



Prediction of walking occurrence and frequency among the residents of city community housing projects using machine learning techniques

Rosilawati Zainol^{a,f,g,*}, Tin Tin Su^b, Hazreen Abdul Majid^c,
Cheong Peng Au Yong^d, Mahdi Aghaabbasi^e, Nithiah Thangiah^c, Dan He^a,
Fadly Usman^h

^a Department of Urban and Regional Planning, Centre for Sustainable Urban Planning and Real Estate (SUPRE), Faculty of Built Environment, Universiti Malaya, Malaysia

^b Jeffrey Cheah School of Medicine and Health Sciences, Victorian Heart Institute (VHI), Monash University, Malaysia

^c Department of Social and Preventive Medicine, Faculty of Medicine, Universiti Malaya, Malaysia

^d Department of Building Surveying, Faculty of Built Environment, Universiti Malaya, Malaysia

^e Urban Digital Twin Lab, School of Modeling, Simulation, and Training, University of Central Florida, 3100 Technology Parkway, Orlando, 32826, Florida, USA

^f Centre for Civilisational Dialogue, Universiti Malaya, Malaysia

^g Centre for Transportation Research (CTR), Faculty of Engineering, Universiti Malaya, Malaysia

^h Department of Urban and Regional Planning, Universitas Brawijaya, Indonesia

ARTICLE INFO

Keywords:

Active transport
Physical activity
Bayesian network model
Walking behavior
Built environment and walkability

ABSTRACT

Introduction: Walking is one of the most affordable and accessible forms of daily physical activity, yet it remains an unsafe and under-supported mode of transport in many low-income urban neighborhoods. Residents of public housing communities often face inadequate pedestrian infrastructure, traffic safety risks, and limited access to essential services, all of which discourage walking and exacerbate health inequalities. Understanding the complex interplay of these factors is critical for improving mobility and public health outcomes among vulnerable urban populations.

Methods: This study applies a Bayesian Network model to predict the occurrence and frequency of walking among 680 residents across five low-income public housing projects (PPRs) in Kuala Lumpur, Malaysia. Using cross-sectional data on individual demographics, travel behavior, built environment perceptions, and health status, the model identifies probabilistic relationships between key factors influencing pedestrian activity.

Results: Results indicate that gender, employment status, and perceived walkability significantly affect the likelihood and frequency of walking trips. Notably, the perceived quality of pedestrian infrastructure and accessibility to public facilities also emerged as influential predictors.

Conclusions: The results highlight the urgent need to design safer and more inclusive walking environments in low-income urban areas. The study demonstrates the practical utility of Bayesian

* Corresponding author. Department of Urban and Regional Planning, Centre for Sustainable Urban Planning and Real Estate (SUPRE), Faculty of Built Environment, Universiti Malaya, Malaysia.

E-mail addresses: rosilawatizai@um.edu.my (R. Zainol), TinTin.Su@monash.edu (T.T. Su), hazreen@um.edu.my (H.A. Majid), auyongcp@um.edu.my (C.P. Au Yong), mahdi.ghaabbasi@ucf.edu (M. Aghaabbasi), nithiah@um.edu.my (N. Thangiah), s2038027@siswa.um.edu.my (D. He), fadlypwk@ub.ac.id (F. Usman).

<https://doi.org/10.1016/j.jth.2026.102340>

Received 30 June 2025; Received in revised form 6 April 2026; Accepted 2 May 2026

Available online 10 May 2026

2214-1405/© 2026 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

modeling in capturing interdependent factors influencing walking behavior and provides actionable insights for planners and policymakers seeking to promote equitable, health-oriented, and pedestrian-friendly urban mobility.

1. Introduction

Research on walking is an intersection between transport planning, public health, and sustainability science and many studies have been conducted on this topic across the globe (Abusaada and Elshater, 2024; Řišová and Madajová, 2020). Walking serves not only as a form of physical activity but also as a healthy and sustainable mode of transport (Abusaada and Elshater, 2024; Aghaabbasi et al., 2017a). Walking can help reduce the negative impacts of modern sedentary lifestyles, such as overweight and obesity (Larsen, 2015; McDonald, 2007). In addition to physical health benefits, walking has been linked to improved mental well-being and reduced stress levels (Guo et al., 2019). Walkable environments also strengthen social cohesion by facilitating everyday social interactions and community connectedness (Leyden, 2003). These broader benefits are consistent with the global sustainability agenda, particularly SDG 3 and SDG 11, which highlight the importance of promoting active and healthy urban lifestyles (Lee et al., 2016).

Governments pay special attention to the people with low socioeconomic status across the globe to provide them with more fair living conditions. One of the most common solutions to support the lower-income groups is to provide them with cheap housing units. In Malaysia, the government has launched the Community Housing Projects (PPRs) since 1996. The PPRs were only provided to households with earnings below MYR 2,500, at least one child, and those who do not have a home. Totally, 63,377 units were constructed within the PPRs scheme where 34,106 (53.81%) of these units were built in Kuala Lumpur (Su et al., 2019). While it is essential to provide lower-income people with affordable houses, it is also vital to ensure that these residential units are able to form a suitable society and supporting sustainable living conditions. A limited number of studies have been conducted on the life quality of PPRs' residents. Furthermore, studies rarely examine their travel behaviour specifically through the lens of walking as a sustainable and health-promoting mode of transport.

The body of literature is large on the issues of walking and pedestrian-built environment. These studies mostly employed the traditional statistical methods. Traditional statistical methods are based on several rigid assumptions—such as linear relationships among variables and the absence of outliers (Rashidi et al., 2014; Stylianou et al., 2019)—which are often violated when applied to travel mode choice datasets due to the large scale and complexity of such data (Aghaabbasi et al., 2020). Machine learning techniques are suitable alternatives of traditional statistical methods which remedy all above-mentioned shortcomings of the conventional statistical techniques. However, to date, a very limited number of studies applied these novel techniques for analysing the issues related to walking and pedestrian environment. Specifically, to the authors' best of knowledge, no study has employed these techniques for predicting the probability of walking occurrence and frequency.

1.1. Influential factors on mode choice and walking

Extensive transport studies have explored various determinants influencing individuals' choices of travel modes. Commonly analyzed predictors include sociodemographic attributes, household profiles, travel-related purposes and patterns, modal characteristics, and elements of the built environment. Key sociodemographic factors such as age, gender, education, vehicle ownership, and income are among the most influential variables shaping travel behavior (Bernetti et al., 2008; Cheng et al., 2016; Kim and Ulfarsson, 2004; Lind et al., 2015). These attributes also influence destination choices and trip frequency, as shown by Yang et al. (2010), who found that Chinese women were more likely than men to select workplaces near home due to family responsibilities.

Built environment attributes play a crucial role in promoting walking and cycling as active modes of transport. Macro-scale factors such as land-use density and diversity, and micro-scale elements such as sidewalks, amenities, and safety features, jointly determine walkability (Aghaabbasi et al., 2017b). Literature emphasizes that micro-scale design factors, such as sidewalk continuity, crossings, and perceived safety, are more critical in shaping pedestrian behavior than macro-scale urban form (De Cambra, 2012; Southworth, 2005). Consequently, residents' perceptions of safety, comfort, and accessibility can strongly affect their decision to walk.

1.2. Studies on walking behaviours in Malaysia

Despite increasing global attention to walking as an equitable and sustainable transport mode, research on walking behavior in Malaysia remains limited. Only a few empirical studies have examined walking in Malaysian neighborhoods. For instance, Wan Omar et al. (2013) identified health beliefs and safety perceptions as key determinants of walking. Tareef Hayat Khan et al. (2015) analyzed how social interaction and urban design influence pedestrian movement in terrace housing estates. Mahmoud et al., (2016) assessed walkability and resident satisfaction in planned neighborhoods, finding low levels of walking despite favorable design attributes.

These studies suffer from small sample sizes (e.g., only 30 respondents in Mahmoud et al., 2016) and rely on traditional statistical methods with limited predictive power. As a result, their findings cannot be generalized to broader populations. Moreover, no prior research has examined the walking behavior of residents in Malaysia's PPR public housing developments, where low-income populations face unique challenges related to the built environment and mobility. Understanding walking behavior in these communities is essential for improving transport equity, safety, and public health outcomes. These limitations highlight a critical knowledge gap in understanding walking behaviour among low-income PPR residents, especially using predictive and data-driven analytical

approaches.

1.3. Walking behaviour and machine learning

Machine learning (ML) methods have gained prominence for modeling complex travel behavior, addressing the limitations of traditional statistical approaches that assume linearity and homogeneity. Previous studies have mainly used ML for pedestrian detection, crash prediction, and crossing behavior, employing algorithms such as ANN, SVM, RF, DT, and XGBoost (Table 1). However, very few have applied these methods to analyze the factors influencing walking mode choice and frequency.

To fill this methodological and contextual gap, this study employs a Bayesian Network (BN) model to predict the probability of walking occurrence and frequency among residents of PPR housing in Kuala Lumpur. This approach enables capturing interdependencies between demographic, behavioral, and built environment factors.

Furthermore, by focusing on low-income urban communities, the study aligns with global frameworks such as the WHO Global Action Plan on Physical Activity (2018–2030), which promotes inclusive and active transport environments, and the United Nations Sustainable Development Goals (SDG 3: Good Health and Well-Being; SDG 11: Sustainable Cities and Communities). Positioning the research within these frameworks highlights its contribution to advancing equity and sustainability in urban transport planning.

In Malaysia, the government provided cheap housing units to people with low socioeconomic status under the Community Housing Projects (PPRs). Despite the large population residing in these developments, limited research has examined their residents' mobility patterns or the quality of their pedestrian environment. Walking represents not only a sustainable mode of transport but also a key indicator of residents' health and well-being. To date, no study has assessed the probability of walking occurrence and frequency among PPR residents or investigated the factors influencing these behaviors using advanced analytical approaches. Therefore, this study aims to identify the key determinants of walking behavior among PPR residents using a Bayesian Network model, providing evidence-based insights to guide policymakers in enhancing walkability and public health outcomes in low-income urban communities.

2. Methods

2.1. Location and description of study site

The data of this study was collected from Community Housing Projects (PPRs) of Lembah Pantai, Kuala Lumpur, Malaysia. PPRs are low-cost, high-rise residential units initiated by the National Housing Department (Jabatan Perumahan Negara, JPN) in Kuala Lumpur since 1998. Four developments have been selected for data collection, including Kampung Limau, Sri Pantai, Sri Cempaka, and Pantai Ria which are located along the New Pantai Expressway (NPE) (Fig. 1). The authors of this study employed a face-to-face interview approach for collecting data. These developments mostly located within areas with poor walking facilities, including narrow sidewalks, uneven sidewalk surface, and disconnected sidewalks. The existing condition of the sidewalk network indicates that the neighbourhood cannot efficiently inspire walking.

A random sampling scheme was employed for interviewing the residents of four PPRs. Two conditions had been satisfied for each interviewee, first, the interviewees must be 18 years old or above, and they should be able to walk. Finally, 680 complete questionnaires were collected from the case study developments. It should be noted that household heads and married respondents were more likely to be available at home during the data collection period, leading to a slight overrepresentation of these groups in the final sample. This potential sampling bias has been acknowledged and considered in the interpretation of results.

The questionnaire used in this study included two main parts: (1) individual and household characteristics and (2) satisfaction of PPRs residents about their pedestrian environment. The initial section of the questionnaire gathered demographic information, such as

Table 1

Some recent studies on the application of machine learning techniques on the walking and pedestrian issues.

Author (year)	Study aim	ML technique(s) used
Kadali et al. (2014)	To investigate pedestrian gap acceptance behaviour, particularly in relation to how individuals decide whether to accept or reject vehicular gaps at mid-block crossings.	ANN
Navarro et al. (2017)	To detect pedestrians for autonomous vehicles utilising high-definition 3D range data.	KNN, NBC, and SVM.
Ding et al. (2018)	To explore different outcomes connected with pedestrian crashes.	MAPRT
Cohen and Dalyot (2020)	To examine machine-learning techniques for developing a computer-assisted model to predict pedestrian traffic flow levels.	ANN, SVM, SGD, AdaBoost, DT, and RF
Das et al. (2020)	To produce a structure for implementing machine-learning models to classify crash varieties from the unorganised textual content.	XGBoost, RF, and SVM
Effati and Saheli (2022)	To investigate high-risk pedestrian areas in densely populated rural districts, with a particular focus on roadside land use variables and the spatial distribution of access roads.	CART
Guo et al. (2021)	To classify older pedestrian traffic crashes into three levels of severity using crash data from Colorado, United States.	XGBoost

ANN = Artificial Neural Network; SVM = Support Vector Machine; RF = Random Forest; CART = Classification and Regression Tree; DT = Decision Tree; SGD = Stochastic Gradient Descent; MAPRT = Multiple Additive Poisson Regression Trees; KNN = K-Nearest Neighbours; NBC = Naïve Bayes classifier.



Fig. 1. A map of Peninsular Malaysia showing the locations of case study PPRs Source: Authors' elaboration based on [Google Maps \(2025\)](#).

participants' age, gender, ethnicity, religious background, educational qualifications, marital status, and employment status, the major daily trips' destination, usual travel mode to the major daily destination, and travel frequency to the major daily destination. In addition, the first part comprised some household attributes, including some items, including average monthly expenditure, monthly household income, count of household members, count of household members with the age of over 18, and family type. The second part of the questionnaire assessed the satisfaction level of PPRs residents toward their pedestrian environments. This assessment was conducted regarding the following issues: pathways, zebra crossings, street furniture and signage, personal safety, adjacent traffic fellow, aesthetics and amenities, sensory quality (sight/visual), sensory quality (Noise), sensory quality (Smell), sensory quality (feel), sensory quality (taste), and sensory quality (Mobility). Complete information concerning the variables used in this study is presented in [Table 1](#).

2.2. Statistical analysis

The objective of this study is predictive rather than causal. Specifically, we aim to estimate the probability of walking occurrence and frequency under specific combinations of socio-demographic and environmental conditions among PPR residents — rather than testing whether individual variables are statistically associated with walking. After data cleaning, 658 complete responses were used for the walking occurrence model, while 320 samples (representing respondents who walked at least once) were used for the frequency model.

It is very common to use traditional statistical methods, including the binary logit model, mixed logit model, and multinomial logit model for analysing the predictors of travel behaviour outcomes. However, regression models carry various robust statistical assumptions, including linearity in modelling associations, and the data is assumed to be outlier-free, which is hardly true for travel behaviour data ([Rashidi et al., 2014](#); [Stylianou et al., 2019](#)). Besides, handling variables with several classes is inefficient in regression models ([Karlaftis and Golias, 2002](#); [Yan et al., 2010](#)). In the present study, these limitations are particularly relevant. The dataset contains multiple categorical predictor variables with several response categories, such as marital status, job status, and satisfaction ratings on a 1–10 scale, which are inherently unsuitable for linear modelling. Furthermore, the research objective requires estimating probabilities under specific combinations of these variables simultaneously, a question that is structurally beyond what regression coefficients can answer.

This study therefore adopted the Bayesian Network (BN) technique, which allows the modelling of interrelationships among multiple predictors without assuming linearity or variable independence. Unlike regression methods, the BN framework explicitly represents probabilistic dependencies through a network structure, enabling a more flexible exploration of how socio-demographic and perceptual factors jointly influence walking behaviour. ML techniques, including BN, are assumption-free regarding variable distributions; therefore, prior probabilistic knowledge on travel behaviour is not required. Another advantage is their ability to efficiently handle outliers and variables with many classes (Ahmad et al., 2018; Breiman et al., 1984; Friedman et al., 1997; Khatami et al., 2017; Utkin et al., 2019). The method provides conditional probabilities that can be directly interpreted by planners and policymakers in designing targeted interventions.

Although the dataset size is moderate, BN is well suited for such analyses because it can effectively handle categorical variables and partial dependencies even with limited samples, provided model complexity is controlled. The Tree-Augmented Naïve Bayes (TAN) structure inherently limits complexity by restricting each predictor node to at most two parents — the class variable and one other attribute, which prevents the high variance that typically degrades more complex BN structures when sample sizes are moderate (Friedman et al., 1997). This is consistent with published applications of TAN models in transport and health behaviour research with comparable sample sizes (Aghaabbasi et al., 2020). In the present study, the number of inputs was further reduced from 91 candidate variables to 53 and 32 respectively through likelihood-ratio chi-square feature selection, ensuring model complexity remained proportionate to the available sample. A 10-fold cross-validation was then applied to mitigate overfitting and ensure generalizability.

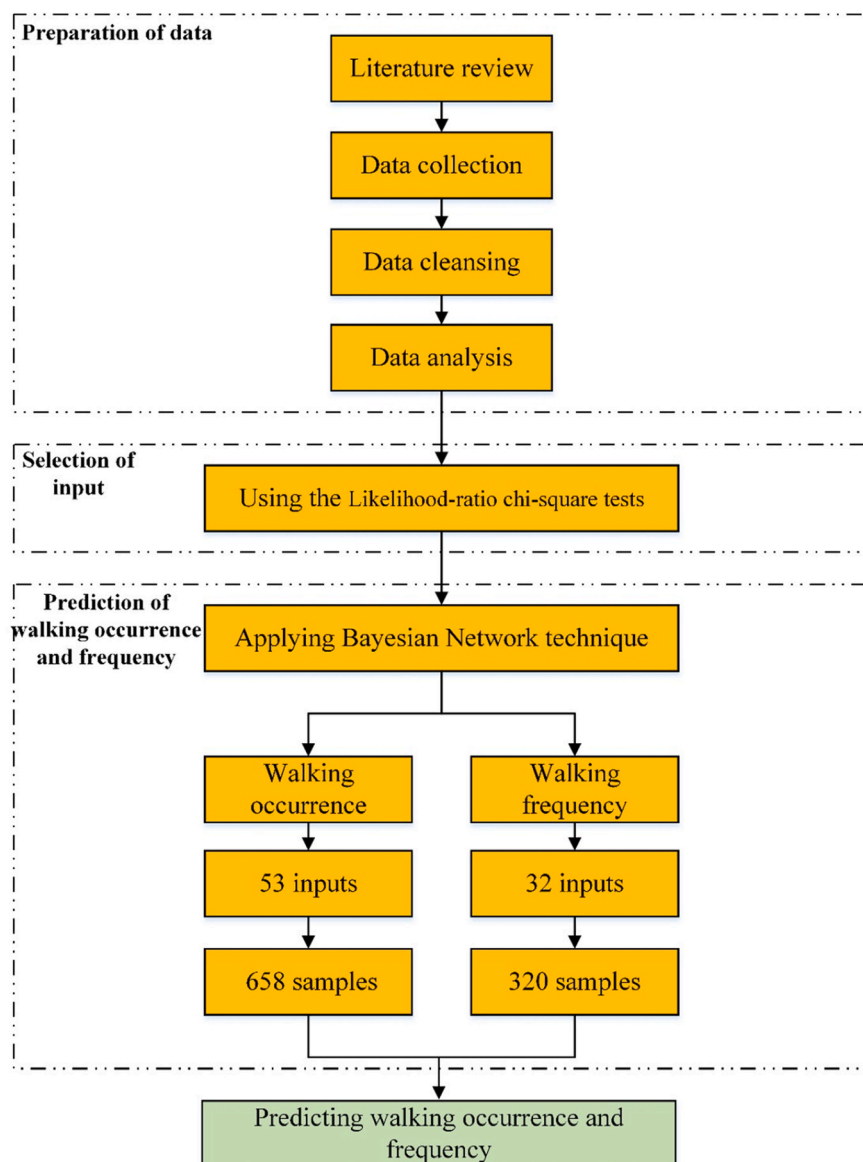


Fig. 2. Process of this study.

The Likelihood-ratio chi-square test was applied to identify the most influential predictors and reduce the number of input variables for the BN model by evaluating their dependency with the target variable. Fig. 2 displays the present study process.

The Bayesian Network framework is grounded in Bayesian theory, which integrates prior and joint probability distributions to estimate posterior probabilities for each variable. In simpler terms, the BN approach can be viewed as a graphical model that illustrates how various factors, such as demographics, employment status, and perceptions of walkability, are connected and jointly influence walking occurrence and frequency. This allows the model to capture both direct and indirect dependencies among variables, which traditional statistical models often fail to represent. In this study, a TAN model structure was adopted because it efficiently captures relationships between predictor variables while maintaining computational simplicity. The significance level for feature selection was set at 0.05, and the conditioning set size was determined based on model fit performance.

3. Results

A total sample of 658 households was involved in applying the ML model for walking occurrence and frequency prediction. To provide a clearer overview of the data used in the Bayesian Network (BN) analysis, the descriptive statistics of the key variables are summarised here. Among the respondents, the majority were married (74.3%) and employed in the private sector (31.8%). In terms of walking behaviour, 48.6% of respondents reported walking as their most frequent travel mode within the neighbourhood, while the weekly walking frequency ranged from 1 to 50 times ($N = 656$). For perceptual variables, the mean satisfaction scores were 5.51 for path condition (Q55.3), 5.41 for path connectivity (Q55.5), 3.93 for zebra crossing visibility (Q56.3; after excluding non-applicable codes 97 and 98), and 4.89 for traffic conditions at pedestrian crossings (Q59.2). These descriptive results provide the baseline profile of the respondents and the distribution of variables prior to the variable selection and BN model development (Table 2).

A Likelihood-ratio chi-square test was applied to select the most relevant inputs for walking occurrence and frequency in the case study neighbourhoods. Among 91 candidate inputs, this test selected 53 and 32 inputs as the most relevant ones for predicting the walking occurrence and walking frequency, respectively. These inputs belonged to both sociodemographic and perceptual groups. These selected inputs then were employed to develop the BN predictive models. For ease of interpretation, descriptive labels for all variable codes (e.g., Q24, Q26, Q55.3) used in the Bayesian Network models are provided in Appendix A (Codebook of Variables).

The authors of this study used several parameters and settings to develop the BN predictive models. The structure type of these models was the Tree Augmented Naïve Bayes model (TAN), parameter learning method was set as maximum likelihood, independent test was set as likelihood ratio, significance level was set as 0.01, and maximal conditioning set size was set as five.

The BN models identified the three best predictors of each walking occurrence and frequency models in the case study neighbourhoods. These inputs and their importance are shown in Fig. 3. For walking occurrence (Fig. 3a), the model identified the job status (Q26) as the most important predictor followed by marital status (Q24) of the participants. The level of satisfaction regarding the traffic conditions at pedestrian crossings (Q59.2) was identified as the least important input. Regarding the walking frequency (Fig. 3b), the second BN model identified the satisfaction towards visibility of zebra crossing (Q56.3) as the most important predictor followed by the path condition (Q55.3) and the path connectivity (Q55.5) which achieved the same importance value. These results implied that the sociodemographic characteristics of participants were more influential than their perception of their physical environment on their decision to take a walk. On the other hand, participants' perceptions towards their physical environment were dominant factors for determining the frequency of walking. The results of the walking frequency model also indicated that the sociodemographic characteristics may not be suitable predictors of walking frequency.

The BN model consists of a directed acyclic graph structured by nodes and connected edges, along with a conditional probability table indicating the likelihood of each node based on the states of its parent nodes. Fig. 4 displays the resulting network diagrams. These diagrams reveal the association between each target variable and its input variables. As previously mentioned, the authors applied the BN technique to predict the walking occurrence (Fig. 4a) and walking frequency (Fig. 4b).

The first model which predicts the walking occurrence was applied to a dataset with 658 samples. In this model, every predictor (namely the job status, marital status, and the level of satisfaction regarding the traffic conditions at pedestrian crossings) possesses the target variable (namely walking occurrence) functioning as a parent and can possess an extra predictor as a parent. The network of the

Table 2
Descriptive statistics of variables used in the Bayesian Network models.

Variable code	Variable description	N	Min	Max	Mean	SD
Q55.3	Satisfaction with path condition	654	1	10	5.51	1.97
Q55.5	Satisfaction with path connectivity	656	1	10	5.41	1.98
Q56.3	Satisfaction with zebra crossing visibility ^a	333	1	10	3.93	2.65
Q59.2	Satisfaction with traffic conditions at crossings	656	1	10	4.89	2.04
walk_freq	Weekly walking frequency (times per week) ^b	656	1	50	–	–
walk_occ	Walking as primary mode (1 = yes, 0 = no) ^c	658	–	–	–	–
Q24	Marital status (categorical)	657	–	–	–	–
Q26	Job status/employment category (categorical)	658	–	–	–	–

Notes.

^a non-applicable values (97, 98) were treated as missing before analysis.

^b Walking frequency is reported using frequency distribution in the text.

^c Binary/categorical variables are presented in frequency tables rather than mean/SD.

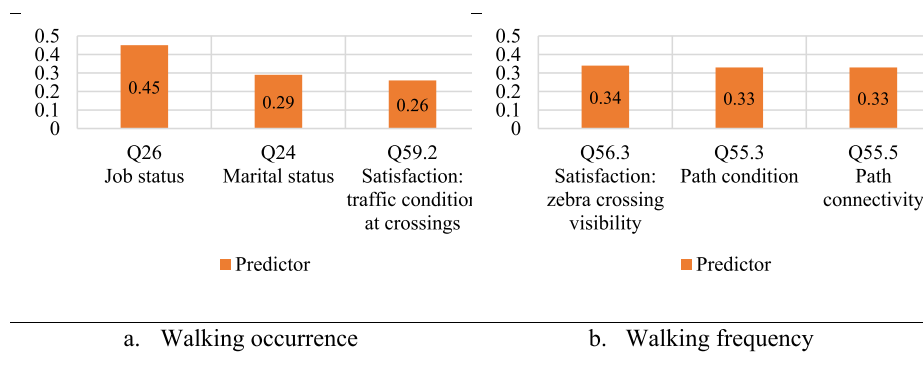


Fig. 3. Importance of walking occurrence predictors.

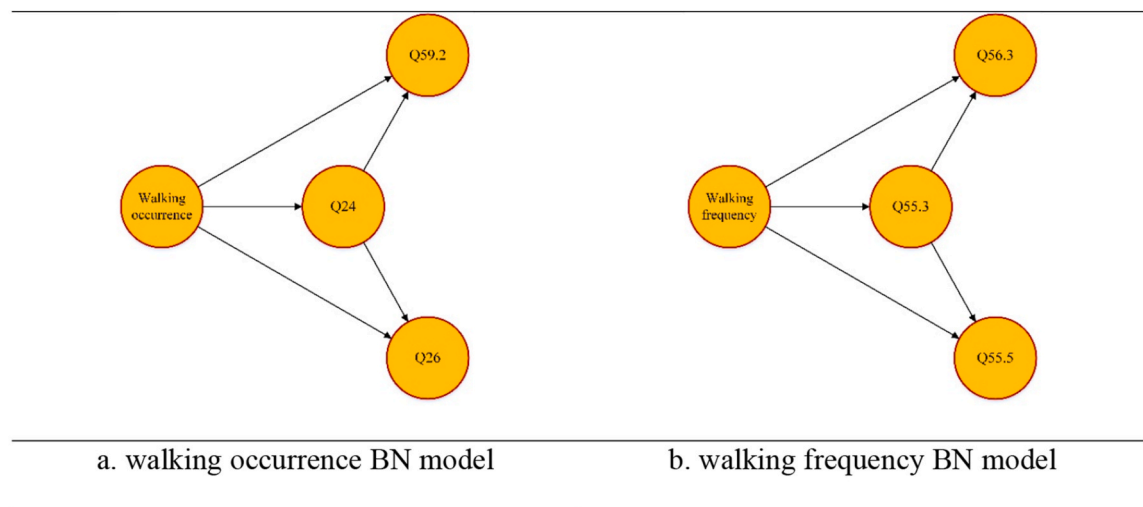


Fig. 4. The BN model for predicting walking occurrence and frequency.

first BN model includes four nodes, three predictors and one target variable. Additionally, this network depicts the association between the predictors. As mentioned previously, the BN model produces a conditional probability table (CPT) for each connected node. It derives the joint probability distribution by combining the conditional probabilities of all nodes, based on the states of their respective parent nodes.

Table 3 presents the conditional probabilities associated with each category of participants' marital status in relation to walking occurrence. The conditional probabilities of participants' marital status suggest that the occurrence of walking was more probable than the non-occurrence of walking for the married and divorced participants. Alternatively, the non-occurrence of walking was more likely than the occurrence of walking for the widows/widowers and singles.

Table 4 indicates the conditional probabilities of participants' job status, taking into consideration the influence of the participants' marital status. It was highly possible for walking occurrence if the participant was a widow/widower and was a housewife (1.00). Instead, if the participant was a divorcee and was self-employed the likelihood of not taking a walk (0.75) was more than other conditions.

Table 5 displays the conditional probabilities of participants' satisfaction regarding the traffic conditions at pedestrian crossings considering the influence of the participants' marital status. It was profoundly possible for not occurring of walking if the participant

Table 3
Participants' marital status (Q24)/walking occurrence conditional probabilities.

Walking occurrence	Probability of Q24			
	Single	Married	Divorcee	Widow/Widower
Yes	0.12	0.79	0.08	0.01
No	0.28	0.70	0.01	0.02

Table 4

Participants' job status (Q26)/participants' marital status (Q24) conditional probabilities.

Parents		Probability of Q26								
Marital status	Walking occurrence	1	2	3	4	5	6	7	8	9
Single	Yes	0.03	0.47	0.16	0.03	0.05	0.00	0.11	0.03	0.13
Single	No	0.04	0.46	0.18	0.00	0.00	0.01	0.23	0.00	0.08
Married	Yes	0.08	0.19	0.20	0.06	0.01	0.00	0.00	0.34	0.11
Married	No	0.11	0.39	0.21	0.05	0.05	0.00	0.00	0.16	0.02
Divorce	Yes	0.00	0.20	0.20	0.04	0.00	0.00	0.00	0.20	0.36
Divorce	No	0.00	0.00	0.75	0.00	0.00	0.00	0.00	0.00	0.25
Widow/Widower	Yes	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00
Widow/Widower	No	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.50	0.17

was a divorcee and was greatly satisfied with the traffic conditions at pedestrian crossings (0.75). Instead, if the participant was a widow/widower, the likelihood of walking (0.33) regardless of his/her perception towards the traffic conditions at pedestrian crossings.

The second model which predicts the walking frequency was applied to a dataset with 320 samples. The second model was applied only to those samples that reported the use of walking for their major destinations. In the second BN model, each predictor (namely satisfaction towards visibility of zebra crossing, path condition, and path connectivity) possesses the target variable (namely walking frequency). The network of the second BN model includes four nodes, three predictors and one target variable. Similarly to the first model, the BN model generates some conditional probability tables for all connected nodes.

Table 6 shows the conditional probability for every value of participants' satisfaction regarding the path condition across all walking frequencies. The regular walking was more likely among those participants that were satisfied about the path condition. The frequent walking was more probable among those participants that had a medium satisfaction regarding the path condition. Those participants that had a low level of satisfaction regarding the path condition fell under the infrequent walkers.

Table B1 indicates the conditional probabilities of participants' satisfaction regarding the path connectivity considering the influence of participants' satisfaction regarding the path condition. The infrequent walking was greatly probable (1.00) if a participant had low satisfaction with the path connectivity and path condition. The highest probability for frequent walking was achieved (0.80) when participants had medium satisfaction regarding the path connectivity and path condition. In addition, if participants had great satisfaction regarding the path connectivity and path condition, a perfect likelihood (1.00) for regular walking was achieved.

Table B2 displays the conditional probabilities of participants' satisfaction towards visibility of zebra crossing considering the influence of participants' satisfaction regarding the path condition. The infrequent walking was greatly probable (1.00) if a participant had low satisfaction with the visibility of zebra crossing and path condition. If the participants had a high perception towards the path condition but had low perception towards the visibility of zebra crossing, then the frequent walking was highly probable (1.00). Furthermore, regular walkers had a high satisfaction level regarding both visibility of zebra crossing and path condition.

4. Discussion

This study primarily intended to identify the most influential factors on walking occurrence and frequency among residents of a community housing project in the metropolis of Kuala Lumpur, Malaysia. In addition, this study modelled the associations among these factors using the Bayesian Network technique. The first BN model showed that participants' marital status, job status, and satisfaction regarding the traffic conditions at pedestrian crossings were the most important factors to the walking occurrence. The second BN model indicated that participants' satisfaction towards visibility of zebra crossing, path condition and path connectivity were the most influential factors to the walking frequency. These findings emphasize the complex and interrelated nature of socio-demographic and perceptual influences on active travel behaviour in low-income urban populations.

Consistent with earlier studies, marital status and job status were two of the most important factors of the travel mode choice and frequency (Alfonzo, 2005; Buehler, 2012; Jiao and Azimian, 2021; Kim et al., 2015; Lundberg and Weber, 2014; Senbil et al., 2009). Concerning the walking occurrence model, the probabilities of the occurrence and non-occurrence of walking were higher for those

Table 5

Satisfaction regarding the traffic conditions at pedestrian crossings (Q59.2)/participants' marital status (Q24) conditional probabilities.

Parents		Probability of Q59.2									
Marital status	Walking occurrence	1	2	3	4	5	6	7	8	9	10
Single	Yes	0.05	0.03	0.21	0.08	0.16	0.24	0.11	0.08	0.03	0.03
Single	No	0.01	0.06	0.10	0.12	0.17	0.17	0.25	0.10	0.01	0.01
Married	Yes	0.11	0.10	0.10	0.18	0.17	0.23	0.07	0.03	0.01	0.00
Married	No	0.04	0.09	0.08	0.09	0.20	0.17	0.23	0.08	0.00	0.01
Divorce	Yes	0.20	0.20	0.08	0.12	0.04	0.16	0.16	0.00	0.04	0.00
Divorce	No	0.00	0.25	0.00	0.00	0.00	0.00	0.75	0.00	0.00	0.00
Widow/Widower	Yes	0.00	0.00	0.00	0.33	0.00	0.33	0.00	0.33	0.00	0.00
Widow/Widower	No	0.00	0.00	0.17	0.00	0.33	0.17	0.17	0.17	0.00	0.00

Table 6

Participants' satisfaction regarding the path condition (Q55.3)/walking frequency conditional probabilities.

Walking frequency	Probability of Q55.3									
	1	2	3	4	5	6	7	8	9	10
Infrequent	0.07	0.27	0.33	0.07	0.13	0.07	0.00	0.07	0.00	0.00
Frequent	0.09	0.01	0.06	0.03	0.24	0.32	0.16	0.05	0.03	0.01
Regular	0.01	0.07	0.08	0.03	0.09	0.04	0.12	0.20	0.25	0.12

participants that were married; compared to those who were single, divorced or widowed/widower. This indicates that marital status is the most influential for the walking occurrence. An important reason for this result is that the majority of our survey participants were among the households' heads which increased the likelihood of over-presenting the married participants (74.32%). Alternatively, divorcees and widows/widowers were under-represented in our dataset, and it may be a feasible answer for little chances of walking for these marital groups. This sampling imbalance should be acknowledged as a possible source of bias; however, the strong association observed suggests genuine behavioural differences related to household and caregiving responsibilities.

The walking occurrence model identified two associations among the predictors: (1) association between marital status and job status and (2) association between marital status and satisfaction regarding the traffic conditions at pedestrian crossings. For our information, these associations were not investigated by the previous studies. The former association resulted in the occurrence of walking with a perfect probability (1.00) if the participants were widows/widowers and housewife. The age range of widows/widowers that were also housewife was between 61 and 76. Typically, in this age range, the people have less responsibility regarding their families and housework. Besides, people may care more about their health in this age range and allocate more time to walk. Instead, if the people were divorcee and self-employed, the probability of non-occurrence of walking was more than its occurrence. People who divorced and were self-employed were mostly in the age range of 33-65. Thus, they are of working age and they were busy with their businesses. In addition, these people normally had a good income level (up to 3000 RM) and could afford a private vehicle. The third reason can be that these people belonged to nuclear and extended families. Thus, the household's heads had much responsibility for other family members, especially for taking children or spouse to the school or office. This finding highlights the social equity dimension of walking behaviour. For example, household structure and caregiving roles can directly limit walking opportunities, reinforcing transport inequities within lower-income communities.

If a participant was divorced and highly satisfied with traffic conditions at pedestrian crossings the probability of non-occurrence of walking was high. The divorced participants belonged to a wide age range (29-80). Also, the majority of them belonged to the nuclear family type and their household count was between two and seven. Thus, this group of participants had a lot of life responsibilities and even with great perception about the traffic conditions at pedestrian crossings, still, the probability of walking among this group is less than the probability of walking. This suggests that improving traffic conditions alone may not be sufficient to increase walking among socioeconomically constrained or time-poor populations; broader livelihood and accessibility interventions may be required.

Regarding the second association, if the participants had great satisfaction regarding the path condition and path connectivity, then the probability of regular walking was perfect (1.00). The proposed BN model also showed that these two associated factors are essential to predict the walking frequency. For our information, the association recognized here was not stated in the literature. The findings of this model imply that to boost the probability of regular walking, people should have a positive attitude towards their neighbourhoods' path conditions and connectivity. The walking frequency model also identified the association of participants' satisfaction towards visibility of zebra crossing and participants' satisfaction regarding the path condition for walking frequency prediction. Similarly to the previous association stated above, the regular walkers had a great satisfaction level regarding both visibility of zebra crossing and path condition. Regarding the association of perceptual factors, it can be said that regular walking can be promoted among people if their perceptions regarding the path condition, path connectivity, and the visibility of zebra crossing are improved. These results are consistent with the principles of the WHO Global Action Plan on Physical Activity (2018-2030), which highlights safe and connected pedestrian infrastructure as essential for promoting physical activity and reducing health inequities.

As mentioned earlier, this present study identified some perceptual factors as the most important ones for predicting the walking occurrence and frequency. Several previous studies confirmed that perception of people regarding their physical environment is important for predicting and analysing travel mode choice and frequency (e.g., [Aghaabbasi et al., 2020](#); [Herrmann-Lunecke et al., 2021](#); [Rišová and Madajová, 2020](#)). In walking occurrence and frequency models, participants' satisfaction towards the traffic conditions at pedestrian crossings, visibility of zebra crossing, path condition and path connectivity were the most influential factors. It was confirmed by [Anciaes et al. \(2019\)](#) that people perceived heavy traffic and fast traffic speed as a barrier to walking in the local areas. In addition, the majority of our case study neighbourhoods were located near busy roads which may be associated with different levels of walking ([Giles-Corti and Donovan, 2002](#)). In other words, the need to pass busy streets may discourage walking. This insight reinforces the importance of designing pedestrian-friendly urban spaces as part of sustainable development goals (SDG 3 on Good Health and Well-being and SDG 11 on Sustainable Cities and Communities).

Previous studies also confirmed that path conditions including its maintenance, surface and the materials were assumed by pedestrians as an important factor of safety (e.g., [Aghaabbasi et al., 2018](#)). Thus, it makes sense that people perception towards the factors that influence their safety from traffic and falling influence their frequency of walking. In addition, people with different perceptions regarding the path condition may have different levels of physical activity, particularly walking. In several studies, it was shown that a positive perception regarding the path condition can encourage people to take physical activity and walk more frequently ([Cerin et al., 2013](#)). Regarding the path connectivity, it always has been recognized as one of the most prominent and supportive

factors to encourage walking (Boulange et al., 2017; Cervero and Kockelman, 1997; King et al., 2015; McCormack et al., 2006). Thus, it is not surprising that the satisfaction of the participants of our study became important for predicting the walking frequency. These findings collectively underline that improving micro-scale infrastructure, like sidewalk continuity, surface quality, and crosswalk visibility, can generate measurable public health co-benefits.

Based on the findings of this study, we proposed a framework according to the importance of predictors for encouraging people to select walking as the principal mode of travel and repeat this usage regularly (Fig. 5). Our analyses of walking occurrence suggest that policy interventions aimed at addressing the well-established links between sociodemographic and perception of people regarding their physical environment and the walking occurrence of residents living in PRPs housing projects should aim to increase both the satisfaction level of residents about their living environments and residents' opportunities to walk. Enhancing both aspects can be accomplished through a mix of policy interventions—for instance, implementing economic or regulatory strategies to discourage private car usage while promoting alternative transport modes to reduce traffic volume. Simultaneously, measures such as speed limits and traffic calming techniques can be applied to lower vehicle speeds. Moreover, enhancing the quality of the paths or increasing signalized crossing facilities may prevent the negative attitudes of street traffic conditions. It is also possible the government consider the “neighbourhood residential street” for the future PRPs projects which has been implemented in South Korea successfully (Jung et al., 2015). The implementation of these streets in apartment complexes may encourage people to walk more frequently, raise the interactions among residents of the neighbourhood, and in turn, improve the neighbourhood environment quality. Such integrated design and policy approaches can help achieve sustainable urban mobility and public health outcomes in alignment with both local planning goals and international health frameworks.

Regarding the opportunities for walking, the government should help the adult residents to care more about themselves. Those who work for public and private sectors should possess a more flexible work schedule. An example of this flexibility can be the right of choosing work starting/finishing time. In addition, the government can provide more kindergartens in the PRPs housing projects which can help the parents to allocate sufficient time for walking or other physical activities. The government also can provide schools near these housing units which can benefit both parents and children and encourage them to walk to school. These recommendations echo the global call for “Health in All Policies” by WHO, recognizing that transport, housing, and labor policies together shape daily opportunities for physical activity.

There are some limitations to our investigation. Because of the sample adoption approach that takes mostly heads of household, the sample has a greater possibility to over-represent married participants. This circumstance demands care when interpreting the overall walking occurrence and frequency from this dataset. Nevertheless, it is expected that a large sample size mitigates potential biases arising from sample imbalance, as even the subsamples representing under-represented groups may be sufficient to yield statistically significant results. Participants of this study did not represent all PPRs residents across Malaysia. This present investigation employed the self-report data of walking occurrence and frequency which is another limiting point. Future studies can supplement self-report with observations of trips. The perception of PPRs' residents regarding their health condition is not considered in this study; thus, future studies can include this factor in their analysis. Other non-motorized travels such as cycling were not examined in this study. Further studies can also predict the probability of cycling occurrence and frequency using the BN methods within the PPRs' projects. Ultimately, this study was conducted in the context of a developing country characterized by a high rate of vehicle ownership. Furthermore, the overall quality of public transit and active transportation infrastructure is suboptimal. Therefore, applying the findings of this study to developed countries should be approached with caution. Despite these limitations, the findings provide actionable insights for creating equitable, health-promoting, and sustainable transport systems for low-income urban populations.

5. Conclusion

This study aimed to identify and model the key factors influencing walking occurrence and frequency among residents of low-income community housing projects (PPRs) in Kuala Lumpur, Malaysia. Using the Likelihood-ratio chi-square tests results, 53 and 32 inputs were identified as the most important predictors of the walking occurrence and walking frequency, respectively. Among these, job status, marital status, and satisfaction with traffic conditions at pedestrian crossings were the strongest determinants of walking occurrence, while satisfaction with zebra crossing visibility, path condition, and path connectivity most strongly influenced walking frequency.

Although the sample size was moderate, it was sufficient for the Bayesian Network analysis because the model complexity was carefully controlled, and predictor selection was performed through the likelihood-ratio chi-square test. The BN approach was chosen not merely for its novelty but because it effectively captures interdependencies among socio-demographic and perceptual variables, relationships that conventional regression models may oversimplify or overlook. The cross-validation process further ensured that the findings were robust and generalizable.

This study represents one of the few efforts globally, and the first within Malaysia's PPR context, to employ a probabilistic machine learning framework to predict walking behavior in low-income urban communities. Three specific findings emerge that, to the best of the authors' knowledge, have not been previously reported in the literature.

First, the study reveals a clear divergence between the predictors of walking occurrence and walking frequency. The decision of whether to walk at all is governed primarily by socio-demographic factors—specifically job status and marital status—while how often a person walks is governed entirely by perceptual factors related to the pedestrian environment, namely satisfaction with zebra crossing visibility, path condition, and path connectivity. This two-domain split has not been previously identified in the Malaysian context or among PPR residents, and it carries a direct policy implication: interventions targeting walking uptake should address social and livelihood conditions, while those targeting walking frequency should focus on the physical quality of pedestrian infrastructure.

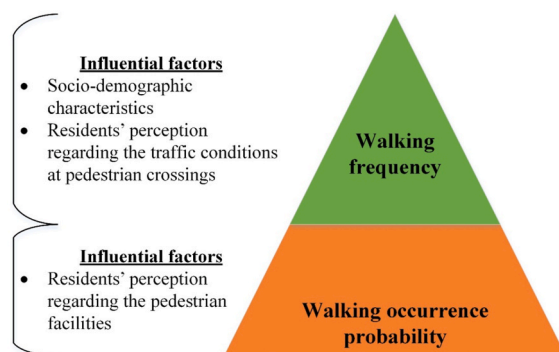


Fig. 5. Framework according to the importance of variables for encouraging people to select walking as the principal mode of travel and repeat this usage regularly.

Second, the Bayesian Network model uncovered conditional probability thresholds that regression-based methods cannot produce. Specifically, when residents reported high satisfaction with both path condition and path connectivity simultaneously, the probability of regular walking reached 1.00. This threshold effect—where two perceptual factors must jointly reach a satisfactory level to promote habitual walking—is invisible to conventional regression models, which produce independent coefficients for each predictor rather than joint conditional outcomes. This finding suggests that improving one factor in isolation may be insufficient; both path condition and connectivity must be addressed together to achieve meaningful increases in regular walking.

Third, the interaction between marital status and job status produced striking scenario-level predictions unique to this community. Widow/widower residents who were housewives had a perfect probability (1.00) of walking, reflecting the combination of low time pressure, reduced family caregiving responsibilities, and health-oriented behaviour in older age. In contrast, divorced self-employed residents had a 0.75 probability of not walking, reflecting competing demands of livelihood, family responsibilities, and private vehicle access. These contrasting profiles, shaped by the joint influence of household structure and employment type, could not have been identified through individual predictor coefficients alone.

Practically, this study demonstrates that data-driven, probabilistic modeling can support evidence-based transport and health planning. The results can guide urban planners and policymakers in designing equitable, health-promoting, and sustainable environments for low-income residents. Targeted interventions—such as improving path connectivity and surface quality, enhancing visibility and safety at crosswalks, and promoting flexible working arrangements—can reduce walking barriers and foster more active, inclusive mobility patterns.

Despite its contributions, this study has limitations. The dataset was cross-sectional and collected from a limited number of PPR sites in Kuala Lumpur; thus, generalizations to other contexts should be made cautiously. The reliance on self-reported walking frequency and perceptions may also introduce response bias. Future research could expand the sample, incorporate objective measures of walking behavior, and explore other non-motorized modes such as cycling using similar probabilistic frameworks.

Overall, this research contributes both methodologically and empirically to the growing body of knowledge on active mobility in developing urban contexts. It underscores that improving pedestrian infrastructure and residents' perceptions of safety, comfort, and accessibility can significantly encourage walking as a daily mode of transport—advancing progress toward public health and sustainability goals. These insights reinforce the need for integrated transport, housing, and health policies that prioritize walkability as a foundation for equitable urban development.

CRediT authorship contribution statement

Rosilawati Zainol: Writing – review & editing, Visualization, Validation, Software, Resources, Project administration, Methodology, Formal analysis, Conceptualization. **Tin Tin Su:** Funding acquisition, Conceptualization. **Hazreen Abdul Majid:** Funding acquisition, Conceptualization. **Cheong Peng Au Yong:** Writing – review & editing, Visualization. **Mahdi Aghaabbasi:** Writing – review & editing, Writing – original draft, Visualization, Validation, Formal analysis. **Nithiah Thangiah:** Methodology, Data curation, Conceptualization. **Dan He:** Writing – review & editing, Visualization. **Fadly Usman:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors thank the authorities from City Hall Kuala Lumpur, Community Housing Management Committee, participants of the study, and the “PARTNER” research team. This research was supported by University of Malaya, Flagship Project FL009- 2011.

Appendices.

Appendix A. Codebook of Variables Used in the Results Section

Code	Variable Description
Q24	Marital status of the respondent (single, married, divorcee, widow/widower)
Q26	Job status/employment category of the respondent
Q55.3	Satisfaction with path condition (surface quality, smoothness, maintenance)
Q55.5	Satisfaction with path connectivity (continuity, network links, access)
Q56.3	Satisfaction with visibility of zebra crossings
Q59.2	Satisfaction with traffic conditions at pedestrian crossings

Appendix B. Conditional Probability Tables

Table B1

Participants' satisfaction regarding the path connectivity (Q55.5)/Participants' satisfaction regarding the path condition (Q55.3) conditional probabilities

Parents		Probability Q55.5									
Walking frequency	Q55.3	1	2	3	4	5	6	7	8	9	10
Infrequent	1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Infrequent	2	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Infrequent	4	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Infrequent	5	0.00	0.00	0.00	0.00	0.50	0.50	0.00	0.00	0.00	0.00
Infrequent	6	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Infrequent	7	0.20	0.00	0.00	0.00	0.00	0.00	0.60	0.00	0.00	0.20
Infrequent	8	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
Frequent	1	0.71	0.00	0.00	0.00	0.00	0.14	0.14	0.00	0.00	0.00
Frequent	3	0.00	0.20	0.60	0.00	0.20	0.00	0.00	0.00	0.00	0.00
Frequent	4	0.00	0.00	0.50	0.50	0.00	0.00	0.00	0.00	0.00	0.00
Frequent	5	0.00	0.11	0.00	0.11	0.53	0.26	0.00	0.00	0.00	0.00
Frequent	6	0.00	0.00	0.04	0.00	0.12	0.80	0.04	0.00	0.00	0.00
Frequent	7	0.00	0.00	0.00	0.08	0.08	0.00	0.69	0.15	0.00	0.00
Frequent	8	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.50	0.00	0.00
Frequent	9	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.50	0.00
Frequent	10	0.14	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00	0.71
Regular	1	0.67	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Regular	2	0.06	0.62	0.12	0.06	0.06	0.00	0.06	0.00	0.00	0.00
Regular	3	0.06	0.18	0.41	0.12	0.12	0.06	0.06	0.00	0.00	0.00
Regular	4	0.00	0.04	0.12	0.42	0.19	0.12	0.04	0.04	0.00	0.04
Regular	5	0.00	0.02	0.02	0.07	0.80	0.09	0.00	0.00	0.00	0.00
Regular	6	0.00	0.00	0.04	0.02	0.16	0.75	0.04	0.00	0.00	0.00
Regular	7	0.00	0.00	0.00	0.04	0.14	0.25	0.36	0.21	0.00	0.00
Regular	8	0.00	0.00	0.05	0.00	0.10	0.00	0.30	0.45	0.05	0.05
Regular	9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.62	0.12
Regular	10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Table B2

Participants' satisfaction towards visibility of zebra crossing (Q56.3)/Participants' satisfaction regarding the path condition (Q55.3) conditional probabilities

Parents		Probability Q56.3									
Walking frequency	Q55.3	1	2	3	4	5	6	7	8	9	10
Infrequent	1	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Infrequent	2	0.25	0.75	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Infrequent	4	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Infrequent	5	0.00	0.50	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00
Frequent	1	0.71	0.14	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00
Frequent	3	0.40	0.20	0.20	0.00	0.20	0.00	0.00	0.00	0.00	0.00
Frequent	4	0.50	0.00	0.00	0.00	0.00	0.50	0.00	0.00	0.00	0.00
Frequent	5	0.58	0.00	0.00	0.05	0.21	0.16	0.00	0.00	0.00	0.00
Frequent	6	0.84	0.00	0.04	0.00	0.08	0.04	0.00	0.00	0.00	0.00
Frequent	7	0.77	0.08	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00
Frequent	8	0.25	0.00	0.00	0.00	0.25	0.25	0.25	0.00	0.00	0.00
Frequent	9	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00	0.00

(continued on next page)

Table B2 (continued)

Parents		Probability Q56.3									
Frequent	10	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Regular	5	0.04	0.22	0.09	0.07	0.02	0.56	0.00	0.00	0.00	0.00
Regular	6	0.00	0.00	0.07	0.04	0.02	0.00	0.04	0.00	0.00	0.84
Regular	7	0.00	0.04	0.00	0.04	0.00	0.14	0.07	0.11	0.61	0.00
Regular	8	0.05	0.00	0.10	0.10	0.00	0.10	0.05	0.10	0.50	0.00
Regular	9	0.12	0.00	0.00	0.00	0.00	0.00	0.12	0.12	0.25	0.38
Regular	10	0.14	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.71

Data availability

Data will be made available on request.

References

- Abusaada, H., Elshater, A., 2024. Revisiting urban street planning and design factors to promote walking as a physical activity for middle-class individuals with metabolic syndrome in Cairo, Egypt. *Int. J. Environ. Res. Publ. Health* 21 (4), 402.
- Aghaabbasi, M., Moeinaddini, M., Shah, M.Z., Asadi-Shekari, Z., 2017a. A new assessment model to evaluate the microscale sidewalk design factors at the neighbourhood level. *J. Transport Health* 5 (1), 97–112. <https://doi.org/10.1016/j.jth.2016.08.012>.
- Aghaabbasi, M., Moeinaddini, M., Zaly Shah, M., Asadi-Shekari, Z., 2017b. A new assessment model to evaluate the microscale sidewalk design factors at the neighbourhood level. *J. Transport Health* 5 (1), 97–112. <https://doi.org/10.1016/j.jth.2016.08.012>.
- Aghaabbasi, M., Moeinaddini, M., Asadi-Shekari, Z., Zaly Shah, M., 2018. The equitable use concept in sidewalk design. *Cities*. <https://doi.org/10.1016/j.cities.2018.10.010>.
- Aghaabbasi, M., Shekari, Z.A., Shah, M.Z., Olakunle, O., Armaghani, D.J., Moeinaddini, M., 2020. Predicting the use frequency of ride-sourcing by off-campus university students through random forest and Bayesian network techniques. *Transport. Res. Pol. Pract.* 136, 262–281.
- Ahmad, I., Bashari, M., Iqbal, M.J., Rahim, A., 2018. Performance comparison of support vector machine, random forest, and extreme learning machine for intrusion detection. *IEEE Access* 6, 33789–33795.
- Alfonzo, M.A., 2005. To walk or not to walk? The hierarchy of walking needs. *Environ. Behav.* 37 (6), 808–836.
- Anciaes, P.R., Stockton, J., Ortegón, A., Scholes, S., 2019. Perceptions of road traffic conditions along with their reported impacts on walking are associated with wellbeing. *Travel behaviour and society* 15, 88–101.
- Bernetti, G., Longo, G., Tomasella, L., Violin, A., 2008. Sociodemographic groups and mode choice in a middle-sized European city. *Transp. Res. Rec.: J. Transport. Res. Board* 2067 (1), 17–25. <https://doi.org/10.3141/2067-03>.
- Boulange, C., Gunn, L., Giles-Corti, B., Mavoa, S., Pettit, C., Badland, H., 2017. Examining associations between urban design attributes and transport mode choice for walking, cycling, public transport and private motor vehicle trips. *J. Transport Health* 6, 155–166.
- Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and regression trees. *Wadsworth Int. Group* 37 (15), 237–251.
- Buehler, R., 2012. Determinants of bicycle commuting in the Washington, DC region: the role of bicycle parking, cyclist showers, and free car parking at work. *Transport. Res. Transport Environ.* 17 (7), 525–531.
- Cerin, E., Lee, K.-y., Barnett, A., Sit, C.H., Cheung, M.-c., Chan, W.-m., 2013. Objectively-measured neighborhood environments and leisure-time physical activity in Chinese urban elders. *Prev. Med.* 56 (1), 86–89.
- Cervero, R., Kockelman, K., 1997. Travel demand and the 3Ds: density, diversity, and design. *Transp. Res., Part D Transp. Environ.* 2 (3), 199–219.
- Cheng, L., Chen, X., Yang, S., Wang, H., Wu, J., 2016. Modeling mode choice of low-income commuters with sociodemographics, activity attributes, and latent attitudinal variables: case study in Fushun, China. *Transp. Res. Rec.: J. Transport. Res. Board* 2581 (1), 27–36. <https://doi.org/10.3141/2581-04>.
- Cohen, A., Dalyot, S., 2020. Machine-learning prediction models for pedestrian traffic flow levels: towards optimizing walking routes for blind pedestrians. *Trans. GIS* 24 (5), 1264–1279.
- Das, S., Le, M., Dai, B., 2020. Application of machine learning tools in classifying pedestrian crash types: a case study. *Transportation Safety and Environment* 2 (2), 106–119. <https://doi.org/10.1093/tse/tdaa010>.
- De Cambra, P.J.M., 2012. *Pedestrian Accessibility and Attractiveness Indicators for Walkability Assessment* Thesis for the Master Degree (Msc) in Urban Studies and Territorial Management].
- Ding, C., Chen, P., Jiao, J., 2018. Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: a machine learning approach. *Accid. Anal. Prev.* 112, 116–126.
- Effati, M., Saheli, M.V., 2022. Examining the influence of rural land uses and accessibility-related factors to estimate pedestrian safety: the use of GIS and machine learning techniques. *Int. J. Transp. Sci. Technol.* 11 (1), 144–157.
- Friedman, N., Geiger, D., Goldszmidt, M., 1997. Bayesian network classifiers. *Mach. Learn.* 29 (2), 131–163. <https://doi.org/10.1023/a:1007465528199> [journal article].
- Giles-Corti, B., Donovan, R.J., 2002. The relative influence of individual, social and physical environment determinants of physical activity. *Soc. Sci. Med.* 54 (12), 1793–1812.
- Google Maps, 2025. Map of Peninsular Malaysia showing the locations of case study PPRs. Google Maps. Available at: <https://www.google.com/maps>.
- Guo, Y., Chan, C.H., Chang, Q., Liu, T., Yip, P.S., 2019. Neighborhood environment and cognitive function in older adults: a multilevel analysis in Hong Kong. *Health Place* 58, 102146.
- Guo, M., Yuan, Z., Janson, B., Peng, Y., Yang, Y., Wang, W., 2021. Older pedestrian traffic crashes severity analysis based on an emerging machine learning XGBoost. *Sustainability* 13 (2), 926. <https://www.mdpi.com/2071-1050/13/2/926>.
- Hayat Khan, Tareef, Danladi Isah, Abubakar, Anjomshoa, Elahe, Sabri, S., 2015. Users' Perceptions on Pedestrian Prioritized Neighborhoods: a Study on Terrace Row Housing Estates in Malaysia.
- Herrmann-Lunecke, M.G., Mora, R., Vejares, P., 2021. Perception of the built environment and walking in pericentral neighbourhoods in Santiago, Chile. *Travel Behaviour and Society* 23, 192–206.
- Jiao, J., Azimian, A., 2021. Exploring the factors affecting travel behaviors during the second phase of the COVID-19 pandemic in the United States. *Transportation Letters* 13 (5–6), 331–343. <https://doi.org/10.1080/19427867.2021.1904736>.
- Jung, E., Lee, J., Kim, K., 2015. The relationship between pedestrian environments and sense of community in apartment complexes in Seoul, Korea. *J. Asian Architect. Build Eng.* 14 (2), 411–418.
- Kadali, B.R., Rathi, N., Perumal, V., 2014. Evaluation of pedestrian mid-block road crossing behavior using an Artificial Neural Network (ANN). In: *CICTP 2014*, pp. 1911–1922. <https://doi.org/10.1061/9780784413623.184>.

- Karlaftis, M.G., Golias, I., 2002. Effects of road geometry and traffic volumes on rural roadway accident rates. *Accid. Anal. Prev.* 34 (3), 357–365.
- Khatami, A., Khosravi, A., Nguyen, T., Lim, C.P., Nahavandi, S., 2017. Medical image analysis using wavelet transform and deep belief networks. *Expert Syst. Appl.* 86, 190–198.
- Kim, S., Ulfarsson, G.F., 2004. Travel mode choice of the elderly effects of personal, household, neighborhood, and trip characteristics. *Transp. Res. Rec.: J. Transport. Res. Board* 1894 (1), 117–126.
- Kim, D., Ko, J., Park, Y., 2015. Factors affecting electric vehicle sharing program participants' attitudes about car ownership and program participation. *Transport. Res. Transport Environ.* 36, 96–106. <https://doi.org/10.1016/j.trd.2015.02.009>.
- King, T.L., Bentley, R.J., Thornton, L.E., Kavanagh, A.M., 2015. Does the presence and mix of destinations influence walking and physical activity? *Int. J. Behav. Nutr. Phys. Activ.* 12 (1), 115. <https://doi.org/10.1186/s12966-015-0279-0>.
- Larsen, K., 2015. Safety and School Travel: How does the Built Environment Relate to Correlates of Safety, Mode of Travel and Physical Activity Levels to and from School.
- Lee, B.X., Kjaerulf, F., Turner, S., Cohen, L., Donnelly, P.D., Muggah, R., Davis, R., Realini, A., Kieselbach, B., MacGregor, L.S., 2016. Transforming our world: implementing the 2030 agenda through sustainable development goal indicators. *J. Publ. Health Pol.* 37 (Suppl. 1), 13–31.
- Leyden, K.M., 2003. Social capital and the built environment: the importance of walkable neighborhoods. *Am. J. Publ. Health* 93 (9), 1546–1551.
- Lind, H.B., Nordfjærn, T., Jørgensen, S.H., Rundmo, T., 2015. The value-belief-norm theory, personal norms and sustainable travel mode choice in urban areas. *J. Environ. Psychol.* 44, 119–125. <https://doi.org/10.1016/j.jenvp.2015.06.001>.
- Lundberg, B., Weber, J., 2014. Non-motorized transport and university populations: an analysis of connectivity and network perceptions. *J. Transport Geogr.* 39 (1), 165–178. <https://doi.org/10.1016/j.jtrangeo.2014.07.002>.
- Mahmoud, Z., Ahmad, Y., Bintah Daud, Mahmoud, E., 2016. Examining the walkability of planned neighbourhood in Malaysia: outcome of a pilot study. *OIDA International Journal of Sustainable Development* 9 (1), 85–98.
- McCormack, G.R., Mäse, L.C., Bulsara, M., Pikora, T.J., Giles-Corti, B., 2006. Constructing indices representing supportiveness of the physical environment for walking using the Rasch measurement model. *Int. J. Behav. Nutr. Phys. Activ.* 3 (1), 44. <https://doi.org/10.1186/14795868-3-44>.
- McDonald, N.C., 2007. Active transportation to school: trends among US schoolchildren, 1969–2001. *Am. J. Prev. Med.* 32 (6), 509–516.
- Navarro, P.J., Fernández, C., Borraz, R., Alonso, D., 2017. A machine learning approach to pedestrian detection for autonomous vehicles using high-definition 3D range data. *Sensors* 17 (1), 18. <https://www.mdpi.com/1424-8220/17/1/18>.
- Rashidi, S., Ranjitkar, P., Hadas, Y., 2014. Modeling bus dwell time with decision tree-based methods. *Transp. Res. Rec.: J. Transport. Res. Board* 2418 (1), 74–83. <https://doi.org/10.3141/2418-09>.
- Rišová, K., Madajová, M.S., 2020. Gender differences in a walking environment safety perception: a case study in a small town of Banská Bystrica (Slovakia). *J. Transport Geogr.* 85, 102723.
- Senbil, M., Kitamura, R., Mohamad, J., 2009. Residential location, vehicle ownership and travel in Asia: a comparative analysis of kei-han-shin and Kuala Lumpur metropolitan areas. *Transportation* 36 (3), 325–350.
- Southworth, M., 2005. Designing the walkable city. *J. Urban Plann. Dev.* 131 (4), 246–257.
- Stylianou, K., Dimitriou, L., Abdel-Aty, M., 2019. Big data and road safety: a comprehensive review. *Mobility Patterns, Big Data and Transport Analytics* 297–343. <https://doi.org/10.1016/b978-0-12-812970-8.00012-9>.
- Su, T.T., Azzani, M., Adewale, A.P., Thangiah, N., Zainol, R., Majid, H., 2019. Physical activity and health-related quality of life among low-income adults in metropolitan Kuala Lumpur. *J. Epidemiol.* 29 (2), 43–49.
- Utkin, L.V., Konstantinov, A.V., Chukanov, V.S., Kots, M.V., Ryabinin, M.A., Meldo, A.A., 2019. A weighted random survival forest. *Knowl. Base Syst.* 177, 136–144.
- Wan Omar, W.R., Patterson, I., Pegg, S., 2013. Using a health belief model to investigate the walking behaviour of residents living in Kuala Lumpur, Malaysia. *Annals of Leisure Research* 16 (1), 16–38. <https://doi.org/10.1080/11745398.2013.769422>.
- Yan, X., Richards, S., Su, X., 2010. Using hierarchical tree-based regression model to predict train-vehicle crashes at passive highway-rail grade crossings. *Accid. Anal. Prev.* 42 (1), 64–74. <https://doi.org/10.1016/j.aap.2009.07.003>.
- Yang, M., Wang, W., Chen, X., Wang, W., Xu, R., Gu, T., 2010. Modeling destination choice behavior incorporating spatial factors, individual sociodemographics, and travel mode. *J. Transport. Eng.* 136 (9), 800–810.