The Journal of Transport and Land Use jtlu.org
Vol. 17 No. 1 [2024] pp. 215–239

Non-linear effects of built environment factors on mode choice: A tour-based analysis

Jia Fang
University of Florida
fangjiaseven@gmail.com

Xiang Yan
University of Florida
xiangyan@ufl.edu

Tao Tao
Carnegie Mellon University
taot@andrew.cmu.edu

Changjie Chen
University of Florida
chj.chen@ufl.edu

Abstract: Understanding the connections between the built environment and travel mode choice is a major research topic in transportation. However, existing studies usually examine the relationship through trip-based analyses rather than tour-based approaches. A tour consists of multiple trips that originate and end at the same place, which is increasingly considered the more appropriate analysis unit for travel behaviors. Applying a tour-based approach, this study employs random forest to investigate the non-linear impacts of built environment factors and tour attributes on different mode combinations of a tour. We find that tour attributes and connectivity-related variables (e.g., block size and intersection density) have a strong association with the use of active travel modes when their values are within a certain threshold. In addition, capturing mode change behaviors offers more nuanced understanding of how various built environment variables shape people's decision to combine modes in a tour.

Keywords: Non-linear effects, built environment, tour-based mode choice, SHAP, random forest

Article history:
Received: June 18, 2023
Received in revised form: October 15, 2023
Accepted: November 13, 2023
Available online: March 20, 2024

1 Introduction

As land-use policies are frequently promoted as strategies to support sustainable transportation, considerable empirical work has been done to investigate the link between the built environment and travel mode choice. Existing research in this area has mostly adopted a trip-based approach to perform the analysis by assuming that single trip characteristics (i.e., mode and destination) is a function of some explanatory variable related to this trip. Recently, trip chaining has become an essential part of people's daily travel. Many studies found that tour-based approaches can better capture and represent the travel patterns as tour-based approaches study reflects trip chaining behavior and tour formation (Daisy et al., 2018; Ho & Mulley, 2013; Van Acker & Witlox, 2011). Usually, a tour consists of multiple trips that originate and end at the same place, such as someone's home (Axhausen, 2007). It assumes that people do not make separate

https://doi.org/10.5198/jtlu.2024.2403
ISSN: 1938-7849

The *Journal of Transport and Land Use* is the official journal of the World Society for Transport and Land Use (WSTLUR) and is published and sponsored by the University of Minnesota Center for Transportation Studies.

decisions at the trip level, but schedule activities at the tour level given their spatiotemporal constraints. For example, a person chooses to drive to their nearby workplace instead of walking because they must run some errands far away afterward. In such a case, the commute mode choice is not determined by characteristics of the individual trip (i.e., the location of workplace), but involves a complex planning process for engaging multiple activities with time-space constraints in a tour. Hence, tour-base approaches can provide more comprehensive understanding of the connection between the built environment and travel behavior (Frank et al., 2008; Lee, 2016).

Despite recent efforts to incorporate tour-based approaches into studying the relationship between the built environment and mode choice, two major research gaps remain. First, a tour may involve the use of multiple modes. Most studies simplify the mode choice by using the most representative mode, which means that a single travel mode is assumed to represent all legs of the tour (Daisy et al., 2018; Ho & Mulley, 2013). By assuming away the possibility that people change travel mode in a tour, these studies have disregarded the interdependence among different facets of mode selection (Hasnine & Nurul Habib, 2021). Considering mode change behavior allows new insights into the travel patterns of using different modes at the tour level. For instance, by investigating all the combinations of mode choices in a tour, Miriam and Marco (2016) found that transit is more likely to be combined with other modes than used alone, which indicates that researchers should go beyond examining a single travel mode when public transit is part of the analysis. Moreover, understanding the travel patterns of different mode combinations with transit will better inform planning efforts that aim to promote sustainable travel.

Second, many previous studies have operated under the assumption of linear relationships between built environment variables and travel behavior, proven to be over-simplified according to recent studies. Using machine learning (ML) methods, researchers have discovered non-linear and threshold effects of the built environment on travel behavior (Ding et al., 2021; Hong, 2017; Tao, Wang, et al., 2020). For example, while conventional wisdom has suggested that increasing land-use diversity can promote transit ridership, Shao et al. (2020) have shown that the land-use entropy needs to be greater than 0.5 to have an effective impact on transit use. Traditional statistical models such as linear or logistic regression tend to focus on whether and how much built-environment variables shape travel; by revealing non-linear and threshold relationships, machine learning methods can provide supplementary insights such as informing planning practitioners of the optimal range of the built environment on promoting active travel modes. Despite an increasing number of studies that apply machine learning methods to examine the built-environment and travel-behavior connection, existing work still needs to integrate these efforts with the tour-based approach discussed above.

This study fills these research gaps by employing random forest (RF) to examine the non-linear impacts of the built environment factors on mode choice using a tour-based approach. It contributes to the existing literature in two aspects. First, assessing the different combinations of modes in a tour allows us to account for the interdependency of various modes, thus providing a comprehensive picture of how the built environment affects people's decision to combine various modes at the tour level. Second, this study enriches the understanding of travel behavior research by investigating the threshold impacts of the built environment on tour-based mode choice. It interprets the models by presenting the relative contribution of each variable with SHapley Additive exPlanations (SHAP) method and illustrating the non-linear built environment-active travel associations with accumulated local effects (ALE) plots.

The rest of the paper is organized as follows. In Section 2, we review the literature on the relationship between the built environment and mode choice using a tour-based approach and identify research gaps. In Section 3, we introduce the data and the method. In Section 4, we discuss the results. In the last section, we summarize the key findings and discuss the study's implications for planning practice.

2 Literature review

2.1 Tour-based approaches to study the built-environment and mode-choice relationship

The link between the built environment and mode choice decision has been studied extensively over the past several decades, but the results have been mixed. Some authors, such as Duncan (2016) and Stead and Marshall (2001) concluded that the influence of the built environment is expected to be minor. On the other hand, Ewing and Cervero (2017) provided strong evidence that a dense area with high connectivity generally reduces car driving. For example, while many studies confirm the role of density in promoting non-auto modes (Barnes, 2001; Chatman, 2003; Concas & DeSalvo, 2012; Gehrke & Welch, 2017), some contend that the impact of density is limited (Ewing & Cervero, 2001, 2010, 2017; Handy, 1997). This is likely because the features that come with density (i.e., limited parking spaces and congestion) do not have as much of an effect on non-auto travel. Such inconsistent results may result from using data from different geographical areas with different lifestyle and demographic backgrounds, different time horizons, various ways of measuring the built environment, and different units of analysis.

Recent studies found that tour-based approaches can better capture and represent the relationship of mode choice and built environment than trip-based approaches (Antipova & Wang, 2010; Daisy et al., 2018; Ho & Mulley, 2013). Lund et al. (2004) found that more mixed-use environments near residences promote chain trips by transit and walking, implying that people would like to walk to take transit when living in a mixed land-use environment. Fang (2022) identified the positive indirect effect of density and design on non-auto use through the analysis of tours with multiple activities clustered at one destination. Her study found that a dense development with a quality street design at destination encourages people to carry out activities in proximity, thus resulting in greater non-auto use.

However, previous tour-based studies have only considered the main mode as a dependent variable in estimating the role of the built environment on travel. They either decide on the main mode based on a hierarchy of modes (Chowdhury & Scott, 2018) or using the mode of the main activity of a tour (Ho & Mulley, 2013; Primerano et al., 2008). These practices essentially use trip-based method to address a tour-base problem (Hasnine & Nurul Habib, 2021). Notably, tour-based mode choice has been explored thoroughly in the travel demand modeling studies. Some studies adopt combinatorial tour-based mode choice. For example, Bhat (2004) predicted combinatorial tour-based mode choice using multinomial logit models. However, this method is only suitable for simpler trip combinations. Recently, some advanced models have captured all trips within a tour by considering the dynamics of interdependence among various aspects of mode choices (Cirillo & Axhausen, 2010; Han et al. 2021; Hasnine, 2019; Saleem et al., 2018). These models allow estimating the impact of each trip destination on the corresponding trip mode in a tour sequentially. They assume that people decide the trip mode during a tour when travel patterns unfold; thus, each trip mode choice is assessed on the information for that trip and adjacent trips. However, this approach might not accurately represent some travelers' decision-making processes. In many cases, travelers have to make decisions in advance, as some travel options (i.e., mode and activity) might become unavailable in the middle of the tour (Primerano et al., 2008). Fang et al. (2022)

found that individual's main mode choice is influenced not only by the built environment of the main destination, but is also strongly affected by the activity place with the least compact built environment. For example, people may use transit over biking because the last destination (absent of bike lane) is unsafe for biking.

2.2 Non-linear effects in the built environment and travel behavior connection

Recently, the non-linear associations between built environment factors and travel have attracted growing research attention (Xu et al., 2021; Wang et al., 2022; Yang, Ao et al., 2021; Yang et al., 2022). For example, Ding et al. (2021) examined the non-linear association between built environment factors and public transit use for commuting purpose using a semi-parametric multilevel mixed logit model. Their results show a non-linear relationship and spatial heterogeneity across traffic analysis zones (TAZ). Tao et al. (2020) adopted the gradient boosting decision tree technique to examine the correlates between built environment factors and walking distance to transit service. They found that built environment factors have the greatest predictive power on walking distance among the examined factors, and the effects are non-linear. Other studies investigated the non-linear effects between the built environment and other travel outcomes, including transportation emissions (Hong, 2017), rail transit ridership (Ding et al., 2019), travel time allocation (Wang & Ozbilen, 2020), and urban vitality (Xiao et al., 2021; Yang, Cao et al., 2021).

Compared to ML models, well-developed traditional statistical models have the advantage of being simpler to understand, user-friendly outcomes of correlation and significance, and a more straightforward model structure. However, since researchers usually assume linear relationships when applying statistical models to study the built-environment and travel-behavior relationship, variable associations are not well represented if the actual relationship is non-linear (Ley et al., 2022). ML methods usually contain less assumptions about the data relationships, allowing them to explore more flexible modeling structures. This capability enables ML to often capture the variable associations more effectively. A notable application of ML in the built-environment and travel-behavior literature is the identification of threshold effects, which can help planners and policymakers determine the most effective range of these variables to promote sustainable travel. For instance, Wali et al. (2021) found that walking can be effectively encouraged with the intersection density in the range of 100-200 intersections/km^2, and Tao et al. (2020) indicated that the use of active travel modes decreases substantially when the distance to park increases from 1 to 0.2 miles. These findings offer planning and design guidance for effective land-use strategy implementation to support health travel.

Some recent studies have concentrated on examining the non-linear impacts of the built environment on travel mode. Park et al. (2020) employed a generalized additive model (GAM) to identify an optimal range of the built-environment factors that maximize the probability of walking and transit use. This study shows that the possibility of walking rather than driving peaks at an activity density of 40,000 while controlling for other variables. Liu et al. (2021) examined the non-linear associations between the built environment and walking/biking for work and shop purposes. Their results show that there are non-linear associations between all built environment factors and the walk/bike mode, and the connections for commuting tend to be U or V-shaped. Those studies reveal the intricate effects of the built environment on mode choice and indicate that decision-makers should meticulously disentangle this complexity to come to more effective planning policies.

While most studies examined the non-linear association at the trip level, Kim (2021) found that tour-related attributes exhibit significant interactions with travel modes and hold substantial importance in understanding travel behavior. His results imply that the tour-based analysis is essential for a better understanding of a traveler's mode choice decision. However, little effort has been made to understand the non-linear influence of the built environment on mode choice at the tour level. This study aims to fill the gap by investigating the non-linear effects of built-environment variables and tour-based attributes on tour-based mode choice decisions.

3 Data

3.1 Study area and data sources

The area selected for this study is the area of Portland Metropolitan located in Oregon, which cover parts of Clackamas, Multnomah, and Washington Counties (Figure 1). Portland, as the core of the area, is served by a comprehensive public transportation system that includes bus service, a light rail network, a streetcar, and a bikesharing system. The city has a strong emphasis on fostering transit-oriented development (TOD) and promoting sustainable travel, and so examining the impacts of the built environment on travel mode has great policy significance.

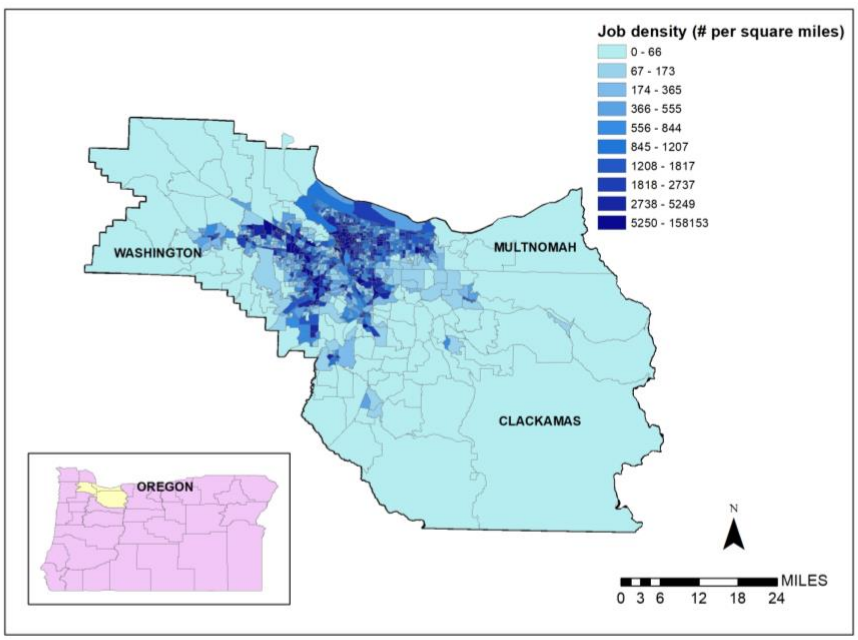

Figure 1. Study area

Data on travel and activities were obtained from the 2011 Oregon Travel and Activity Survey (OTAS), which is the most recent and comprehensive household travel survey conducted in Oregon since 2011. After data cleaning, a total of 9,969 home-based tours were included in the final analysis.

3.2 Travel mode combinations

This study captured the mode change behavior in a tour. Initially, we had a comprehensive set of travel modes for a trip, including car driver, car passenger, transit, bike, and walk. There are 25 mode combinations in our dataset, as shown in Table 1. For modeling, we merged some trip modes to increase the sample sizes of the minority classes with the following rules: 1) car driver and car passenger were jointly considered as the car alternative; 2) walk and bike were jointly considered as active mode (WB); 3) car & transit & WB was merged into car & transit. Finally, we obtained six mode combinations: car, WB, transit, car & transit, car & WB, and transit & WB (See Table 2).

Table 1. All mode combinations in the sampled data

Mode Combinations	Count	Mode Combinations	Count	Mode Combinations	Count
D	7618	DW	181	BTW	6
P	2386	DT	79	DPT	5
W	1088	PTW	64	DPTW	4
T	920	BW	44	BDW	3
B	394	BP	25	BDT	1
PT	365	DPW	24	BPTW	1
TW	217	DTW	23	BPW	1
PW	214	BD	19		
DP	183	BT	19		

Note: D: car driver; P: car passenger; W: walk; T: Transit; B: bike.

Table 2. Distribution of merged mode combinations by tour complexity

Merged Mode Combinations	Count	Share	Simple Tours	Complex Tours	Average Number of Trips
Car	7585	78.2%	56.0%	44.0%	1.9
WB	991	10.2%	77.5%	22.5%	1.4
Transit	452	4.7%	88.9%	11.1%	1.1
Car & WB	261	2.7%	10.0%	90.0%	3.7
Car & Transit	224	2.3%	26.8%	73.2%	2.7
Transit & WB	181	1.9%	8.8%	91.2%	3.0

These descriptive statistics allow one to better understand the distribution of the mode combinations in the dataset. The percentages presented in Table 2 summarize the multimodality behaviors. As expected, people typically rely on a single mode when they travel. The car alternative is the predominant mode accounting for 78.2%, and WB and transit account for 10.2% and 4.7%, respectively. By contrast, the share of each multi-modal combination is below 3%. In addition, we find that trip frequency is an essential determinant of mode choice, which the trip-based models have generally ignored. Simple tours consist of one activity, while complex tours include multiple ones. Among car travel, 56.0% are simple tours, and 44.0 % are complex tours. For WB travel, over 77.5% are simple tours, which is reasonable as people often have low accessibility to various activities within walking or biking distance. Transit travel has the highest percentage of simple tours because of the inefficiency and inflexibility of fulfilling multiple activities only by transit. For multi-modal travel, it is expected that car & WB and transit & WB travel will have a high proportion of complex tours. Perhaps, most people carry out sub-tours by walking or biking. However, for car & transit travel, simple tours account for a relatively high percentage (26.8%).

3.3 Built environment factors, tour attributes, and socioeconomic variables

The built environment was assessed within quarter-mile buffers of both the residential locations and trip destinations. We calculated built-environment variables in five dimensions ("5Ds") using Geographic Information System tools (Chen et al., 2022). Household density represents the density dimension. The indicators of diversity include land-use entropy and job-population balance. Design measures consist of intersection density and block size. The indicators of distance to transit include the proximity to the closest transit stop and stop density. The measure of destination accessibility is based on the count of jobs that can be reached within a 45-minute driving time during peak periods.

The explanatory variables used for controlling purposes consist of tour characteristics and socioeconomic factors. Tour attributes include the tour distance, main purpose of a tour, duration of main activity, tour departure time, departure time during peak hours, and tour complexity. The main purpose could be categorized into three regarding the flexibility of scheduling and location (Stopher et al., 1996). They are mandatory, flexible, and optional activities. We captured tour complexity from two aspects: trip frequency and trip distribution type. The latter includes four categories. Travel with activities around home (within walking distance) was coded as MPAH. Single-purpose travel with a single destination far from home was coded as SPSD. Travel with multiple purposes clustered at a single destination far from home was coded as MPSD. Lastly, travel with multiple purposes widely distributed at multiple destinations was MPMD. Socioeconomic variables include individual- and household-level variables. At the household level, household workers, motor vehicle, and bike availability are included, while individual-level variables include age, gender, and driver's license ownership.

The summary statistics for the explanatory variables are presented in

Table 3. After extracting the data, we perform a variance inflation factor (VIF) analysis to evaluate potential multicollinearity between the independent variables. The VIF value of each variable is well below 5, and the tolerance statistics are well above 0.2, indicating that there is little concern for multicollinearity within our data.

Table 3. Summary statistics for the explanatory variables

Variables	Description	Average	Std dev	5th percentile	95th percentile
	Tour attributes				
NUM_STOP	Number of stops in a tour	1.9	1.43	1	5
TOURDIST	Total tour distance (in miles)	10.67	10.74	0.59	32.01
PEAKHR	Denote if traveler departs in the peak hours (morning/afternoon)	Yes: 63.5% No: 36.5%			
MAIN_DUR	Activity duration of the main activity (in mins)	210	219	3	590
TYPE	Trip distribution type, coded as three dummy variables in modeling: MPSD, MPMD and MPAH	MPSD: 5.6% MPMD: 33.7% MPAH: 9.2% SPSD: 51.5%			

Variables	Description	Average	Std dev	5th percentile	95th percentile
MAIN_PUR	Main purpose of a tour, including mandatory, optional, and flexible. It is coded as two dummy variables: MANDATORY and OPTIONAL	MANDATORY: 45.6% OPTIONAL: 21.6% FLEXIBLE: 32.8%			
IF_ESCORT	Denote if the tour involves an escort trip	Yes: 18.1% No: 81.8%			
	Socioeconomic variables				
AGE	Traveler's age	50.81	14.59	25	74
LIC	Denote if the traveler owns a driver's license	Yes: 95.6% No: 4.4%			
GEND	Donate if the travel is male	Male: 44.6% Female: 55.4%			
HHSIZ	Number of family members in the household	2.79	1.31	1	5
HHVEH	Number of cars in the household	2.06	1.08	1	4
BIKES	Number of bikes in the household	1.60	1.86	0	5
INCOME	Family income 1=$0 - $34,999; 2=$35,000 - $99,999; 3=$100,000 or more	2.17	0.66	1	3
HHLIC	Amount of family members with driver licenses in the household	2	0.76	1	3
HHWRK	Number of employees in the household	1.62	0.86	0	3
	Built environment variables				
STOPDEN_O	Density of transit stops at the origin (per square mile)	17.86	21.40	0	56.05
NEARSTOP_O	Proximity to the closest transit stop at the origin (in mile)	0.9	2.59	0.03	4.04
INSDEN_O	Density of intersections at the origin (per square mile)	112	57	10	214
HOUDEN_O	Number of households per square mile at origin	2412	2401	82	5213
BLOCKSIZE_O	Average block sized within 1/4-mile radius of origin (in mi2)	0.08	0.37	0	0.33
ENTROPY_O	Entropy for land-use types at the origin	0.13	0.14	0	0.39
ACCJOB_O	Job accessibility at origin. It is the count of jobs that can be accessed within a 45-minute travel time during peak periods.	869259	156261	562696	977567

Variables	Description	Average	Std dev	5th percentile	95th percentile
JPRATIO_O	Job-population balance at origin. The closer to zero the index is, the more balanced the area is. A higher index indicates a higher job proportion.	-0.77	0.54	-1.56	0.15
STOPDEN_D	Density of transit stops at the destination (per square mile)	36.46	37.35	0	112.1
NEARSTOP_D	Proximity to the closest transit stop at the destination (in mile)	0.48	1.98	0.01	1.31
INSDEN_D	Density of intersections at the destination (per square mile)	118	84	10	290
HOUDEN_D	Number of households per square mile at destination	2796	3402	24	10914
BLOCKSIZE_D	Average block sized within 1/4-mile radius of destination	0.05	0.19	0	0.17
ENTROPY_D	Entropy for land-use types at the destination	0.23	0.14	0	0.48
ACCJOB_D	Job accessibility at destination. It is the count of jobs that can be accessed within a 45-minute travel time during peak periods.	895639	164642	584905	988764
JPRATIO_D	Job-population balance at destination. The closer to zero the index is, the more balanced the area is. A higher index indicates a higher job proportion.	0.14	0.97	-1.11	1.81

4 Method

In this chapter, we detail our use of ML models and model-agnostic interpretation methods to understand the threshold impacts of built-environment and tour-related attributes in tour-based mode choice. Figure 2 is a workflow diagram demonstrating the process of model training, validation, and interpretation.

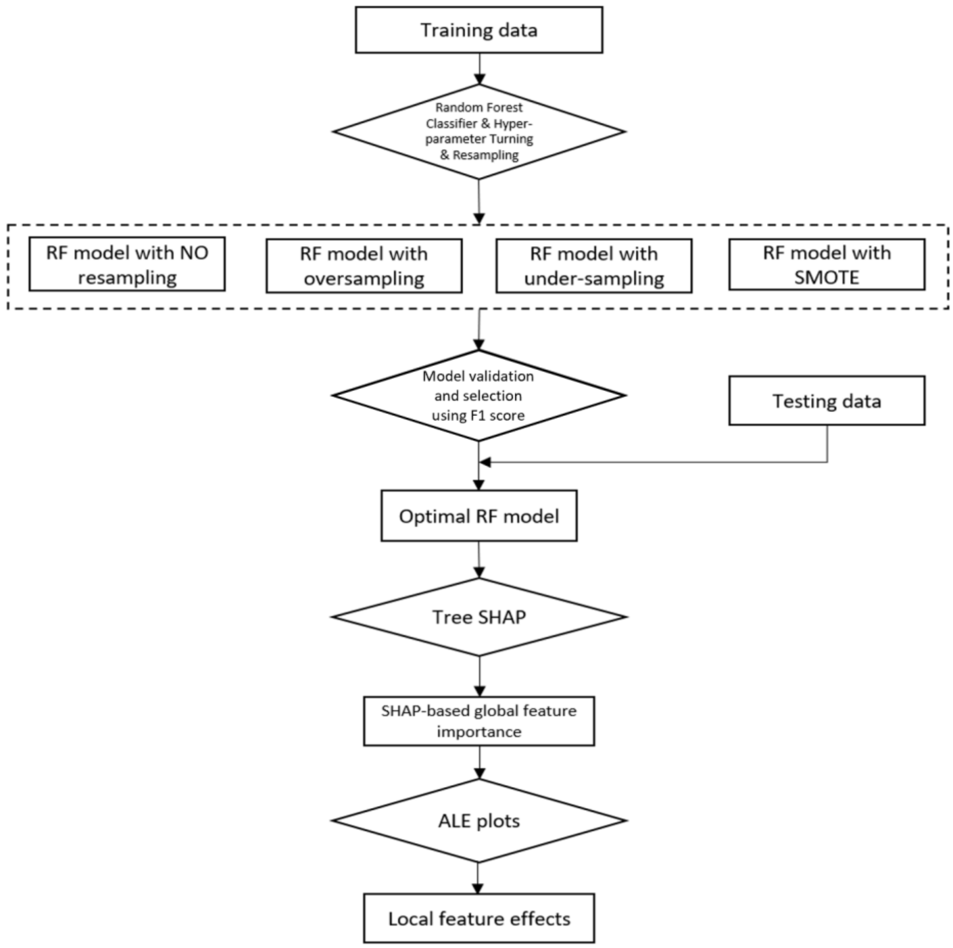

Figure 2. Workflow diagram

4.1 Machine learning model for predicting travel mode

This study employed RF to unravel the intricate relationship between the built environment and mode choice. RF is a commonly utilized and widely acknowledged tree-based method. One of the major advantages is its highly accurate prediction, even for skewed distributions, large proportion of missing data, and irrelevant variables (Breiman, 2001). Moreover, RF can accommodate the non-linear relationship between the variables with much concern for overfitting. Some studies found that RF models outperform other machine learning models in predicting mode choice (Hagenauer & Helbich, 2017; Zhao et al., 2020).

Tuning hyperparameters in RF helps to increase the predictive capabilities of the models. We used a 10-fold cross-validated grid search with a repetition of three to tune the following set of hyperparameters:

- Total count of trees
- Maximum count of attributes to be evaluated for dividing a node
- Maximum count of layers in each decision tree
- Minimum count of data points required for a node to be divided

In machine learning, the classification performance of a model is affected by dataset imbalances (Thabtah et al., 2020). To address the imbalanced mode shares in our dataset, we applied data resampling inside the cross-validation procedure. Three traditional resampling methods were compared: (1) classic oversampling, which oversamples the minority class; (2) under-sampling, which under-samples the majority class; and (3) the synthetic minority oversampling technique (SMOTE), a method of oversampling generating synthetic samples for the minority class. A standardized collection of hyperparameter combinations is considered for each RF model applied with different resampling methods.

Regarding performance metrics in model comparison, we applied the F1 score as it is more reliable than other metrics (i.e., recall and precision) when we have an uneven class distribution. F1 score was calculated at two levels. One was with mode-specific metrics, which measure the fraction of accurate predictions for a specific mode. The other was with average metrics, which can be calculated in two ways: macro and weighted averages. The former takes the average over mode-specific metrics, thus treating all mode classes equally, while the latter combines the contributions of all classes to compute the average metric. Because we valued the minority class, we focused on the macro-averaged F1 score. Based on these metrics, we obtained the best model for result interpretation.

4.2 Model-agnostic interpretation methods

This study used SHAP method to conduct the global model interpretation. SHAP is an approach that combines the conventional Shapley values from coalitional game theory to provide explanations for individual predictions (Lundberg & Lee, 2017). As the only consistent and locally accurate method, SHAP values are theoretically optimal for tree ensemble feature attribution but challenging to compute (Lundberg et al., 2019). Thus, we adopted a tree SHAP algorithm introduced by Lundberg et al. (2019), which calculates the SHAP values of tree-based models at a high speed. Features with high absolute SHAP values indicate that they have large contribution in estimating the dependent variable. The SHAP feature importance is computed by taking the mean of the absolute Shapley values $\emptyset_i$ for per feature i, shown in Equation 1:

$$\frac{1}{n} I_i = \sum_{j=1}^{n} \left| \emptyset_i^{(j)} \right|. \tag{1}$$

Compared to the traditional variable importance approaches using simple global approximations, this method provides a richer and more accurate picture of the model's global pattern. It utilizes numerous high-quality local explanations to reflect the overall structure of the global model, all while maintaining a faithful representation of the original model at the local level. (Lundberg et al., 2020).

Finally, we used accumulated local effects (ALE) plots to visualize the impact of features on the prediction of the outcome. Generally, partial dependence plots (PDPs) are used to demonstrate the marginal effect of features. However, PDPs cannot be trusted if the features are correlated (Molnar, 2021). The issue is resolved by ALE plots through the computation of prediction differences instead of means, thus enabling the capture of the pure impact of features.

5 Results and discussions

5.1 Predictive performance

After the hyper-parameter tuning and resampling process, we obtained four best-performing RF models with different resampling methods (Table 4). Concerning mode-specific metrics, the F1 score values of all the models exhibit consistent patterns. That is, the most accurate predictions are observed for car travel, followed by WB and transit travel. Compared to the majority class (exclusively using cars), the models show poor performance in predicting the minority classes. However, all the resampling methods significantly improve the prediction of all the minority classes. For instance, while the predictive accuracy of car & WB travel in the non-resampled model is 5.7%, the oversampling method increases this figure to 28.6%. The values for the SMOTE and under-sampling methods are 26.2% and 16.9%, respectively. Another vast improvement is observed in car & transit travel, as the predictive accuracy is increased by 18.2% with the SMOTE method. Regarding the average metrics, the SMOTE method has the best fitness to the data, with the highest macro and weighted F1 score. Therefore, we chose the RF model with the SMOTE method as our final model. Its optimal combination of hyper-parameters is listed below.

- Total count of trees: 190
- Maximum count of attributes to be evaluated for dividing a node: 5
- Maximum count of layers in each decision tree: 60
- Minimum count of data points required for a node to be divided: 10

Table 4. Performance of RF models using various resampling methods

Methods	Car	WB	Transit	Car & Transit	Car & WB	Transit & WB	Macro F1	Weighted F1
No resampling	0.918	0.638	0.569	0.154	0.057	0.500	0.473	0.824
Under-sampling	0.699	0.539	0.460	0.274	0.169	0.431	0.429	0.642
Classic oversampling	0.912	0.642	0.624	0.312	0.286	0.554	0.555	0.833
SMOTE	0.909	0.658	0.629	0.336	0.262	0.575	0.561	0.833

5.2 Variable importance

We analyzed the results with SHAP values to find the features with high contributions to travel mode prediction. Table 5 displays the overall relative importance of the variables as well as the collective relative importance by category. Mode choice decision is mainly influenced by tour-related attributes. with many individual variables (i.e., tour distance, trip frequency, and traveling around home) ranked at the top. Vehicle ownership is the most influential variable among demographic attributes, followed by driver's license ownership. Moreover, Our findings indicate that built-environment variables measured at destination have higher ranks than those measured at origin, which indicates that the built environment at destination exerts a greater influence on mode choice compared to those at origin, consistent with the literature (Zhang, 2004). Among the individual built-environment variables, connectivity-related ones, i.e., block size and intersection density are ranked higher than others. Regarding local diversity, land-use mix at both origin and destination has trivial impacts on mode choice. By contrast, job-population balance has a stronger influence than land-use mix, consistent with results of the meta-analysis by Ewing and Cervero (2010).

5.2.1 Non-linear relationship between variables and travel modes

Variable importance provides a first indication of the relationship between a variable and mode choice. To further visualize the exact form of the relationship, we used ALE plots (Figure 3-Figure 8). Note that, for easier comparison, we standardized the y-axis scale for the same variables across different modes.

Table 5. Overall relative importance of variables

Variables	Relative importance	Rank	Sum by variable category
Tour attributes			
TOURDIST	11.7%	1	
NUM_STOP	7.2%	2	
MPAH	6.9%	3	
MAIN_DUR	5.1%	6	
MPMD	4.8%	8	
DEP_TIME	3.2%	11	43.9%
MPSD	2.3%	15	
MANDTORY	1.4%		
IF_ESCORT	0.9%		
OPTIONAL	0.2%		
PEAKHR	0.1%		
Socioeconomic variables			
HHVEH	6.6%	4	
LIC	2.5%	14	
BIKES	2.0%		
AGE	1.5%		
HHLIC	1.1%		15.6%
INCOME	0.7%		
HHSIZ	0.5%		
GEND	0.4%		
HHWRK	0.4%		
Built environment variables at origin			
BLOCKSIZE_O	3.1%	12	
INSDEN_O	2.2%		
HOUDEN_O	2.1%		
ACCJOB_O	1.8%		14.0%
NEARSTOP_O	1.5%		
JPRATIO_O	1.2%		
ENTROPY_O	1.1%		
STOPDEN_O	1.1%		
Built environment variables at destination			
BLOCKSIZE_D	6.3%	5	
INSDEN_D	5.0%	7	
STOPDEN_D	4.6%	9	
ACCJOB_D	3.2%	10	26.5%
JPRATIO_D	2.9%	13	
HOUDEN_D	2.3%		
NEARSTOP_D	1.3%		
ENTROPY_D	0.9%		

5.2.2 Tour-related variables

As discussed in the variable importance section, several tour-related variables have high rankings overall. This section focuses on these most important tour-related variables, including tour distance, trip distribution type, and trip frequency.

Figure 3 shows that tour distance exhibits a positive relationship with car or transit mode combinations, while it demonstrates a negative relationship with WB travel., which indicates that a longer tour distance tends to encourage travelers to use a car or ride transit. The impacts of tour distance are much stronger on car and WB travel than on the other alternatives, which is consistent with the literature (Liu et. al., 2022). However, these associations become insignificant when the tour distance exceeds 2 miles (the curve becomes flat) except for the transit alternative: the association between tour distance and riding transit remains positive.

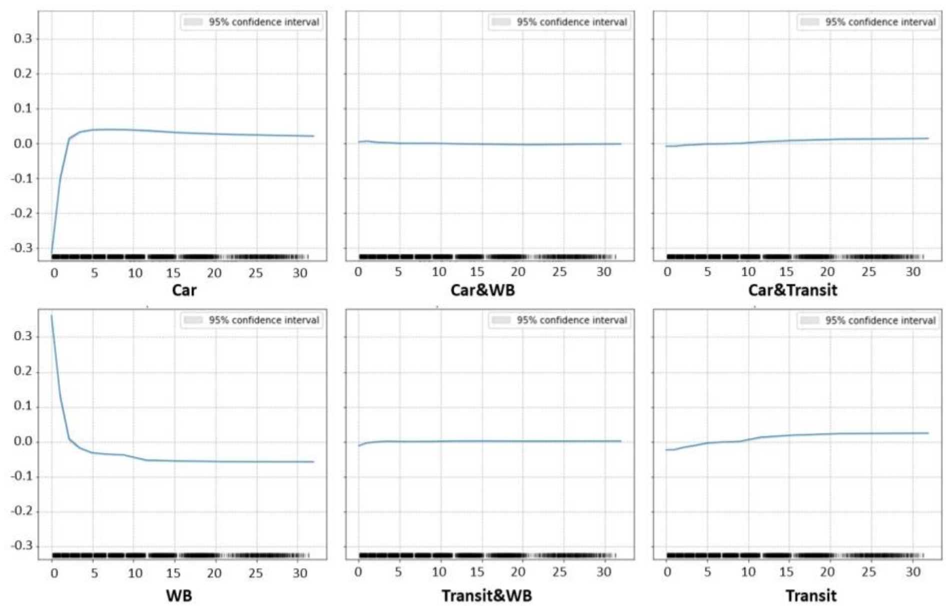

Figure 3. Non-linear associations between tour distance and mode choice

Figure 4 presents that people are likely to use car travel for simple tours or tours with dispersive activities, while people tend to walk/bike if they travel in relatively proximity to home, consistent with previous studies (Harding et al., 2015; Ho & Mulley, 2013). Besides, it is interesting to find that people tend to walk/bike if they make MPSD tours than simple tours even though simple tours tend to have a shorter distance. Transit travel is related to simple tours while car & transit is related to MPMD tours, which is consistent with the descriptive findings discussed above. Moreover, it is reasonable that car & WB and transit & WB travel are highly related to MPSD tours. That is, Individuals may travel by car or public transportation to reach their destination, and then engage in multiple activities within walking distance.

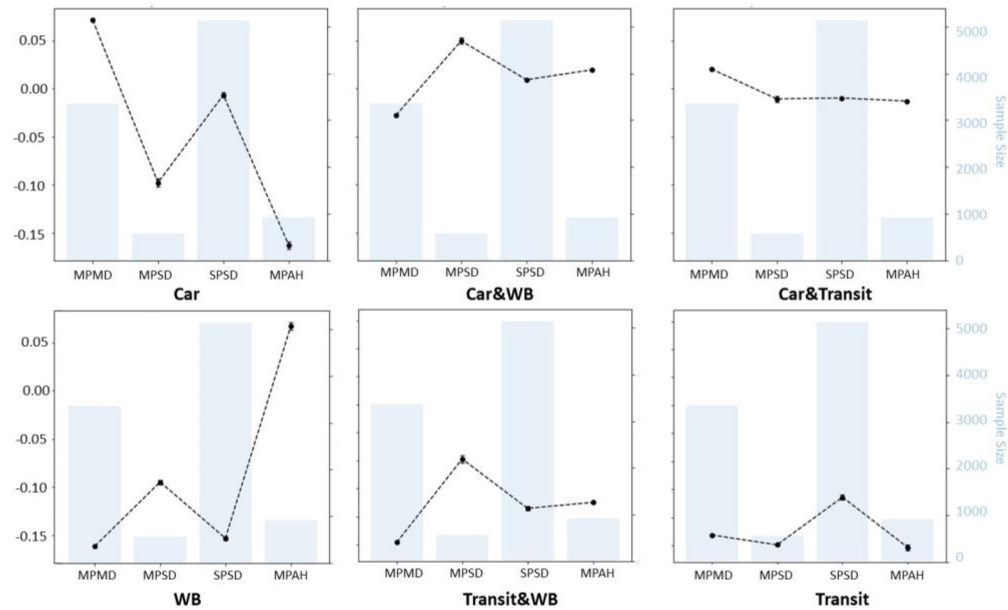

Figure 4. Non-linear associations between tour complexity and mode choice

Concerning trip frequency, we do not focus on WB and car & transit travel, as the associative variables have low importance rankings and show unclear patterns with large confidence intervals. For the remaining four mode combinations, there are two major findings from Figure 5. Previous studies suggest that when the trip frequency pf a tour increases, people tend to use car over transit (Cirillo & Toint, 2001; Huang et al., 2021; Vande Walle & Steenberghen, 2006). However, this study shows that this observation is not necessarily true. While a tour's trip frequency is negative associated with car and transit, its associations with car & WB and transit & WB are positive. Considering together with our findings about MPSD tours, we interpret this finding as suggesting that, when they travel to a dense area by car or transit, people often take advantage of high walking/biking accessibility by combining more trips in a tour.

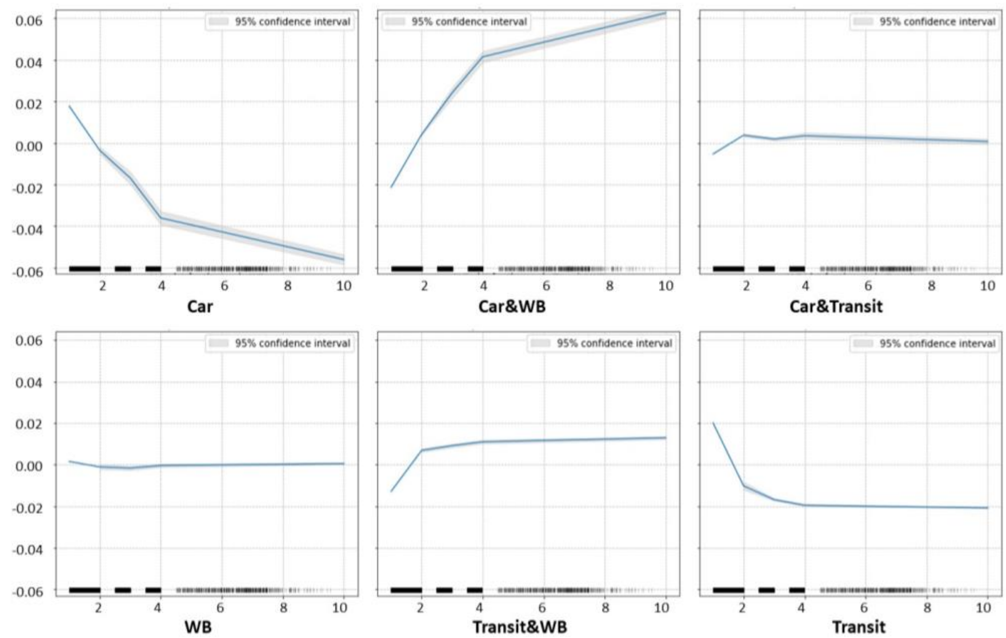

Figure 5. Non-linear associations between trip frequency and mode choice

Finally, it is worth noting that tour-related attributes are not capable of distinguishing the choice between car and transit, which is consistent with the findings of Kim (2021). Probably, given the high quality of transit service in the Portland Metropolitan Area, transit is competitive with cars. In such circumstances, the built environment can significantly contribute to encouraging the use of transit over driving alternatives.

5.2.3 Built environment variables

In this section, we only presented and discussed the ALE plots of the built environment variables with larger relative importance (namely, the top three variables). They are block size at destination, intersection density at destination, and transit stop density at destination.

Figure 6 presents the relationships between block size at destinations and mode choice. Generally, when its value ranges from 0 to 0.018 mi2, a larger block size at destination is positively related to car travel, while an opposite relationship is observed for the other alternatives. These findings are intuitive because short block size is an essential feature related to transit/walking friendless by offering direct travel and various route options (Anabtawi & Scoppa, 2022; Lu et al., 2018, Reilly & Landis, 2002; Saelens et al., 2003). Notably, block size has stronger impacts on WB than other alternatives, indicating that increasing block size will efficiently promote active travel modes.

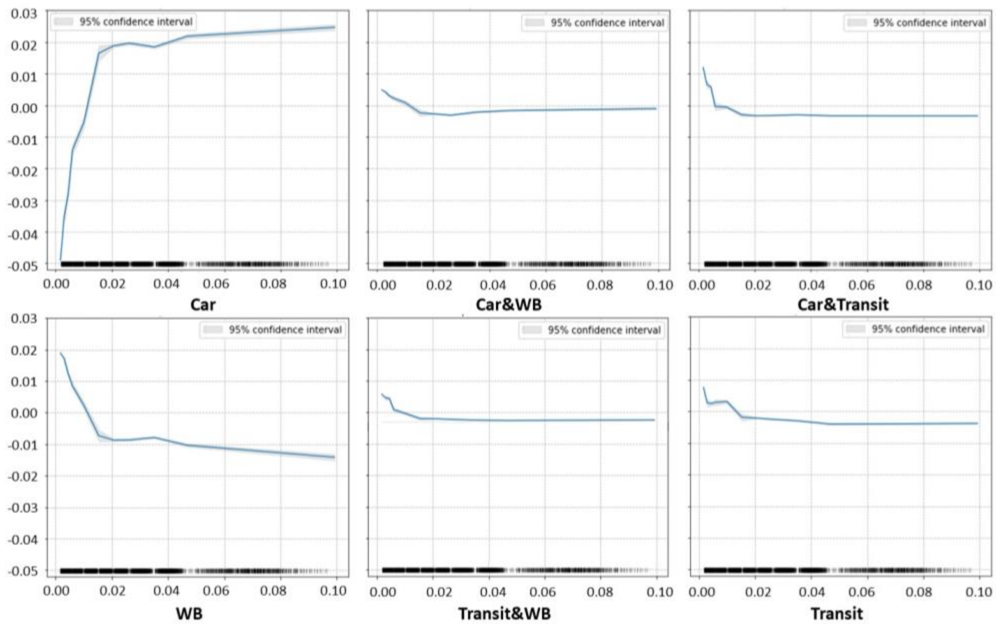

Figure 6. Non-linear associations between block size at destination and mode choice

The overall relationship between intersection density at destination is generally negative with car travel, while its relationships with the remaining alternatives are positive (Figure 7), which is consistent with prior studies (Daisy et al., 2020; Ewing & Cervero, 2010; Lu et al., 2018). However, when intersection density is within 45 /mi^2, we observed a positive relationship between this variable and car use. This is because the benefits of enhanced network connectivity also depend on the scale: when considering a smaller scale, increased connectivity primarily benefits pedestrians; at a larger scale, they benefit auto-users (Reilly & Landis, 2002). These findings suggest that intersection density should reach at least 45 /mi^2 to have the desired effects in discouraging exclusive car use in a tour. Like block size at destination, increasing intersection density will effectively promote the use of active travel modes.

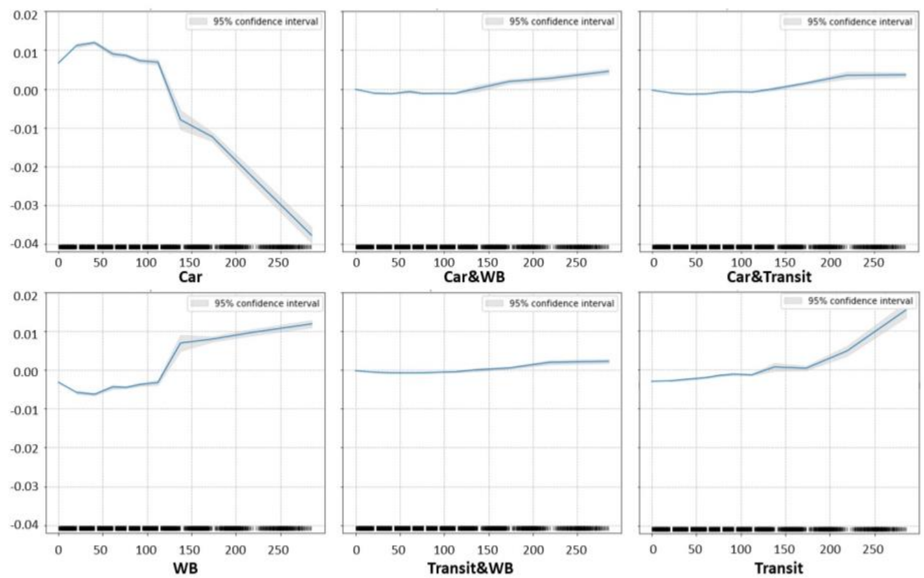

Figure 7. Non-linear associations between intersection density at destination and mode choice

Moreover, transit stop density at destination is negatively related to car travel (Figure 8), which is reasonable, as some drivers tend to switch to using transit when there is a quality transit service (Azimi et al., 2021; Elldér, 2020). The most effective range to discourage car travel over other transit-involved alternatives is 40 to 65 stops/mi2. Surprisingly, stop density at destination shows relatively high importance on car & WB travel, with a positive influence. One possible explanation is that high stop density often goes together with increased connectivity.

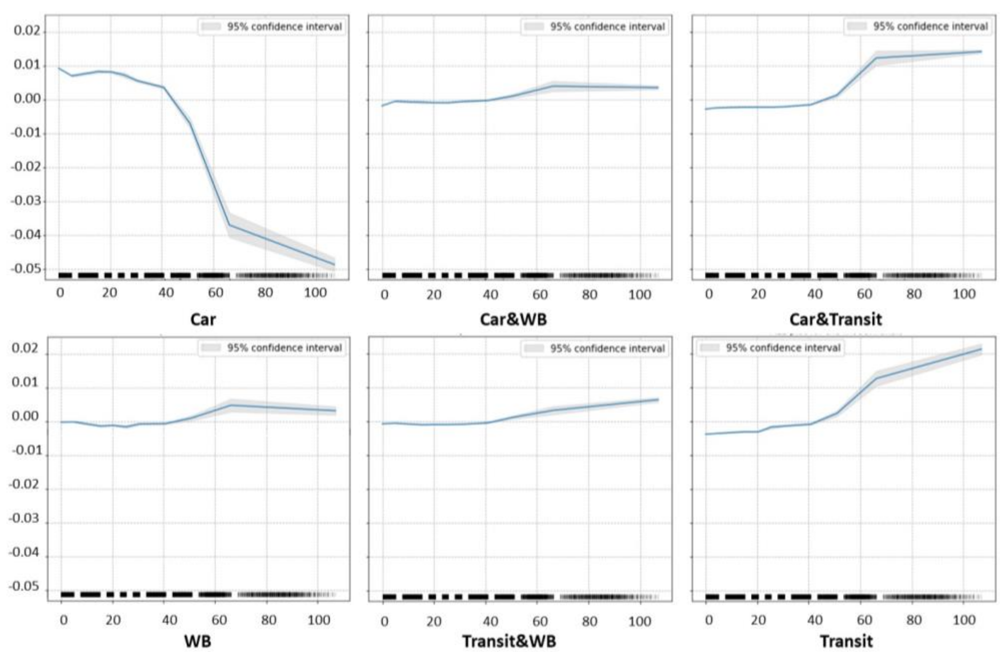

Figure 8. Non-linear associations between transit stop density at destination and mode choice

6 Conclusion and discussion

Using the 2011 Oregon Travel and Activity Survey, the study applied a random forest approach to examine the non-linear effects of the built environment and tour-related attributes on tour-based mode choice decisions, while accounting for sociodemographic variables. We employed the SMOTE method to tackle the data imbalance issue and enhance the accuracy of travel mode choice prediction. We used machine learning interpretation approaches, including variable importance and accumulated local effects, to interpret the model results.

6.1 Contributions to the literature

This study makes two important contributions to current literature. First, adopting ML methods (i.e., RF) advances the tour-base analysis by identifying the non-linear impacts of tour-related attributes and built environment factors on mode choice. With the strongest influence on mode choice, tour distance is positively associated with exclusive car travel when its value is below 2 miles. However, its association with transit use is positive even after its value exceeds 2 miles, which demonstrates that transit is a competitive travel mode in the study region for longer distance travel (i.e., trips over 2 miles). Besides, built-environment variables at both the trip origin and destination have non-linear effects on tour-based mode choice. While the prior studies using the traditional statistical models found the positive relationship between intersection density and non-auto use, our study enriches the literature by identifying the threshold effects of intersection density. To achieve the desired effects to promote non-auto use, intersection density should reach at least 45/mi2; otherwise, the effect is opposite.

In addition, the study generates a clear picture of how the built environment and tour-related attributes affects individual's decision to combine various travel modes in a tour. Under a tour-based analytical framework, previous studies have generally assessed the main mode of a tour by grouping exclusive car travel and car & WB into a single car alternative; in other words, they assumed that car and car & WB have similar travel patterns. However, our finding indicates that, in many cases, the association between the two alternatives and some built environment variables can have opposite signs. For example, a compact urban form is negatively associated with exclusive car travel but has a positive relationship with car & WB travel. Thus, the two modes should be treated differently. Even though both car and car & WB modes involve auto use, the latter is preferable as travelers achieve their daily travel objectives with fewer vehicle miles traveled (VMT) and more active travel. Additionally, some previous studies have found a negative association between trip frequency of a tour and transit use, implying that having more trips in a tour can be a barrier to riding transit (Cirillo & Toint, 2001; Huang et al., 2021; Vande Walle & Steenberghen, 2006). Distinguishing between exclusive transit travel and transit & WB travel, our study provides evidence to challenge this conclusion: findings on transit &WB suggest that many transit users engage in multiple activities within walking distance at some transit-rich destinations.

6.1.1 Policy implications

Our study results have three important policy implications for transportation and land-use planning. First, the model results show that built-environment variables at destinations have higher variable importance than those at origins in estimating travel modes. This implies that policymakers aiming for promoting green travel with land-use strategies are likely to see great policy impact from non-residential areas, like downtowns and employment centers. Notably, block size at destination has the highest variable importance ranking among all built-environment variables, and policymakers should give more attention to this variable.

Second, this study provides the threshold effects of land-use variables on mode choice. The most effective value range for block size influencing car use in a tour is from 0 to 0.018 mi2. Besides, we found that intersection density is negatively related with car use only when its value exceeds 45 ins/mi2; when its value is below this threshold, the association is negative. These findings suggest that a grid street network characterized with small block size and dense intersection tends to encourage greater use of active travel modes. In addition, transit stop density at destination has a significant impact on promoting transit ridership. The most effective value range for this impact is between 40 to 65 stops per square mile. Overall, these findings imply that expanding transit service areas and improving street connectivity are critical for promoting sustainable travel, consistent with the principles of transit-oriented development (Ding et al., 2019).

Third, the results suggest that compact development is effective in encouraging sustainable travel. A travel with different types of purposes completed by active modes near a transit station is a sustainable and efficient chained tour (Harding et al., 2015; Ho & Mulley, 2013). For example, people take transit to the workplace and make multiple trips by foot in proximity before, between, and after work (i.e., café, lunch, grocery, and gym). While lower block sizes, higher intersection densities, and higher transit stop densities lead to combined transit and walk mode shares, proper coordination of land use and transit provision help promoting this type of complex tour by allowing efficient chained trips.

7 Limitations

This study has some limitations. First, we have considered the tour distance for each alternative but not their respective travel times and monetary costs. These travel attributes of non-chosen alternatives are difficult to infer for a tour-based approach when mode change is considered, which warrants further research. Second, it is important to acknowledge that the matter of residential self-selection has not been adequately addressed in this context. Future studies that account for people's attitudes and lifestyles or use longitudinal datasets could mitigate the self-selection bias. Third, the threshold impacts of built environment factors might be context specific. This study is based on the area of Portland Metropolitan located in Oregon, with a supportive environment for walking, biking, and public transportation. Therefore, more studies based on different geographical areas are needed to enhance the transferability and generalizability of the study findings.

Lastly, this study has essentially assumed the tour attributes to be exogenous variables in the mode choice model, which is somewhat problematic as mode choice also shapes tour characteristics. This simultaneity problem is expected to have a larger impact on the magnitude of variable effects than their relative importance. Additionally, some studies have found that the prevailing causal pattern within the population is one where the complexity of the trip chaining pattern influences mode choice (Krygsman et al., 2006; Ye, 2007). Therefore, we consider the results of variable importance to be more robust than the estimated nonlinear effects. That said, we believe that the bias caused by simultaneity is quantitative, not qualitative; in other words, we believe that even though the true nonlinear patterns can differ from what was estimated from our model, nonlinear effects do exist.

References

Anabtawi, R., & Scoppa, M. (2022). Measuring street network efficiency and block sizes in superblocks—Addressing the gap between policy and practice. *Buildings*, *12*(10), 1686. http://doi.org/10.3390/buildings12101686

Antipova, A., & Wang, F. (2010). Land use impacts on trip-chaining propensity for workers and nonworkers in Baton Rouge, Louisiana. *Annals of GIS, 16*(3), 141–154. https://doi.org/10.1080/19475683.2010.513150

Axhausen, K. W. (2007). Definition of movement and activity for transport modelling. In D. A. Hensher & K. J. Button (Eds.), *Handbook of transport modelling* (handbooks in transport) (2nd ed.). Amsterdam: Elsevier.

Azimi, G., Rahimi, A., Lee, M., & Jin, X. (2021). Mode choice behavior for access and egress connection to transit services. *International Journal of Transportation Science and Technology*, *10*(2), 136–155. https://doi.org/10.1016/j.ijtst.2020.11.004

Barnes, G. (2001). *Population and employment density and travel behavior in large U.S. cities.* St. Paul, MN: Minnesota Department of Transportation.

Bhat, C., Guo, J., Srinivasan, S., & Sivakumar, A. (2004). A comprehensive econometric microsimulator for daily activity-travel patterns. *Transportation Research Record, 1894,* 57–66. https://doi.org/10.3141/1894-07

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chatman, D. (2003). How density and mixed uses at the workplace affect personal commercial travel and commute mode choice. *Transportation Research Record, 1831*(1), 193–201. https://doi.org/10.3141/1831-22

Chen, C., Judge, J., & Hulse, D. (2022). PyLUSAT: An open-source Python toolkit for GIS-based land use suitability analysis. *Environmental Modelling & Software, 151*, 105362. https://doi.org/10.1016/j.envsoft.2022.105362

Chowdhury, T., & Scott, D. M. (2018). Role of the built environment on trip-chaining behavior: An investigation of workers and non-workers in Halifax, Nova Scotia. *Transportation*, *47*, 737–761. https://doi.org/10.1007/s11116-018-9914-3

Cirillo, C., & Axhausen, K. (2010). Dynamic model of activity-type choice and scheduling. *Transportation, 37*(1), 15–38. https://doi.org/DOI: 10.1007/s11116-009-9218-8

Cirillo, C., & Toint, P. (2001). An activity-based approach of the Belgian National Travel Survey. Namur, Belgium: FUNDP, Faculté des Sciences. Département de Mathématique.

Concas, S., & DeSalvo, J. (2012). *The effect of density and trip-chaining on the interaction between urban form and transit demand.* Retrieved from https://doi.org/10.2139/ssrn.2133602

Daisy, N. S., Liu, L., & Millward, H. (2020). Trip chaining propensity and tour mode choice of out-of-home workers: Evidence from a mid-sized Canadian city. *Transportation, 47*(2), 763–792. https://doi.org/10.1007/s11116-018-9915-2

Daisy, N. S., Millward, H., & Liu, L. (2018). Trip chaining and tour mode choice of non-workers grouped by daily activity patterns. *Journal of Transport Geography, 69,* 150–162. https://doi.org/10.1016/j.jtrangeo.2018.04.016

Ding, C., Cao, X., & Liu, C. (2019). How does the station-area built environment influence Metrorail ridership? Using gradient boosting decision trees to identify non-linear thresholds. *Journal of Transport Geography, 77*, 70–78. https://doi.org/10.1016/j.jtrangeo.2019.04.011

Ding, C., Cao, X., Yu, B., & Ju, Y. (2021). Non-linear associations between zonal built environment attributes and transit commuting mode choice accounting for spatial heterogeneity. *Transportation Research Part A: Policy and Practice, 148*, 22–35. https://doi.org/10.1016/j.tra.2021.03.021

Duncan, M. (2016). How much can trip chaining reduce VMT? A simplified method. *Transportation, 43*(4), 643–659. https://doi.org/10.1007/s11116-015-9610-5

Elldér, E. (2020). What kind of compact development makes people drive less? The "Ds of the built environment" versus neighborhood amenities. *Journal of Planning Education and Research, 4*0(4), 432–446. https://doi.org/10.1177/0739456X18774120

Ewing, R., & Cervero, R. (2001). Travel and the built environment: A synthesis. *Transportation Research Record, 1780*(1), 87–114. https://doi.org/10.3141/1780-10

Ewing, R., & Cervero, R. (2010). Travel and the built environment: A meta-analysis. J*ournal of the American Planning Association, 76*(3), 265–294. https://doi.org/10.1080/01944361003766766

Ewing, R., & Cervero, R. (2017). Does compact development make people drive less? The answer is yes. *Journal of the American Planning Association*, *83*(1), 19–25. https://doi.org/10.1080/01944363.2016.1245112

Fang, J. (2022). Exploring the role of the built environment in studying mode choice using a tour-based approach [PhD dissertation], University of Florida, Gainsville, FL.

Fang, J., Yan, X., Bejleri, I., & Chen, C. (2022). Which trip destination matters? Estimating the influence of the built environment on mode choice for home-based complex tours. *Journal of Transport Geography, 105*, 103474. https://doi.org/10.1016/j.jtrangeo.2022.103474

Frank, L., Bradley, M., Kavage, S., Chapman, J., & Lawton, K. (2008). Urban form, travel time, and cost relationships with tour complexity and mode choice, *Transportation, 5*(1), 37–54.

Gehrke, S., & Welch, T. (2017). The built environment determinants of activity participation and walking near the workplace. *Transportation*, *44*(5), 941–956. https://doi.org/10.1007/s11116-016-9687-5

Hagenauer, J., & Helbich, M. (2017). A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications, 78*, 273–282. https://doi.org/10.1016/j.eswa.2017.01.057

Han, Y., Chen, C., Peng, Z., & Mozumder, P. (2021). Evaluating impacts of coastal flooding on the transportation system using an activity-based travel demand model: A case study in Miami-Dade County, FL. *Transportation, 49*(1), 163-184. https://doi.org/10.1007/s11116-021-10172-w

Handy, S. (1997). Travel behavior issues related to neo-traditional developments—A review of the research. Paper presented at the Urban Design, Telecommuting and Travel Forecasting Conference, October 27–30, 1996, Williamsburg, VA.

Harding, C., Miller, E. J., Patterson, Z., & Axhausen, K. W. (2015). Multiple purpose tours and efficient trip chaining. An analysis of the effects of land use and transit on travel behavior in Switzerland. Paper presented at the 94th International Conference of the Transportation Research Board, January 11–15, Washington, DC.

Hasnine, M. S. (2019). *Tour-based mode choice model in activity-based modelling framework.* Toronto: University of Toronto.

Hasnine, M. S., & Nurul Habib, K. (2021). Tour-based mode choice modelling as the core of an activity-based travel demand modelling framework: A review of state-of-the-art. *Transport Reviews*, *41*(1), 5–26. https://doi.org/10.1080/01441647.2020.1780648

Ho, C., & Mulley, C. (2013). Multiple purposes at single destination: A key to a better understanding of the relationship between tour complexity and mode choice. *Transportation Research Part A: Policy and Practice*, *49*, 206–219. https://doi.org///doi.org/10.1016/j.tra.2013.01.040

Hong, J. (2017). Non-linear influences of the built environment on transportation emissions: Focusing on densities. *Journal of Transport and Land Use, 10*(1), 229–240. https://doi.org/10.5198/jtlu.2017.815

Huang, Y., Gao, L., Ni, A., & Liu, X. (2021). Analysis of travel mode choice and trip chain pattern relationships based on multi-day GPS data: A case study in Shanghai, China. *Journal of Transport Geography*, *93*, 103070. https://doi.org/10.1016/j.jtrangeo.2021.103070

Kim, E. J. (2021). Analysis of travel mode choice in Seoul using an interpretable machine learning approach. *Journal of Advanced Transportation, 2021,* 1–13, https://doi.org/10.1155/2021/6685004

Krygsman, S. C., Arentze, T., & Timmermans, H. (2007). Capturing tour mode and activity choice interdependencies: A co-evolutionary logit modelling approach. *Transportation Research Part A: Policy and Practice, 41*(10), 913–933. https://doi.org/10.1016/j.tra.2006.03.006

Lee, J. (2016). Impact of neighborhood walkability on trip generation and trip chaining: Case of Los Angeles. *Journal of Urban Planning and Development, 142*(3), 1–13. https://doi.org/10.1061/(asce)up.1943-5444.0000312

Ley, C., Martin, R. K., Pareek, A., Groll, A., Seil, R., & Tischer, T. (2022). Machine learning and conventional statistics: Making sense of the differences. *Knee Surgery, Sports Traumatology, Arthroscopy*, *30*(3), 753–757. https://doi.org/10.1007/s00167-022-06896-6

Liu, J., Wang, B., & Xiao, L. (2021). Non-linear associations between built environment and active travel for working and shopping: An extreme gradient boosting approach. *Journal of Transport Geography, 92*, 103034. https://doi.org/10.1016/j.jtrangeo.2021.103034

Liu, S., Yamamoto, T., & Yao, E. (2022). Joint modeling of mode choice and travel distance with intra-household interactions. *Transportation, 50*(5), 1527–1552. https://doi.org/10.1007/s11116-022-10286-9

Lu, Y., Sun, G., Sarkar, C., Gou, Z., & Xiao, Y. (2018). Commuting mode choice in a high-density city: Do land-use density and diversity matter in Hong Kong? *International Journal of Environmental Research and Public Health, 15*(5), 920. http://doi.org/10.3390/ijerph15050920

Lund, H., Cervero, R., & Willson, R. (2004). *Travel characteristics of transit-oriented development in California.* Sacramento: Caltrans.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., … & Lee, S.-I. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence, 2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Lundberg, S. M., Erion, G. G., & Lee, S.-I. (2019). *Consistent individualized feature attribution for tree ensembles*. arXiv.org. https://doi.org/10.48550/arXiv.1802.03888

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4768–4777.

Miriam, P., & Marco, D. (2016). Classification of tours in the U.S. National Household Travel Survey through clustering techniques. *Journal of Transportation Engineering, 142*(6), 04016021. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000845

Molnar, C. (2021). *Interpretable machine learning: A guide for making black box models explainable.* Retrieved from https://christophm.github.io/interpretable-ml-book/index.html

Park, K., Ewing, R., Sabouri, S., Choi, D., Hamidi, S., & Tian, G. (2020). Guidelines for a polycentric region to reduce vehicle use and increase walking and transit use.

Journal of the American Planning Association, 86(2), 236–249. https://doi.org/10.1080/01944363.2019.1692690

Primerano, F., Taylor, M., Pitaksringkarn, L., & Tisato, P. (2008). Defining and understanding trip chaining behavior. Transportation, 35(1), 55–72. https://doi.org/10.1007/s11116-007-9134-8

Reilly, & Landis. (2002). *The influence of built-form and land use on mode choice: Evidence from the 1996 bay area travel survey*. Retrieved from https://escholarship.org/content/qt46r3k871/qt46r3k871_noSplash_4e162f031fb4e3d0da1b5bb571de55f8.pdf

Saelens, B. E., Sallis, J. F., & Frank, L. D. (2003). Environmental correlates of walking and cycling: Findings from the transportation, urban design, and planning literatures. *Annals of Behavioral Medicine*, *25*(2), 80–91. https://doi.org/10.1207/S15324796ABM2502_03

Saleem, M., Västberg, O. B., & Karlström, A. (2018). An activity-based demand model for large scale simulations. *Procedia Computer Science, 130*, 920–925. https://doi.org/10.1016/j.procs.2018.04.090

Shao, Q., Zhang, W., Cao, X., Yang, J., & Yin, J. (2020). Threshold and moderating effects of land use on metro ridership in Shenzhen: Implications for TOD planning. *Journal of Transport Geography, 89*, 102878. https://doi.org/10.1016/j.jtrangeo.2020.102878

Stead, D., & Marshall, S. (2001). The relationships between urban form and travel patterns: An international review and evaluation. *European Journal of Transport and Infrastructure Research, 1*(2), 113–141.

Stopher, P. R., Hartgen, D. T., & Li, Y. (1996). SMART: Simulation model for activities, resources and travel. *Transportation*, *23*(3), 293–312. https://doi.org/10.1007/BF00165706

Tao, T., Wang, J., & Cao, X. (2020). Exploring the non-linear associations between spatial attributes and walking distance to transit. *Journal of Transport Geography, 82*, 102560. https://doi.org/10.1016/j.jtrangeo.2019.102560

Tao, T., Wu, X., Cao, J., Fan, Y., Das, K., & Ramaswami, A. (2020). Exploring the non-linear relationship between the built environment and active travel in the Twin Cities. *Journal of Planning Education and Research*, *43*(3), 637–652 https://doi.org/10.1177/0739456X20915765

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences, 513*, 429–441. https://doi.org/10.1016/j.ins.2019.11.004

Van Acker, V., & Witlox, F. (2011). Commuting trips within tours: How is commuting related to land use? *Transportation, 38*(3), 465–486. https://doi.org/10.1007/s11116-010-9309-6

Vande Walle, S., & Steenberghen, T. (2006). Space and time related determinants of public transport use in trip chains. *Transportation Research Part A: Policy and Practice*, *40*(2), 151–162. https://doi.org///doi.org/10.1016/j.tra.2005.05.001

Wang, K., & Ozbilen, B. (2020). Synergistic and threshold effects of telework and residential location choice on travel time allocation. *Sustainable Cities and Society, 63*, 102468. https://doi.org/10.1016/J.SCS.2020.102468

Xiao, L., Lo, S., Liu, J., Zhou, J., & Li, Q. (2021). Non-linear and synergistic effects of TOD on urban vibrancy: Applying local explanations for gradient boosting decision tree. *Sustainable Cities and Society*, *72*, 103063. https://doi.org/10.1016/j.scs.2021.103063

Xu, Y., Yan, X., Liu, X., & Zhao, X. (2021). Identifying key factors associated with ridesplitting adoption rate and modeling their nonlinear relationships. *Transportation Research Part A: Policy and Practice*, *144*, 170–188.

Wali, B., Frank, L. D., Chapman, J. E., & Fox, E. H. (2021). Developing policy thresholds for objectively measured environmental features to support active travel. *Transportation Research Part D: Transport and Environment*, *90*(January), 102678. https://doi.org/10.1016/j.trd.2020.102678

Wang, W., Rahimi-Golkhandan, A., Chen, C., Taylor, J., & Garvin, M. (2022). Measuring the impact of transportation diversity on disaster resilience in urban communities: Case study of Hurricane Harvey in Houston, TX. *Computing in Civil Engineering, 2019*, 555–562. https://doi.org/10.1061/9780784482445.071

Yang, J., Cao, J., & Zhou, Y. (2021). Elaborating non-linear associations and synergies of subway access and land uses with urban vitality in Shenzhen. *Transportation Research Part A: Policy and Practice*, *144*, 74–88. https://doi.org/10.1016/j.tra.2020.11.014

Yang, L., Ao, Y., Ke, J., Lu, Y., & Liang, Y. (2021). To walk or not to walk? Examining non-linear effects of streetscape greenery on walking propensity of older adults. *Journal of Transport Geography, 94*, 103099. https://doi.org/10.1016/J.JTRANGEO.2021.103099

Yang, Y., Sasaki, K., Cheng, L., & Liu, X. (2022). Gender differences in active travel among older adults: Non-linear built environment insights. *Transportation Research Part D: Transport and Environment, 110*, 103405. https://doi.org/10.1016/J.TRD.2022.103405

Ye, X., Pendyala, R. M., & Gottardi, G. (2007). An exploration of the relationship between mode choice and complexity of trip chaining patterns. *Transportation Research Part B: Methodological*, *41*(1), 96–113. https://doi.org/10.1016/j.trb.2006.03.004

Zhang, M. (2004). The role of land use in travel mode choice: Evidence from Boston and Hong Kong. *Journal of the American Planning Association, 70*(3), 344–360. https://doi.org/10.1080/01944360408976383

Zhao, X., Yan, X., Yu, A., & Van Hentenryck, P. (2020). Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models. *Travel Behaviour and Society, 20,* 22–35. https://doi.org/10.1016/j.tbs.2020.02.003