs to reach PT stations show that if people reside in high-density

areas with integrated public transportation, they will be discouraged from using POVs. Another variable, trip time to work (TIMETOWK), was found to mediate the impact of population density (HBPPOPDN) on using POVs to reach PT stations in the BN model (Figure 5b). The joint impact of these two variables on predicting POV usage to reach PT stations in California is shown in Figure 4b. In high-density locations, using POVs to reach PT stations is less likely (e.g., 7000–30,000 persons per square mile). When the commute time to work is between 42 and 57 min, lower trips using POVs occur. This suggests that TIMETOWK strengthens the HBPPOPDN's negative association with POV usage to reach the transport stations. It is worth noticing that a 57 min commute to work has a greater mediation effect than other TIMETOWK values on the effect of population density on not utilizing POVs to reach PT stations.

## 4. Discussions

The time it takes to walk to PT stops or stations is the most essential factor in people's choice to walk. Furthermore, it was shown that population density acted as a mediator for this effect. The POV model revealed population density as the most relevant component, which was mediated by the commute time to work. The findings are crucial because they show that planners should concentrate on population density, public transportation, and job locations when contemplating the replacement of POVs with walking to commute to PT stations for work. However, it is critical that the BN's outcomes are unaffected by the following variables: the health condition attributes, the individual and household attributes, the work trip attributes (excluding the trip time to work and trip time to PT stations), and the work attributes. Previous studies have deemed these factors relevant [17,29,31–33,36,47]. However, this research suggests that they may not be necessary to take into consideration. In terms of health condition attributes, a very limited number of studies show that this factor is essential in travel mode choice [1,47,48]. Additionally, this study did not find a substantial effect of these factors on mode choice for PT stations. This may be due to the fact that health conditions may have a greater impact on leisure walking in California than on work-related walking.

As mentioned above, population density emerged as the most important factor of POV usage to reach PT stations and a mediator of the effects of travel time to PT stations on walking to PT stations. This finding reflects the importance of this factor in studying of travel mode choice to PT stations. Several previous studies also stressed the relevance of this factor on travel mode choice [38,39]. Furthermore, according to Nigro, Bertolini and Moccia [99], population density has a significant impact on the mode of transportation to the nearest PT facility.

The density of the population is seen as a crucial component in the success of a PT operation [101]. Population density, particularly for pedestrians, is commonly cited as a factor that encourages more people to use PT. However, research has shown inconsistent outcomes. Higher densities tend to have a more compact land use and closer destinations, which makes walking more possible and beneficial. However, although some research suggests that short-distance walking to reach PT stations is dependent on density, wealth and other societal variables are progressively taking precedence after populational density reaches a certain level [102,103].

## 5. Conclusions and Recommendations

This study employed a Bayesian Network model to examine the relative importance of health condition, work trip, work, and individual and household attributes in trip mode choice to transit stations to go to work and their complex relationships with travel mode choice to transit stations to go to work in California, using data from 2017 NHTS. It is among the first to investigate how population density in California mediates the effects of time to transit stations on walking to transit stations to go to work and how time to go to work mediates the influences of population density on POV usage to reach PT stations to go to work. The findings provide positive consequences regarding densifying population

and land uses around transit stations for walking level growth to reach public transport stations in developed countries' cities, especially car-oriented ones.

The outcomes indicate that work trip attributes play a dominant role in walking to reach PT stations in California. People that have a short trip time to transit stations to go to work are more likely to walk to PT stations. This variable is the most important predictor in the walking model, contributing to more than 0.40 of the predictive power. With a decrease in the trip time to transit stations in California, the walking level for first-mile connections to reach the workplace is expected to grow faster. The determination of efficacious approaches to accelerate growth is key to sustainable transportation in California.

The factors that affect PT-related walking are similar to those that impact urban walking in general, especially in terms of built environment features [104,105]. The appealing aspects of PT, as well as the PT services offered and the transportation options available to individuals, influence how far someone is willing to walk to reach public transportation.

Land use and transportation strategies can be utilized by planners to change the built environment. The setting wherein PT-related walking takes place is defined by non-modifiable characteristics (e.g., alternative travel alternatives, culture, purpose, physical ability, and the weather). However, urban or transportation planning experts can employ changeable influences (such as density, land use, infrastructure quality, and trip length) to impact the distances people would walk to PT stations to reach the workplace.

Planners should consider promoting high-density development because this has a strong effect on PT-related walking lengths. This development makes the origins and endpoints much closer and increases the transit stations density. These may reduce the distance that individuals must walk to transit stations. Density has also been connected to enhanced walkability, which can attract more walkers by raising the proportion of people who walk to reach transit stations or broadening the catchment area around a transit station.

Typically, people prefer to walk to transit stations through more walkable routes [106]. The higher level of walkability and, in turn, shorter PT-related walking can be achieved through a higher level of street connectivity and lower-level detours [17,18,87,89]. In addition, the tolerable walking travel time of pedestrians to transit stations can be increased if the walkability at the micro-level is improved [107]. Street elements, such as lighting, seating areas, trees, and width of sidewalk, may increase the distances people are willing to walk [22,108,109].

Along with these built-environment solutions, various car-restrictive policies could assist to reduce the use of POVs for general use and reaching transit stations. These regulations can be implemented particularly well in high-density areas, as low-density areas may lack enough PT and walking infrastructures. As a result, the only way to reach transit stations is by using a POV.

This study has a few limitations that deserve comment. First, this study utilized the 2017 NHTS dataset, which does not consider biking, micro-mobility, and ride-sourcing exclusively as modes to reach PT stations. Thus, future studies can apply the BN algorithm considering these modes and using different datasets. Second, the NHTS includes a few variables of the built environment. Hence, it is recommended that future studies develop BN models using more comprehensive datasets. Finally, this study was conducted in a car-oriented setting. Thus, people who use walking to reach public transit stations were underrepresented. Therefore, the outcomes of this study should be transferred cautiously to other cities, especially European ones.

## References

1. Martin, A.; Panter, J.; Suhrcke, M.; Ogilvie, D. Impact of changes in mode of travel to work on changes in body mass index: Evidence from the British Household Panel Survey. *J. Epidemiol. Community Health* **2015**, *69*, 753–761. [CrossRef] [PubMed]
2. Chinnam, R.B.; Murat, A.E. Community-Aware Charging Station Network Design for Electrified Vehicles in Urban Areas: Reducing Congestion, Emissions, Improving Accessibility, and Promoting Walking, Bicycling, and Use of Public Transportation. 2016. Available online: https://rosap.ntl.bts.gov/view/dot/31590 (accessed on 1 January 2022).
3. McKenzie, B. *Modes Less Traveled: Bicycling and Walking to Work in the United States, 2008–2012*; US Department of Commerce, Economics and Statistics Administration: Washington, DC, USA, 2014.
4. Macioszek, E.; Kurek, A. The analysis of the factors determining the choice of park and ride facility using a multinomial logit model. *Energies* **2021**, *14*, 203. [CrossRef]
5. Henry, E.; Furno, A.; El Faouzi, N.-E.; Rey, D. Locating park-and-ride facilities for resilient on-demand urban mobility. *Transp. Res. Part E Logist. Transp. Rev.* **2022**, *158*, 102557. [CrossRef]
6. Kumar, P.; Khani, A. Adaptive park-and-ride choice on time-dependent stochastic multimodal transportation network. *Netw. Spat. Econ.* **2021**, *21*, 771–800. [CrossRef]
7. Rezaei, S.; Khojandi, A.; Haque, A.M.; Brakewood, C.; Jin, M.; Cherry, C.R. Park-and-ride facility location optimization: A case study for Nashville, Tennessee. *Transp. Res. Interdiscip. Perspect.* **2022**, *13*, 100578. [CrossRef]
8. Switkes, F. *Getting There is Half the Problem: Removing Obstacles to Accessing Rail Transit*; Massachusetts Institute of Technology: Cambridge, MA, USA, 2003.
9. Kanthavel, R.; Sangeetha, S.; Keerthana, K. Design of smart public transport assist system for metropolitan city Chennai. *Int. J. Intell. Netw.* **2021**, *2*, 57–63. [CrossRef]
10. Rietveld, P. Non-motorised modes in transport systems: A multimodal chain perspective for The Netherlands. *Transp. Res. Part D Transp. Environ.* **2000**, *5*, 31–36. [CrossRef]
11. Weinstein Agrawal, A.; Schlossberg, M.; Irvin, K. How far, by which route and why? A spatial analysis of pedestrian preference. *J. Urban Des.* **2008**, *13*, 81–98. [CrossRef]
12. Tight, M.; Rajé, F.; Timms, P. Car-free urban areas: A radical solution to the last mile problem or a step too far? *Built Environ.* **2016**, *42*, 603–616. [CrossRef]
13. Ali, M.; de Azevedo, A.R.G.; Marvila, M.T.; Khan, M.I.; Memon, A.M.; Masood, F.; Almahbashi, N.M.; Shad, M.K.; Khan, M.A.; Fediuk, R.; et al. The Influence of COVID-19-Induced Daily Activities on Health Parameters—A Case Study in Malaysia. *Sustainability* **2021**, *13*, 7465. [CrossRef]
14. Ali, M.; Dharmowijoyo, D.B.; Harahap, I.S.; Puri, A.; Tanjung, L.E. Travel behaviour and health: Interaction of Activity-Travel Pattern, Travel Parameter and Physical Intensity. *Solid State Technol.* **2020**, *63*, 4026–4039.
15. Zhao, T.-H.; Shi, L.; Chu, Y.-M. Convexity and concavity of the modified Bessel functions of the first kind with respect to Hölder means. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas* **2020**, *114*, 96. [CrossRef]
16. Næss, P.; Peters, S.; Stefansdottir, H.; Strand, A. Causality, not just correlation: Residential location, transport rationales and travel behavior across metropolitan contexts. *J. Transp. Geogr.* **2018**, *69*, 181–195. [CrossRef]
17. Kim, K.W.; Lee, D.W.; Chun, Y.H. A comparative study on the service coverages of subways and buses. *KSCE J. Civ. Eng.* **2010**, *14*, 915–922. [CrossRef]
18. El-Geneidy, A.; Grimsrud, M.; Wasfi, R.; Tétreault, P.; Surprenant-Legault, J. New evidence on walking distances to transit stops: Identifying redundancies and gaps using variable service areas. *Transportation* **2014**, *41*, 193–210. [CrossRef]
19. Patterson, R.; Webb, E.; Millett, C.; Laverty, A. Physical activity accrued as part of public transport use in England. *J. Public Health* **2019**, *41*, 222–230. [CrossRef] [PubMed]
20. Ali, M.; Dharmowijoyo, D.B.E.; de Azevedo, A.R.G.; Fediuk, R.; Ahmad, H.; Salah, B. Time-Use and Spatio-Temporal Variables Influence on Physical Activity Intensity, Physical and Social Health of Travelers. *Sustainability* **2021**, *13*, 12226. [CrossRef]
21. Tabatabaee, S.; Aghaabbasi, M.; Mahdiyar, A.; Zainol, R.; Ismail, S. Measurement Quality Appraisal Instrument for Evaluation of Walkability Assessment Tools Based on Walking Needs. *Sustainability* **2021**, *13*, 11342. [CrossRef]
22. Aghaabbasi, M.; Moeinaddini, M.; Shah, M.Z.; Asadi-Shekari, Z. A new assessment model to evaluate the microscale sidewalk design factors at the neighbourhood level. *J. Transp. Health* **2017**, *5*, 97–112. [CrossRef]
23. Aghaabbasi, M.; Moeinaddini, M.; Zaly Shah, M.; Asadi-Shekari, Z.; Arjomand Kermani, M. Evaluating the capability of walkability audit tools for assessing sidewalks. *Sustain. Cities Soc.* **2018**, *37*, 475–484. [CrossRef]

24. Qian, Y.; Aghaabbasi, M.; Ali, M.; Alqurashi, M.; Salah, B.; Zainol, R.; Moeinaddini, M.; Hussein, E.E. Classification of Imbalanced Travel Mode Choice to Work Data Using Adjustable SVM Model. *Appl. Sci.* **2021**, *11*, 11916. [CrossRef]
25. Zhao, T.-H.; Shen, Z.-H.; Chu, Y.-M. Sharp power mean bounds for the lemniscate type means. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas* **2021**, *115*, 175. [CrossRef]
26. Morency, C.; Trépanier, M.; Demers, M. Walking to transit: An unexpected source of physical activity. *Transp. Policy* **2011**, *18*, 800–806. [CrossRef]
27. Wasfi, R.A.; Ross, N.A.; El-Geneidy, A.M. Achieving recommended daily physical activity levels through commuting by public transportation: Unpacking individual and contextual influences. *Health Place* **2013**, *23*, 18–25. [CrossRef]
28. Zhao, T.-H.; Wang, M.-K.; Zhang, W.; Chu, Y.-M. Quadratic transformation inequalities for Gaussian hypergeometric function. *J. Inequal. Appl.* **2018**, *2018*, 251. [CrossRef]
29. García-Palomares, J.C.; Gutiérrez, J.; Cardozo, O.D. Walking accessibility to public transport: An analysis based on microdata and GIS. *Environ. Plan. B Plan. Des.* **2013**, *40*, 1087–1102. [CrossRef]
30. Johar, A.; Jain, S.; Garg, P.; Gundaliya, P. A study for commuter walk distance from bus stops to different destination along routes in Delhi. *Eur. Transp.* **2015**, *59*, 1–15.
31. Alshalalfah, B.; Shalaby, A.S. Case study: Relationship of walk access distance to transit with service, travel, and personal characteristics. *J. Urban Plan. Dev.* **2007**, *133*, 114–118. [CrossRef]
32. Besser, L.M.; Dannenberg, A.L. Walking to public transit: Steps to help meet physical activity recommendations. *Am. J. Prev. Med.* **2005**, *29*, 273–280. [CrossRef]
33. He, J.; Zhang, R.; Huang, X.; Xi, G. Walking access distance of metro passengers and relationship with demographic characteristics: A case study of Nanjing metro. *Chin. Geogr. Sci.* **2018**, *28*, 612–623. [CrossRef]
34. Durand, C.P.; Gabriel, K.K.P.; Hoelscher, D.M.; Kohl, H.W. Transit use by children and adolescents: An overlooked source of and opportunity for physical activity? *J. Phys. Act. Health* **2016**, *13*, 861–866. [CrossRef] [PubMed]
35. Zhao, T.-H.; He, Z.-Y.; Chu, Y.-M. Sharp bounds for the weighted Hölder mean of the zero-balanced generalized complete elliptic integrals. *Comput. Methods Funct. Theory* **2021**, *21*, 413–426. [CrossRef]
36. Freeland, A.L.; Banerjee, S.N.; Dannenberg, A.L.; Wendel, A.M. Walking associated with public transit: Moving toward increased physical activity in the United States. *Am. J. Public Health* **2013**, *103*, 536–542. [CrossRef] [PubMed]
37. Zhao, T.-H.; Zhou, B.-C.; Wang, M.-K.; Chu, Y.-M. On approximating the quasi-arithmetic mean. *J. Inequal. Appl.* **2019**, *2019*, 1–12. [CrossRef]
38. Sun, G.; Zacharias, J.; Ma, B.; Oreskovic, N.M. How do metro stations integrate with walking environments? Results from walking access within three types of built environment in Beijing. *Cities* **2016**, *56*, 91–98. [CrossRef]
39. Yu, C.-Y.; Lin, H.-C. Exploring factors regarding transit-related walking and walking duration. *J. Phys. Act. Health* **2016**, *13*, 1220–1229. [CrossRef]
40. Zhao, T.-H.; Wang, M.-K.; Chu, Y.-M. Monotonicity and convexity involving generalized elliptic integral of the first kind. *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A. Matemáticas* **2021**, *115*, 46. [CrossRef]
41. Kim, T.; Sohn, D.-W.; Choo, S. An analysis of the relationship between pedestrian traffic volumes and built environment around metro stations in Seoul. *KSCE J. Civ. Eng.* **2017**, *21*, 1443–1452. [CrossRef]
42. Paydar, M.; Fard, A.K.; Khaghani, M.M. Walking toward metro stations: The contribution of distance, attitudes, and perceived built environment. *Sustainability* **2020**, *12*, 10291. [CrossRef]
43. Yu, C.-Y.; Lin, H.-C. Transit-related walking to work in promoting physical activity. *J. Phys. Act. Health* **2015**, *12*, 483–489. [CrossRef]
44. Clark, A.F.; Scott, D.M.; Yiannakoulias, N. Examining the relationship between active travel, weather, and the built environment: A multilevel approach using a GPS-enhanced dataset. *Transportation* **2014**, *41*, 325–338. [CrossRef]
45. Pan, H.; Wei, P. 11. Effects of rail transit proximity emissions in the suburbs of Shanghai on commuter $CO_2$. In *Handbook on Transport and Urban Transformation in China*; Edward Elgar Publishing: Northampton, MA, USA, 2020; p. 182.
46. Wei, P.; Pan, H. Research on individual carbon dioxide emissions of commuting in peri-urban area of metropolitan cities—An empirical study in Shanghai. *Transp. Res. Procedia* **2017**, *25*, 3459–3478. [CrossRef]
47. Day, K.; Loh, L.; Ruff, R.R.; Rosenblum, R.; Fischer, S.; Lee, K.K. Does bus rapid transit promote walking? An examination of New York City's select bus service. *J. Phys. Act. Health* **2014**, *11*, 1512–1516. [CrossRef]
48. Zhao, T.-H.; Yang, Z.-H.; Chu, Y.-M. Monotonicity properties of a function involving the psi function with applications. *J. Inequal. Appl.* **2015**, *2015*, 193. [CrossRef]
49. Sadeghiravesh, M.H.; Khosravi, H.; Abolhasani, A.; Ghodsi, M.; Mosavi, A. Fuzzy Logic Model to Assess Desertification Intensity Based on Vulnerability Indices. *Acta Polytech. Hung.* **2021**, *18*, 7–24. [CrossRef]
50. Rafiei-Sardooi, E.; Azareh, A.; Choubin, B.; Mosavi, A.H.; Clague, J.J. Evaluating urban flood risk using hybrid method of TOPSIS and machine learning. *Int. J. Disaster Risk Reduct.* **2021**, *66*, 102614. [CrossRef]
51. Pinter, G.; Mosavi, A.; Felde, I. Artificial intelligence for modeling real estate price using call detail records and hybrid machine learning approach. *Entropy* **2020**, *22*, 1421. [CrossRef]
52. Holtzclaw, J.; Clear, R.; Dittmar, H.; Goldstein, D.; Haas, P. Location efficiency: Neighborhood and socio-economic characteristics determine auto ownership and use-studies in Chicago, Los Angeles and San Francisco. *Transp. Plan. Technol.* **2002**, *25*, 1–27. [CrossRef]

53. Shay, E.; Khattak, A.J. Household travel decision chains: Residential environment, automobile ownership, trips and mode choice. *Int. J. Sustain. Transp.* **2012**, *6*, 88–110. [CrossRef]
54. Zhang, W.; Zhao, Y.; Cao, X.J.; Lu, D.; Chai, Y. Nonlinear effect of accessibility on car ownership in Beijing: Pedestrian-scale neighborhood planning. *Transp. Res. Part D Transp. Environ.* **2020**, *86*, 102445. [CrossRef]
55. Chen, E.; Ye, Z.; Wu, H. Nonlinear effects of built environment on intermodal transit trips considering spatial heterogeneity. *Transp. Research Part D Transp. Environ.* **2021**, *90*, 102677. [CrossRef]
56. Zhao, T.H.; Khan, M.I.; Chu, Y.M. Artificial neural networking (ANN) analysis for heat and entropy generation in flow of non-Newtonian fluid between two rotating disks. *Math. Methods Appl. Sci.* **2021**. [CrossRef]
57. Zandi, P.; Rahmani, M.; Khanian, M.; Mosavi, A. Agricultural risk management using fuzzy TOPSIS analytical hierarchy process (AHP) and failure mode and effects analysis (FMEA). *Agriculture* **2020**, *10*, 504. [CrossRef]
58. Tavoosi, J.; Zhang, C.; Mohammadzadeh, A.; Mobayen, S.; Mosavi, A.H. Medical image interpolation using recurrent type-2 fuzzy neural network. *Front. Neuroinform.* **2021**, *15*, 667375. [CrossRef] [PubMed]
59. Ma, T.; Aghaabbasi, M.; Ali, M.; Zainol, R.; Jan, A.; Mohamed, A.M.; Mohamed, A. Nonlinear Relationships between Vehicle Ownership and Household Travel Characteristics and Built Environment Attributes in the US Using the XGBT Algorithm. *Sustainability* **2022**, *14*, 3395. [CrossRef]
60. Ding, C.; Cao, X.J.; Næss, P. Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. *Transp. Res. Part A Policy Pract.* **2018**, *110*, 107–117. [CrossRef]
61. Stylianou, K.; Dimitriou, L.; Abdel-Aty, M. Big data and road safety: A comprehensive review. In *Mobility Patterns, Big Data and Transport Analytics*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 297–343.
62. Mousavi, S.M.; Ghasemi, M.; Dehghan Manshadi, M.; Mosavi, A. Deep learning for wave energy converter modeling using long short-term memory. *Mathematics* **2021**, *9*, 871. [CrossRef]
63. Meiabadi, M.S.; Moradi, M.; Karamimoghadam, M.; Ardabili, S.; Bodaghi, M.; Shokri, M.; Mosavi, A.H. Modeling the producibility of 3D printing in polylactic acid using artificial neural networks and fused filament fabrication. *Polymers* **2021**, *13*, 3219. [CrossRef]
64. Khosravi, K.; Golkarian, A.; Booij, M.J.; Barzegar, R.; Sun, W.; Yaseen, Z.M.; Mosavi, A. Improving daily stochastic streamflow prediction: Comparison of novel hybrid data-mining algorithms. *Hydrol. Sci. J.* **2021**, *66*, 1457–1474. [CrossRef]
65. Ghaemi, A.; Zhian, T.; Pirzadeh, B.; Hashemi Monfared, S.; Mosavi, A. Reliability-based design and implementation of crow search algorithm for longitudinal dispersion coefficient estimation in rivers. *Environ. Sci. Pollut. Res.* **2021**, *28*, 35971–35990. [CrossRef]
66. Ebrahimi-Khusfi, Z.; Taghizadeh-Mehrjardi, R.; Roustaei, F.; Ebrahimi-Khusfi, M.; Mosavi, A.H.; Heung, B.; Soleimani-Sardo, M.; Scholten, T. Determining the contribution of environmental factors in controlling dust pollution during cold and warm months of western Iran using different data mining algorithms and game theory. *Ecol. Indic.* **2021**, *132*, 108287. [CrossRef]
67. Aghaabbasi, M.; Shekari, Z.A.; Shah, M.Z.; Olakunle, O.; Armaghani, D.J.; Moeinaddini, M. Predicting the use frequency of ride-sourcing by off-campus university students through random forest and Bayesian network techniques. *Transp. Res. Part A Policy Pract.* **2020**, *136*, 262–281. [CrossRef]
68. Peng, Y.; Ghahnaviyeh, M.B.; Ahmad, M.N.; Abdollahi, A.; Bagherzadeh, S.A.; Azimy, H.; Mosavi, A.; Karimipour, A. Analysis of the effect of roughness and concentration of $Fe_3O_4$/water nanofluid on the boiling heat transfer using the artificial neural network: An experimental and numerical study. *Int. J. Therm. Sci.* **2021**, *163*, 106863. [CrossRef]
69. Panahi, F.; Ehteram, M.; Ahmed, A.N.; Huang, Y.F.; Mosavi, A.; El-Shafie, A. Streamflow prediction with large climate indices using several hybrid multilayer perceptrons and copula Bayesian model averaging. *Ecol. Indic.* **2021**, *133*, 108285. [CrossRef]
70. Mousavi, S.P.; Atashrouz, S.; Nait Amar, M.; Hemmati-Sarapardeh, A.; Mohaddespour, A.; Mosavi, A. Viscosity of Ionic Liquids: Application of the Eyring's Theory and a Committee Machine Intelligent System. *Molecules* **2021**, *26*, 156. [CrossRef] [PubMed]
71. Chen, Y.; Aghaabbasi, M.; Ali, M.; Anciferov, S.; Sabitov, L.; Chebotarev, S.; Nabiullina, K.; Sychev, E.; Fediuk, R.; Zainol, R. Hybrid Bayesian Network Models to Investigate the Impact of Built Environment Experience before Adulthood on Students' Tolerable Travel Time to Campus: Towards Sustainable Commute Behavior. *Sustainability* **2022**, *14*, 325. [CrossRef]
72. Musarat, M.A.; Alaloul, W.S.; Rabbani, M.B.A.; Ali, M.; Altaf, M.; Fediuk, R.; Vatin, N.; Klyuev, S.; Bukhari, H.; Sadiq, A.; et al. Kabul River Flow Prediction Using Automated ARIMA Forecasting: A Machine Learning Approach. *Sustainability* **2021**, *13*, 10720. [CrossRef]
73. Nafees, A.; Amin, M.N.; Khan, K.; Nazir, K.; Ali, M.; Javed, M.F.; Aslam, F.; Musarat, M.A.; Vatin, N.I. Modeling of Mechanical Properties of Silica Fume-Based Green Concrete Using Machine Learning Techniques. *Polymers* **2022**, *14*, 30. [CrossRef]
74. Zhao, T.-H.; Wang, M.-K.; Chu, Y.-M. Concavity and bounds involving generalized elliptic integral of the first kind. *J. Math. Inequal.* **2021**, *15*, 701–724. [CrossRef]
75. Tavoosi, J.; Suratgar, A.A.; Menhaj, M.B.; Mosavi, A.; Mohammadzadeh, A.; Ranjbar, E. Modeling renewable energy systems by a self-evolving nonlinear consequent part recurrent type-2 fuzzy system for power prediction. *Sustainability* **2021**, *13*, 3301. [CrossRef]
76. Tareeq, S.M.; Inamura, T. A sample discarding strategy for rapid adaptation to new situation based on Bayesian behavior learning. In Proceedings of the 2008 IEEE International Conference on Robotics and Biomimetics, Bangkok, Thailand, 22–25 February 2009; pp. 1950–1955.
77. Wu, Q.; Yang, C.; Gao, X.; He, P.; Chen, G. EPAB: Early pattern aware bayesian model for social content popularity prediction. In Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM), Singapore, 17–20 November 2018; pp. 1296–1301.

78. Liu, T.; Zhang, Y.; Chen, J.; Shen, H. Discovery of association rule of learning action based on Bayesian network. In Proceedings of the 2018 9th International Conference on Information Technology in Medicine and Education (ITME), Hangzhou, China, 19–21 October 2018; pp. 466–470.

79. Tao, W.; Aghaabbasi, M.; Ali, M.; Almaliki, A.H.; Zainol, R.; Almaliki, A.A.; Hussein, E.E. An Advanced Machine Learning Approach to Predicting Pedestrian Fatality Caused by Road Crashes: A Step toward Sustainable Pedestrian Safety. *Sustainability* **2022**, *14*, 2436. [CrossRef]

80. Zhao, T.-H.; Wang, M.-K.; Chu, Y.-M. A sharp double inequality involving generalized complete elliptic integral of the first kind. *AIMS Math.* **2020**, *5*, 4512–4528. [CrossRef]

81. Taghizadeh-Mehrjardi, R.; Emadi, M.; Cherati, A.; Heung, B.; Mosavi, A.; Scholten, T. Bio-inspired hybridization of artificial neural networks: An application for mapping the spatial distribution of soil texture fractions. *Remote Sens.* **2021**, *13*, 1025. [CrossRef]

82. Ratanawaraha, A.; Chalermpong, S.; Chullabodhi, C. Walking distance of commuters after modal shift to rail transit in Bangkok. *J. East. Asia Soc. Transp. Stud.* **2015**, *11*, 1467–1478.

83. Rodriguez-Gonzalez, M.; Aguero-Valverde, J. Walking distances from home to bus stops in San José, Costa Rica: Real, perceived, and stated-preference distances. *Adv. Transp. Stud.* **2017**, *43*, 63–74.

84. Townsend, C.; Zacharias, J. Built environment and pedestrian behavior at rail rapid transit stations in Bangkok. *Transportation* **2010**, *37*, 317–330. [CrossRef]

85. Vandebona, U.; Tsukaguchi, H. Impact of urbanization on user expectations related to public transport accessibility. *Int. J. Urban Sci.* **2013**, *17*, 199–211. [CrossRef]

86. Voss, C.; Winters, M.; Frazer, A.; McKay, H. School-travel by public transit: Rethinking active transportation. *Prev. Med. Rep.* **2015**, *2*, 65–70. [CrossRef]

87. Wang, J.; Cao, X. Exploring built environment correlates of walking distance of transit egress in the Twin Cities. *J. Transp. Geogr.* **2017**, *64*, 132–138. [CrossRef]

88. Xi, Y.; Saxe, S.; Miller, E. Accessing the subway in Toronto, Canada: Access mode and catchment areas. *Transp. Res. Rec.* **2016**, *2543*, 52–61. [CrossRef]

89. Zacharias, J.; Zhao, Q. Local environmental factors in walking distance at metro stations. *Public Transp.* **2018**, *10*, 91–106. [CrossRef]

90. Zhao, J.; Deng, W.; Song, Y.; Zhu, Y. What influences Metro station ridership in China? Insights from Nanjing. *Cities* **2013**, *35*, 114–124. [CrossRef]

91. Zuo, T.; Wei, H.; Rohne, A. Determining transit service coverage by non-motorized accessibility to transit: Case study of applying GPS data in Cincinnati metropolitan area. *J. Transp. Geogr.* **2018**, *67*, 1–11. [CrossRef]

92. Friedman, N.; Geiger, D.; Goldszmidt, M. Bayesian network classifiers. *Mach. Learn.* **1997**, *29*, 131–163. [CrossRef]

93. Murphy, P.M. UCI repository of machine learning databases. 1994. Available online: https://archive.ics.uci.edu/ml/datasets.php (accessed on 1 January 2022).

94. Chow, C.; Liu, C. Approximating discrete probability distributions with dependence trees. *IEEE Trans. Inf. Theory* **1968**, *14*, 462–467. [CrossRef]

95. Cervero, R.; Kockelman, K. Travel demand and the 3Ds: Density, diversity, and design. *Transp. Res. Part D: Transp. Environ.* **1997**, *2*, 199–219. [CrossRef]

96. Targa, F.; Clifton, K.J. Built environment and nonmotorized travel: Evidence from Baltimore city using the NHTS. *J. Transp. Stat.* **2005**, *8*, 55–70.

97. Schneider, R.J. Local environment characteristics associated with walking and taking transit to shopping districts. *J. Transp. Land Use* **2015**, *8*, 125–147. [CrossRef]

98. Chia, J.; Lee, J.; Kamruzzaman, M. Walking to public transit: Exploring variations by socioeconomic status. *Int. J. Sustain. Transp.* **2016**, *10*, 805–814. [CrossRef]

99. Nigro, A.; Bertolini, L.; Moccia, F.D. Land use and public transport integration in small cities and towns: Assessment methodology and application. *J. Transp. Geogr.* **2019**, *74*, 110–124. [CrossRef]

100. Sun, B.; Yin, C. Impacts of a multi-scale built environment and its corresponding moderating effects on commute duration in China. *Urban Stud.* **2019**, *57*, 2115–2130. [CrossRef]

101. Calthorpe, P. *The Next American Metropolis: Ecology, Community, and the American Dream*; Princeton Architectural Press: New York, NY, USA, 1993.

102. Marquet, O.; Miralles-Guasch, C. The Walkable city and the importance of the proximity environments for Barcelona's everyday mobility. *Cities* **2015**, *42*, 258–266. [CrossRef]

103. Forsyth, A.; Oakes, J.M.; Schmitz, K.H.; Hearst, M. Does residential density increase walking and other physical activity? *Urban Stud.* **2007**, *44*, 679–697. [CrossRef]

104. Saelens, B.E.; Handy, S.L. Built environment correlates of walking: A review. *Med. Sci. Sports Exerc.* **2008**, *40*, S550. [CrossRef]

105. Van Holle, V.; Deforche, B.; Van Cauwenberg, J.; Goubert, L.; Maes, L.; Van de Weghe, N.; De Bourdeaudhuij, I. Relationship between the physical environment and different domains of physical activity in European adults: A systematic review. *BMC Public Health* **2012**, *12*, 1–17. [CrossRef]

106. Jiang, Y.; Zegras, P.C.; Mehndiratta, S. Walk the line: Station context, corridor type and bus rapid transit walk access in Jinan, China. *J. Transp. Geogr.* **2012**, *20*, 1–14. [CrossRef]
107. Asadi-Shekari, Z.; Moeinaddini, M.; Aghaabbasi, M.; Cools, M.; Zaly Shah, M. Exploring effective micro-level items for evaluating inclusive walking facilities on urban streets (applied in Johor Bahru, Malaysia). *Sustain. Cities Soc.* **2019**, *49*, 101563. [CrossRef]
108. Aghaabbasi, M.; Shah, M.Z.; Zainol, R. Investigating the Use of Active Transportation Modes Among University Employees Through an Advanced Decision Tree Algorithm. *Civ. Sustain. Urban Eng.* **2021**, *1*, 26–49. [CrossRef]
109. Aghaabbasi, M.; Moeinaddini, M.; Asadi-Shekari, Z.; Shah, M.Z. The equitable use concept in sidewalk design. *Cities* **2019**, *88*, 181–190. [CrossRef]

**MDPI**