## RESEARCH ARTICLE

# Non-Linear Associations Between the Urban Built Environment and Commuting Modal Split: A Random Forest Approach and SHAP Evaluation

**FAIZEH HATAMI** [1], **MD. MOKHLESUR RAHMAN**[2,3], **BEHNAM NIKPARVAR**[4], **AND JEAN-CLAUDE THILL**[1,5]

[1]Department of Geography and Earth Sciences, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
[2]The School of Information Studies, 343 Hinds Hall, Syracuse University, Syracuse, NY 13244, USA
[3]Department of Urban and Regional Planning, Khulna University of Engineering and Technology (KUET), Khulna 9203, Bangladesh
[4]Department of Biology, Center for Infectious Disease Dynamics, Penn State University, University Park, PA 16801, USA
[5]School of Data Science, University of North Carolina at Charlotte, Charlotte, NC 28223, USA

Corresponding author: Jean-Claude Thill (jfthill@uncc.edu)

**ABSTRACT** The study of commuting mode choice is crucial since driving, with all its associated environmental and economic consequences, is the United States' most popular mode of transportation due to urban sprawl, priority to road construction and America's love affair with the automobile. More attention needs to be paid to sustainable modes such as public transit and walking. The built environment is expected to have an impact on commuting mode choice. Built environments with higher density, diversity, intentional design, destination accessibility, and shorter distance to transit (collectively known as the 5 Ds of the built environment) are hypothesized to lead to more sustainable mode choices, including public transit and walking. In this paper, we evaluate the impact of built environment variables on commuting modal split, including the four modes of public transit-bus, public transit-rail, walking, and driving. The study is conducted in Mecklenburg County, North Carolina, at the geographic level of census block groups in year 2015. Given the complexity of relationships in the built environment-travel behavior subject, the random forest method is used to predict aggregated commuting mode choice. Random forest is employed as it is capable of capturing nonlinear relationships and is not constrained by limitations in other widely used methods, such as multinomial logistic regression. After predicting the commuting mode shares, SHAP values (SHapley Additive exPlanations) are used to evaluate the impact of the built environment on commuting mode choices. As an advanced machine learning method, SHAP values adds explainability to the model. This method resolves the known limitation of machine learning methods as being "black boxes" and converts them to "white boxes" by providing interpretability. They provide insights into both the direction and magnitude of the relationships. Thanks to its rigorous ML-based design, our study helps to solidify the state of knowledge with strong evidence that block groups with higher degrees of the 5Ds lead to more choices of public transit and walking modes. We discuss urban policy implications of this study.

**INDEX TERMS** The built environment, commuting mode choice, modal split, random forest, SHAP values, feature importance.

## I. INTRODUCTION
### A. CONTEXT AND OBJECTIVES
The transportation system significantly affects social, economic, and environmental sustainability in many developed

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino.

and developing countries due to rapid urbanization and motorization [1], [2]. According to the 2017 National Household Travel Survey [3], 42.3% and 39.8% of trips were made by cars and SUVs/trucks/vans, respectively, in the United States (US). On the other hand, only 11.5% of trips were made on foot or bicycle, 2.6% by public and para-transit, and 0.7% by shared mobility services. This

overwhelming use of motorized vehicles is causing traffic crashes, congestion, and emissions, influencing land-use patterns, and thereby negatively affecting society, economy, and environment [4].

To address these problems, government agencies have been taking strategic actions by building safer transportation infrastructure, improving system reliability and connectivity, expanding people's accessibility, and expanding zero-emission vehicle usage [5]. In this context, pertinent policies undertaken by state and local governments to promote public and active transportation and to curb private vehicle use could uphold government's strategic vision. This study aims to investigate the impacts of the layout of the built environment and people's socioeconomic factors on travel mode choice behaviors of people in Mecklenburg County, North Carolina, US. Specifically, we measure behaviors aggregated to geographic neighborhoods with the percentage of neighborhood residents using a particular mode of transport for their commute to work. We analyze this so-called modal split using the Random Forest (RF) machine learning technique. Results from this study may draw significant insights and guide policy makers to formulate appropriate actionable interventions to achieve a more efficient and more sustainable transportation system in cities. The following three-fold research questions are formulated to understand the non-linear associations between the built environment, people's socio-economic factors, and the modal split:

1) What are the impacts of the built environment through its various dimensions on the modal split?
2) What are the impacts of socio-economic factors on the modal split?
3) What is the relative importance of independent variables in predicting the modal split in geographic neighborhoods?

Driving comprises the biggest share of travel mode choices in the United States at large and also in the preponderance of more local communities due to rather uncontrolled development patterns. This leads to social, economic, and environmental problems such as social exclusion, increased costs of urban infrastructure, and environmental pollution. As a result, novel approaches to practice in urban planning and design such as smart growth, new urbanism, transit-oriented developments, and sustainable design have emerged, aiming to influence people's travel behaviors. Their main objective is to reduce travel demand and consumption and to stimulate the choice of more sustainable modes, such as public transit, walking, and biking, through built environment interventions. Known as the 5Ds of the built environment, these interventions pertain to a suite of built environment properties, including density, design, diversity [6], distance to transit [7], and destination accessibility [8]. They have been conceived as a comprehensive set of actionable measures over a period of about two decades of research works studying the impact of the built environment on travel behavior. The main contention for the advancement of new urban planning and design practices is that built environment variables that align with the 5Ds reduce travel consumption and its negative externalities. Given the considerable investments required to implement built environment interventions, such as compact developments, public transit infrastructures, and walkable design, it is of paramount importance to empirically assess the impact of these built environment variables on travel behavior outcomes beforehand.

To date, numerous studies have investigated the impacts of the built environment on different aspects of people's mobility and travel behaviors (e.g., [7], [9], [10], [11]). Most of these studies have used statistical and econometric models to gauge these effects [12], [13], [14]. The main problem these studies encounter is that they fail to capture the complex functional relationship between the built environment and travel mode choice, especially non-linearities, as noted in the literature [15]. As a new paradigm of scientific research with robust analytical power, machine learning (ML) techniques can handle the complex and non-linear associations between predictors and response variables [16], [17]. A limited number of studies have used ML approaches to unravel the non-linear associations between the built environment and people's travel behaviors [18], [19], [20]. However, this small body of literature has primarily sought to solve predictive problems and has omitted causal inferences. These studies still suffer from the "black box" condition of ML models, which stipulates that we cannot learn about the underlying connections between the different variables used in the model and gain semantic justification from fitted parameters. Hence, it is critical to improve ML from a "black box" to a "white box." Even when studies purport to investigate causal inference while using ML, they use methods such as permutation [21] or filter and wrapper methods [22], which only provide insight into the ranking of features, while overlooking the importance of the direction of relationships. Considering the limitations of the extant literature, this study seeks to identify the factors that influence commuting mode choices using ML techniques and to evaluate the specific impact of the dimensions of the built environment on commuting mode choices using the SHapley Additive exPlanations (SHAP) evaluative values. This advanced method stands out by its ability to inform us about the relationships between the predictors and the outcome variables in terms of both the magnitude and the direction of the relationships.

### B. BUILT ENVIRONMENT, MOBILITIES AND TRAVEL PATTERNS

It has been argued that the layout of the urban fabric, which is commonly referred to as the built environment, has a structural role in influencing people towards active and non-motorized transportation as well as public transportation, and therefore induces people to refrain from driving. Many studies to date have already investigated the impacts of the built environment on people's travel behavior patterns. A brief discussion of the extant literature is provided here to understand these relationships.

### 1) DENSITY

Most researchers investigating the links between the built environment and travel mode choice have used density to measure the built environment. These studies have found that density significantly influences people's travel mode choices. For example, a study in the Washington metropolitan area investigated the impacts of the built environment on travel mode choice [12]. Using a multilevel integrated multinomial logit (MNL) model and structural equation model (SEM), it was reported that residential and employment density increase transit use, walking, and cycling. High population and employment densities at both trip origin and destination increase transit use and decrease solo car driving significantly [23], [24], [25]. Similarly, compact development (i.e., high density) encourages walking, cycling, ride-sharing, and transit use and discourages car reliance for both work and non-work trips (e.g. trips to the bank, or to a dentist) compared to sprawling development [26], [27], [28]. Analyzing information from the 98 largest cities in India, Ahmad and Puppim de Oliveira [29] reported that private transport is prominent in small and medium-sized cities, whereas people in large cities mainly depend on public transport.

Although many researchers have found a significant impact of density, McKibbin [30] only found a moderate influence of population and employment densities on increased walking and cycling in the greater Sydney region, Australia. Frank and Pivo [31] have determined threshold employment (i.e., 75 employees/acre) and population (i.e., 13 persons/acre) densities to encourage walking and transit use and discourage car use. Therefore, high employment density (20 to 75 employees per acre or over 125 employees per acre) significantly shifts travel mode choice to walking and transit use from the car. Similarly, high population density (>13 persons/acre) significantly persuades to shift from car driving to walking and transit use.

Overall, studies highlight the importance of population and employment density to describe travel mode choice. High population and employment densities reduce car travel and increase walking, cycling, and public transit by reducing travel distance. However, threshold population and employment densities are needed to increase walking, cycling, and transit use and non-linearities exist.

### 2) LAND-USE DIVERSITY

A considerable number of studies have also used land-use diversity to evaluate the influence of the built environment on travel mode choice. They have found that greater diversity in land use (i.e., lower land-use segregation) significantly increases walking, cycling, and transit use and decreases car use by reducing travel distance [23], [25], [28]. Besides, multifamily housing reduces solo driving and increases transit use considerably more compared to single-family housing. In contrast, vacant and undeveloped lands are associated with fewer car trips, while commercial land uses produce more car trips [32].

Some studies also used the job-housing balance to measure the mixing of land uses. These studies reported that a high job-housing balance significantly increases walking, cycling, and transit use and decreases car use by reducing travel distance [33], [34]. A higher job-housing balance encourages people to use non-motorized modes of transportation by placing workplaces, service centers, and convenience retail centers close to residences.

Collectively, studies outline the critical role of land uses on travel mode choice. The evidence so far presented advocates that mixed land uses significantly increase walking, cycling, and public transit and reduce car use.

### 3) DESIGN

Studies have assessed the role of street network connectivity on travel mode choice as a measure of the design dimension of the built environment. They have found a significant influence of this factor. The extant literature shows that grid road networks, short block size, high intersection density, low prevalence of T-intersections, and cul-de-sacs significantly reduce car trips by increasing accessibility to activity locations [12], [24], [25], [26]. Moreover, pedestrian-oriented design (i.e., connected sidewalks and pedestrian paths, access to transit) increases walking and transit trips and decreases car trips.

Researchers have also observed that a higher street density tends to increase the cycling tendency of people around transit stations [24], [28], although bicycle density is the most influential factor. Researchers [29] also reported a dearth of non-motorized transport in many parts of cities due to the lack of supporting infrastructure. They argued that the provision of dedicated walking and cycling lanes would act as a catalyst to increase active travel. Thus, a higher number of walking facilities (e.g. wide footpaths, street lights) and cycling facilities (e.g., shared bike services) can increase ride-sharing, transit, walking, and cycling trips [6], [35].

Studies referenced above describe the influence of street network connectivity on travel mode choice. In a word, one may expect that grid street patterns, smaller blocks, and high availability of walking/bike paths can significantly increase walking, cycling, and public transport and decrease car use.

### 4) DESTINATION ACCESSIBILITY

Destination accessibility indicates the ease of access or proximity to destination [11], [33]. Accessibility is measured using distance to destinations or the number of jobs/facilities that can be reached easily. Some studies have used accessibility to destinations to evaluate the influence of the built environment on travel mode choice. They measure accessibility to destinations using transport network connectivity. Higher network connectivity indicates a higher level of accessibility to destinations. These studies mentioned that better accessibility to destinations increases walking, cycling, and transit use compared to car use [33], [34]. Accessibility is particularly influential on the use of slower and non-motorized modes

of travel for which the friction of distance exerts more of a barrier effect to mobility.

### 5) DISTANCE TO TRANSIT

Researchers have emphasized how critical distance to points of access to transit (stations and stops) is in the relationship between the built environment and travel mode choice. They have found that the distance between activity sites and transit stops is negatively associated with walking, cycling, and public transit use and positively associated with car use. For example, a long distance to transit stops increases car use, while decreasing walking, cycling, and transit use [12], [32], [36]. Conversely for short travel distances [33], [37]. A higher number of bus stops tends to enhance the use of public transportation, but also increases the cycling tendency of people around transit stations [24], [28].

The above studies highlight the important role of distance to transit on household travel mode choice. In summary, short travel distance to transit systems reduces car use and increases other modes. Moreover, high density and mixed land use reduce car use and increase walking and cycling by reducing travel distance to transit stops.

### 6) SUMMARY

In a nutshell, the factors of the built environment significantly influence the travel mode choice behaviors of urban populations. The extant literature shows that high population and employment density, mixed land use, and public transport options encourage people towards sustainable transportation (i.e., high use of public transit, cycling and walking, and low use of automobiles) compared to low density and single land use settings. Higher connectivity and quality of the sidewalks, sidewalk density, bike infrastructure, transit facilities, closer integration with public transportation, and pedestrian-friendly street designs increase walking and public transit trips. Similarly, proximity to public facilities and services (e.g., jobs, parks, schools, hospitals, recreational facilities) increases the walking, cycling, and e-biking tendency of urban residents. Thus, appropriate design of the built environment (i.e., effective sidewalks, separate and protected bike lanes, high density, mixed land uses, and better connectivity) could be an effective strategy to increase active and public transport use and reduce solo driving.

### C. SOCIO-ECONOMIC FACTORS, ATTITUDES, SELF-SELECTION, AND MODE CHOICE

A large number of studies have also investigated the influence of people's socio-economic factors, attitudes, and self-selection on household travel mode choice. Most of these studies mentioned that various socio-economic and demographic factors significantly influence people's travel mode choice behaviors. For example, being male, younger, elderly, having low education, and being in the low or middle-income strata would be more associated with walking, cycling, e-biking, and public transit, while employed workers are less

likely to walk due to longer trip lengths and appreciation for speed to reach the workplace on time [35], [38], [39], [40]. In contrast, high-income people, households with children, and early residents of the city are more prone to travel by car [39], [40], [41].

However, researchers have also observed that affluent people frequently participate in physical activities (e.g., walking, cycling) for the health benefits of physical exercise [35]. Thus, the attitudes and preferences of individuals also influence their travel decisions [42]. De Vos et al. [43] underscored the significance of attitudes and residential self-selection on travel and residential decision-making. Analyzing empirical data, this study found that despite greater travel distances and times, some people are interested in living in suburban areas due to their tolerance for greater travel efforts. Accordingly, they self-select their residential neighborhoods and travel with their preferred mode of transportation. Researchers [40] also mentioned that socio-economic considerations equally contribute to explain people's mode choice behaviors. To sum up, socio-economic features, attitudes, and self-selection have significant impacts on travel decisions.

Also, it has been found that perceived safety influences the travel patterns of people [39], [44]. People are reluctant to let their children walk or bicycle to school due to fear of potentially hazardous situations they could encounter [39]. Thus, increasing the number of police officers, particularly at night can significantly increase people's active travel [45], as could well-lit and unobstructed public spaces. Comfort and convenience also influence people's travel mode choices [46]. Thus, psychological factors have stronger effects on mode choice behaviors compared to people's socio-economic attributes [47].

To sum up, people's socio-economic and demographic attributes, preferences, and perceived safety significantly influence mode choice behaviors. Thus, estimating the impacts of the built environment ignoring socio-economic factors may overstate the influence of the built environment on travel mode share.

### D. TRAVEL FACTORS AND TRAVEL MODE CHOICE

A considerable number of studies also found a significant influence of travel factors on travel mode choice. For example, higher traffic speed (i.e. more than 30 miles/hour) is associated with less walking and cycling to school [37]. People's propensity to use transit is reduced in case of a higher number of transfers, longer travel time, walking time to stops to catch the bus or train, and waiting time [23], [48], [49]. Researchers [36], [50] have found that fuel price and parking costs have a significant influence on vehicle type choice. Households switch from large vehicles to small, more fuel-efficient vehicles to reduce overall travel costs. Furthermore, people often switch to car use due to increased transit travel costs [23], [51]. It is also well known that spikes in gas prices lead to changes in modal split away from driving alone.

The availability of public transportation and restrictions on car use can effectively promote active transportation in the

population [52]. Policies and strategies to enhance people's accessibility to public transportation, to reduce travel time and transfer distance, to create a comfortable travel environment, and to develop an integrated fare system can achieve a more sustainable transportation system [53]. Thus, travel factors have significant impacts on determining modes and mode shares for daily travel purposes.

From the above discussion, a number of observations can be made. Empirical studies support that the built environment has a significant impact on travel mode choice. High density, mixed uses, high street network connectivity, accessibility, and shorter travel distance significantly reduce car travel and increase walking, cycling, and public transport use. Moreover, socioeconomic, self-selection, and transport factors also significantly influence mode choice. Therefore, excluding the latter considerations may establish a spurious relationship between the built environment and travel mode choice.

## II. DATA AND STUDY METHODS
### A. DATA

Data on commuting modal split are sourced from the U.S. Census Bureau's American Community Survey (ACS) at the geographic level of census block groups in Mecklenburg County [54]. Four travel modes are considered, namely bus transit, rail transit, walking, and automobile. The latter mode encompasses both car driving and ride sharing. The streetcar mode and the pooled mode of taxicab, motorcycle, bicycle, or other means are excluded from the analysis because of their low frequency and higher margin of errors in the ACS dataset. Built environment variables, including the bus route density, rail transit proximity, street intersection density, and land-use mix using an entropy index, are calculated in ArcGIS using shapefile data collected from the Mecklenburg County GIS Center [55]. Data on job accessibility are collected from the U.S. Census Bureau's Longitudinal Employer-Household Dynamics (LEHD) data source [56]. Data for additional factors, such as commute time, population density, median age, median gross rent, and percentage of renter-occupied dwellings, are sourced from the ACS [57].

### B. STUDY AREA

The study area is Mecklenburg county, North Carolina, and the year of study is 2015. The study is carried out at the geographic level of neighborhood-representing census block groupings; there are 546 block groups in the county. As the second most populous county in NC, with a population of 1,128,945 in 2020, Mecklenburg County has a strong economic profile in comparison with the rest of the state and the nation. With a value of $65,244, its per capita personal income (PCPI) in 2020 was 129.7% of the state average and 110% of the national average. In addition to its significance to the state's economy, it has grown significantly during the last decade. Since 2010, its real gross domestic product has increased at a compounded annual rate of 3.1%, which is higher than the 1.3% and 1.6% values for the

state and for the country, respectively (Bureau of Economic Analysis, 2022). The county's rapid economic expansion increased its attractiveness and contributed to a predicted 16% increase in population between 2010 and 2021 [58].

Until fairly recently, Mecklenburg County's seat, the city of Charlotte, has been one of the conventional car-oriented monocentric American cities with urban sprawl developments [59], [60]. Due to the county's growing economy and population, strategies for integrating land use and transportation have been developed over recent decades. Examples include the 2030 Transit Corridor System Plan, which was adopted in 2006, and the 2025 Integrated Transit/Land-Use Plan, which was approved in 1998 [61]. The main objective of these plans has been to integrate the residential and economic activities embedded in land uses with transportation through 5 transit corridors and stimulate growth (see Figure 1). These growth transit corridors are planned to connect different parts of Mecklenburg County by multiple transit modes such as Bus Rapid Transit, Light Rail, street car, as well as driving. These plans have led to urban structural changes and the emergence of new secondary employment centers in different parts of the city, such as the University City area to the northeast of the CBD and the Ballantyne and South Park centers to south of the county. These changes have evolved the city from a monocentric structure towards a polycentric one. For the implementation of these transit plans in the city of Charlotte, significant investments in public transportation has taken place, such as the construction of the first phase of the Blue Line of the Lynx light rail system with a finished cost of approximately $462.7 million. This line of light rail opened in 2007, running from the city's center to its southwest through 15 stations. A more recent extension at an estimated
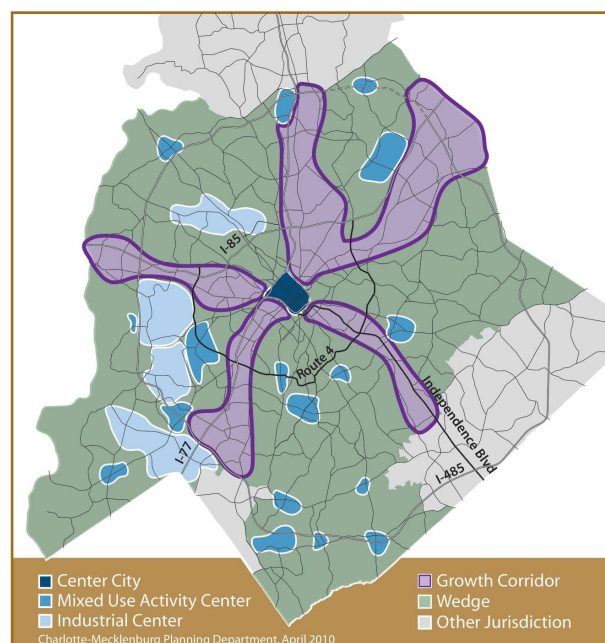


**FIGURE 1.** Transit corridors in Mecklenburg County [62].

cost of about $1 billion opened to the public in 2018 from the city center to the University City area.

As part of the new urban structural changes, the new transit corridors resulted in the growth of new compact and mixed-use developments such as transit-oriented developments (TOD) in different parts of the county, including the university area, the outer parts of the city center, along the LYNX Blue Line and in the south of the county. Mainly occurring in the proximity of transit stations, the new mixed-use developments, including office, residential, and retail spaces, are anticipated to promote the use of public transportation by improving accessibility and bringing trip ends like homes and workplaces closer together.

The year 2015 is selected for our analysis as it is about a decade after the establishment of light rail in Charlotte (LYNX Blue Line). A decade is a good time for built environment and urban structural changes to form in response to transit investments and for travel behaviors to settle into a new equilibrium. Charlotte's land development activities had also regained vigor after the sharp downturn of the Great Recession. In summary, given the new built environment and transportation features in Mecklenburg county such as the availability of multiple transit options (bus and light rail), the polycentric structure of the urban area and improved access to multiple employment centers, increased diversity of mixed use developments, enhanced walkable urban design, and denser urban form, this county is a suitable case for investigating the impact of the factors of the built environment on commuting modal choice behavior.

### C. STUDY METHODS

Data-driven methods like ML are increasingly common in a variety of fields including transportation, as a result of recent improvements in the availability of fine-grain data and of enhanced computational power. Because ML infers the mathematical functions from the data, it is less reliant on subject-matter expertise in a specific field than statistical models that are primarily grounded in theoretical underpinnings. In contrast to statistical models conventionally used in the transportation literature, ML input data require less experimental design. ML has several other advantages over conventional statistical models in several ways. It is not affected by limiting assumptions exhibited by many statistical models, such as the Independence of Irrelevant Alternatives (IIA) in multinomial logistic regression as the most common method in travel mode choice studies, and normality and linearity in other regression-based techniques. Complex phenomena, such as the nonlinear between variables in the built environment-transportation topic, can readily be studied explicitly via ML. ML has been demonstrated to have better predictive accuracy than statistical modeling and is independent of prior parameters.

Like other application domains, transportation research has increasingly turned its attention to ML to enhance its analytic performance and empirical validity [20], [21], [63];

however, the majority of this literature employs ML to solve prediction problems. Investigating the links between components and drawing causal conclusions are not given enough attention in the literature. A very limited number of studies have investigated causal inference in travel behavior studies using machine learning. For example, Zhao et al. [64] uses a number of ML models and employs the tool of feature importance to investigate the relationships between factors. However, they only investigate the feature importance and ranking of predictive factors, regardless of the direction or magnitude of the relationship.

In this study, we use Random Forest (RF) to estimate the commuting modal split in block groups of the study area, Mecklenburg County, NC. In RF modeling, based on the bagging method, an ensemble of multiple regression trees is generated in order to optimize the model performance [65]. RF is chosen as an ML predictive model because it captures the non-linear relationships between predictive factors and commuting mode shares, which is common in socio-economic and behavioral subjects. Partial Dependence Plots (2) show evidence of meaningful non-linear relationships between variables in our data. This is the case for multiple variables –both built environment measures and control variables, and in ways that defy handling via conventional mathematical transformation. RF has also proved to be very effective at modeling data and to routinely outperform other ML methods [64], [66], [67]. In our RF implementation, 70% of the data is used for training the model, and the remaining 30% is used for out-of-sample evaluation of the fitted model. 100 trees are used for the forest ensemble. The model is implemented using the random forest regressor available in the sklearn library.

In our RF model, we have four commuting modes: public transit-bus, public transit-rail, walking and car driving. All four mode shares are modeled jointly. The outcome variables are in share format to reflect the relative choices in block group neighborhoods (Table 1). Our independent variables measuring the built environment are land-use mix using an entropy index, bus route density, population density, street intersection density (as a proxy for urban block size), distance to the closest light rail station, and job accessibility. These variables are calculated with geospatial analysis software ArcGIS using shapefile data, as explained in an earlier study [10]. Since larger block groups may naturally include a larger variety of land uses, they would have higher land-use mix values. In order to control for this bias, the land-use mix values are normalized by dividing by each bloc group's land area. These built environment variables measure each of the 5 Ds of the built environment. Table 2 contains the descriptive statistics of independent variables used in this study for predicting commuting mode shares in block groups.

Once the RF model is fitted, in order to investigate the causal relationships, including the magnitude and direction of the impacts, we use shapely values, which are in the class of additive feature importance measures. Shapely values are model-agnostic, as they require the output of the
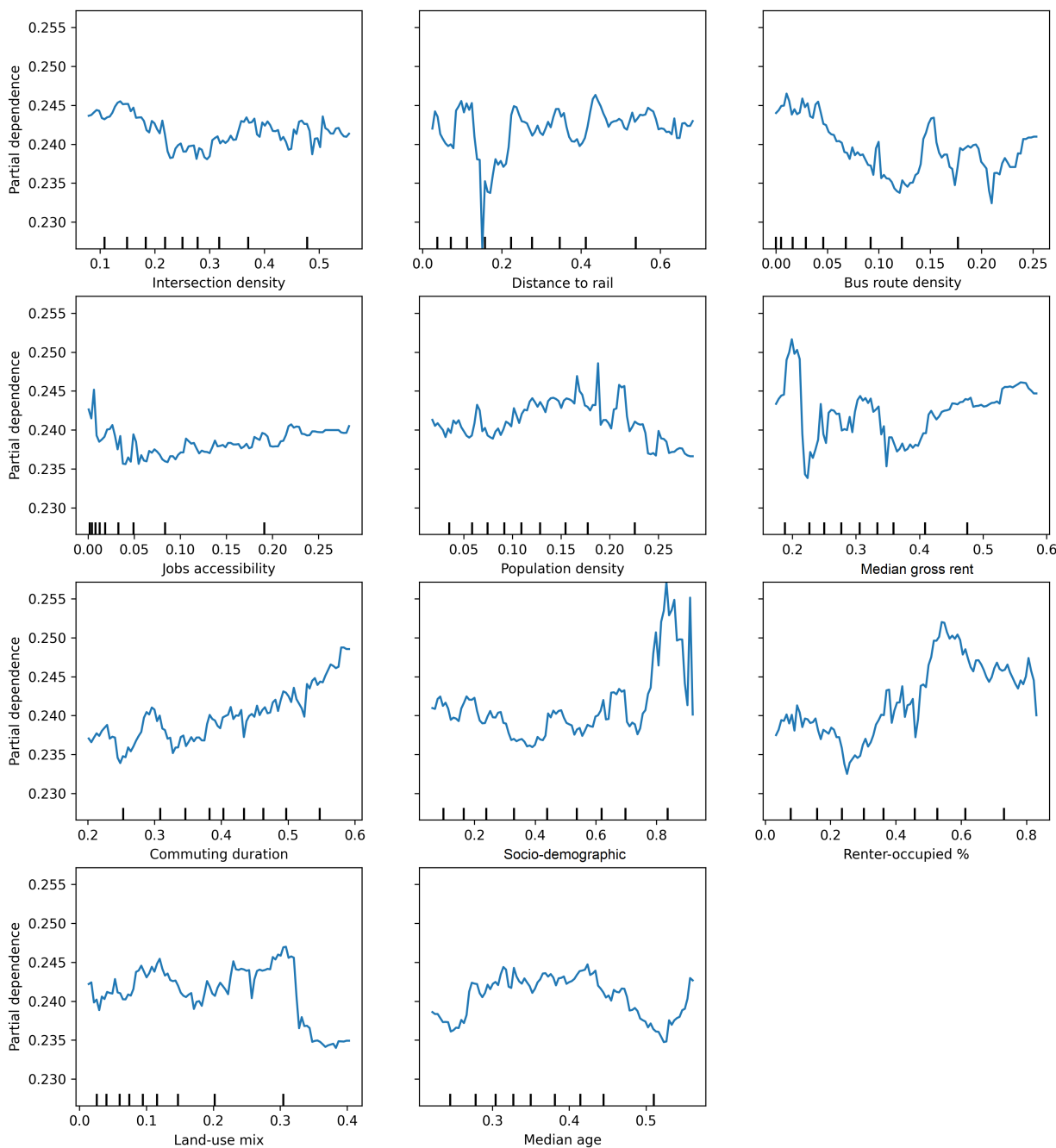
**FIGURE 2.** Partial dependence plots.

**TABLE 1.** Descriptive statistics of dependent variables (n = 546 block groups).

| Dependent Variable | Min | Mean | Standard Deviation | Max |
|---|---|---|---|---|
| **Driving (%)** | 35.11 | 93.98 | 8.55 | 100 |
| **Walking (%)** | 0 | 2.13 | 5.93 | 64.69 |
| **Public transit-bus (%)** | 0 | 3.68 | 6.11 | 36.04 |
| **Public transit-rail (%)** | 0 | 0.21 | 0.88 | 8.43 |

prediction model and observed values, regardless of the ML model type. Other studies have typically employed feature importance techniques such as the mean decrease in impurity or permutation of features [68], [69]. However, the issue with these techniques is that they calculate the feature importance for all outcome variables together, without providing insights into the impacts on each outcome in particular. Shapely values, on the other hand, convey the heterogeneity of impacts of features over different outcome variables. In addition to the importance, the shapely values show the sign of association between the features and outcome variables.

**TABLE 2.** Descriptive statistics of independent variables (n = 546 block groups).

| Independent Variable | Min | Mean | Standard Deviation | Max |
|---|---|---|---|---|
| Land-use mix (normalized entropy index) | 0 | 0.001 | 0.001 | 0.009 |
| Intersection density (units/mi2) | 9.83 | 75.00 | 37.68 | 255.31 |
| Jobs accessibility (jobs) | 0 | 772 | 1531 | 12,768 |
| Bus route density (mi/ mi2) | 0 | 3.40 | 4.57 | 47.10 |
| Population density (population/mi2) | 25.69 | 3148.89 | 2398.12 | 24951.76 |
| Distance to rail line (ft) | 0 | 27613.69 | 20838.45 | 97990.56 |
| Renter-occupied housing (%) | 0 | 38.16 | 24.68 | 100 |
| Commuting duration (mins) | 8.72 | 26.45 | 5.17 | 51.93 |
| PCA component for control | -48.76 | 0 | 30.12 | 62.39 |
| Median age | 10.8 | 36.59 | 7.61 | 80.5 |
| Median gross rent ($) | 253 | 1066.74 | 337.60 | 2719 |

**TABLE 3.** Eigenvalue and explained variance by component.

| | Eigenvalue | Variance (%) | Cumulative variance (%) |
|---|---|---|---|
| Component 1 | 2.2 | 71.9 | 71.9 |
| Component 2 | 0.6 | 19.9 | 91.9 |
| Component 3 | 0.2 | 8.1 | 100 |

For a subset $S$ from the set of all features $F$, a model is trained in the presence and absence of feature $i$ and assigns an importance value to that feature based on the effect on prediction. Training is repeated on all possible subsets $S$, and the average shapely values for each feature are calculated as follows [70]:

$$\Phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)]$$

where the term in the bracket represents the comparison between the prediction for the current feature subset $S$ in the presence and absence of feature $i$. [70] further proposes a unified framework, SHAP (SHapley Additive exPlanations), to calculate shapely values. SHAP values are shapely values of a conditional expectation function of the original model that measure the change in the expected model prediction for a specific feature. Larger SHAP values are indicative of a stronger impact of the feature. Based on Shapley values from game theory, this method provides both global and local explanations of the feature impacts. The local interpretability feature of SHAP provides the feature importance for each individual prediction. The feature importance is implemented here using the SHAP library.

Figure 3 illustrates our conceptual research design and methodology in four major steps and multiple substeps. Grounded in the extant literature on 5D principles and commuting modal choice, we identify relevant measures of the built environment. Next, the outcome variables and control factors are selected and organized in the block group dataset. Third, the RF ensemble model is trained and tested. Lastly, SHAP values are calculated and discussed.

## III. RESULTS

We predict the shares of four transportation modes using the RF method. Table 4 shows the model's predictive performance, with a Mean Absolute Error (MAE) of 0.02 and a Root Mean Squared Error (RMSE) of 0.04.

After fitting the RF model and obtaining predictions, we compute the feature importance using the SHAP library [70]. Figure 4 shows the local impacts of features for each transportation mode share. In these figures, each point is an instance (block group), and its color shows the value of that instance. The horizontal axis indicates the SHAP values of instances for each feature. For example, for the feature of median gross rent in Figure 4a, low values (blue dots) are associated with higher predicted values for the bus mode share (negative contribution to the model predictions). As another example, for the feature of renter-occupied housing percentage, high values (red dots) are associated with higher predicted values for the bus mode share (positive contribution to the model predictions). Moreover, features are sorted on the vertical axis based on their global importance (high to low from the top down). SHAP enhances the transparency of ML models by providing interpretability. It is also worth noting that it shows how each observation (block group in this study) contributes to the model predictions and reveals the heterogeneity of these impacts, which may be of great value for specific purposes. For example, in figure 4a socio-demographic factors and commuting duration have more heterogeneous impact on bus commuting share than renter-occupied housing percentage. In this figure, we can infer more homogeneous impact of renter-occupied housing percentage as points are located closer together. In addition, this plot indicates that two block groups with low commuting duration values are strongly associated with lower bus commuting share. These two block groups have the strongest contribution to the positive relationships between the commuting duration and bus commuting share, among all other block groups with short commuting duration.

In addition to the local interpretations, we can look into the average SHAP values to learn about the overall impact of each feature, which is the main purpose of this study. Figure 5 shows the global impact of features by the four commuting modes. In these figures, the red color depicts the positive impact and the blue color depicts the negative impact. The horizontal line (SHAP values) indicates the magnitude of impacts.
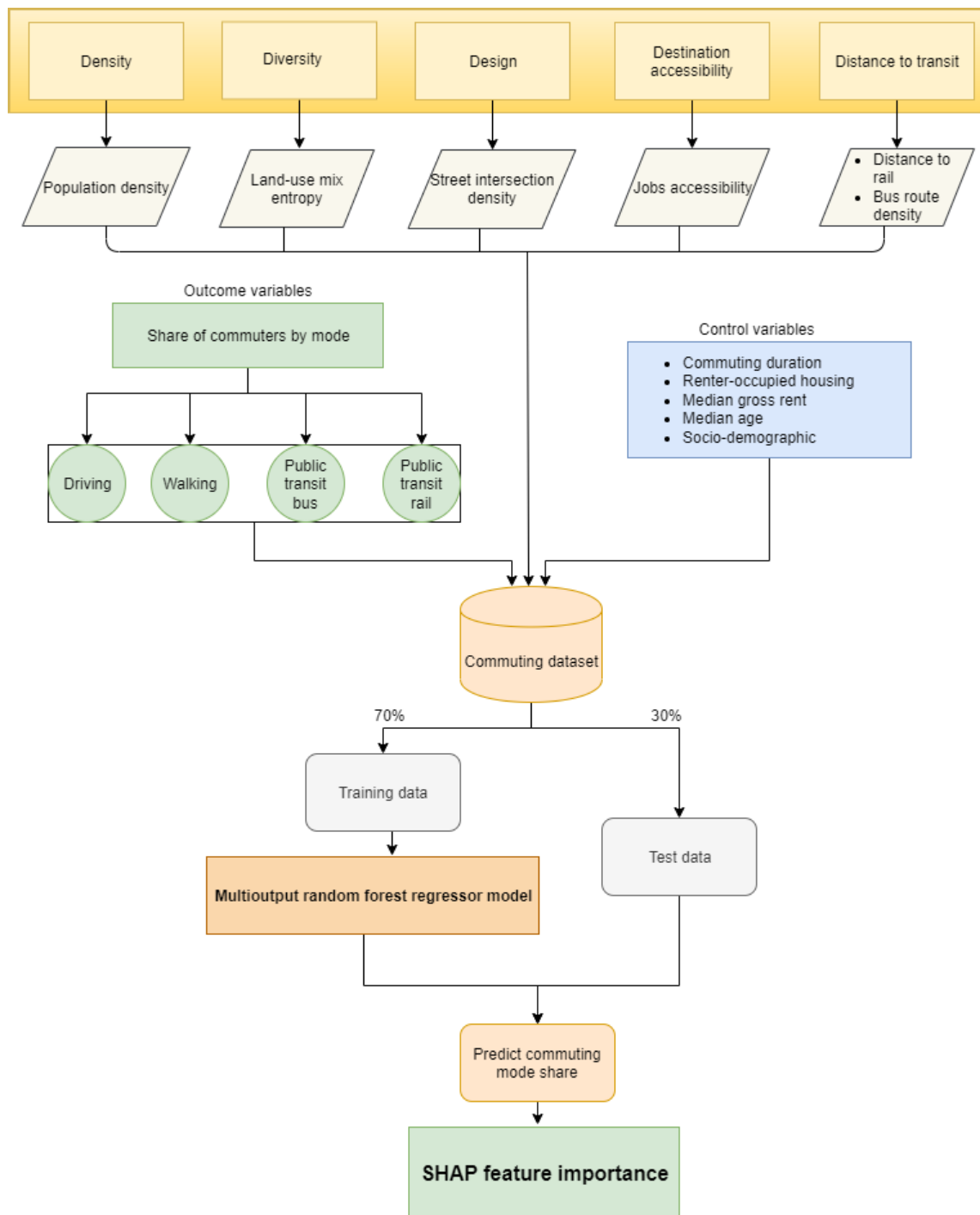
**FIGURE 3.** Research design flowchart.

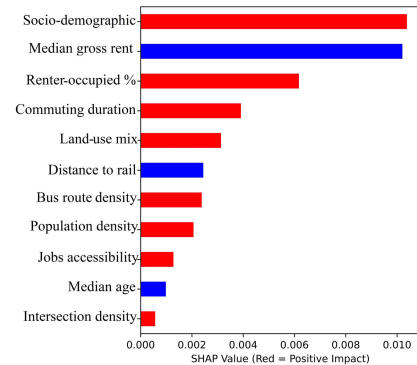**TABLE 4.** Random forest model evaluation metrics.

| MAE | MSE | RMSE | Explained Variance Score | Mean Squared Log Error | Median Absolute Error | R^2 |
|---|---|---|---|---|---|---|
| 0.02378 | 0.001927 | 0.03895 | 0.3301 | 0.000982 | 0.01502 | 0.3193 |

The impact of the built environment on commuters' share of bus transit mode is depicted in Figure 5a. Block groups with a higher land-use 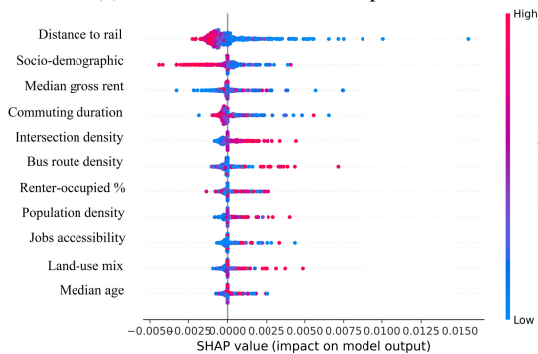mix, denser bus routes, a higher population density, higher job accessibility, and higher intersection density, greater proximity to rail transit, exhibit a larger share of the bus mode. In terms of the control variables,
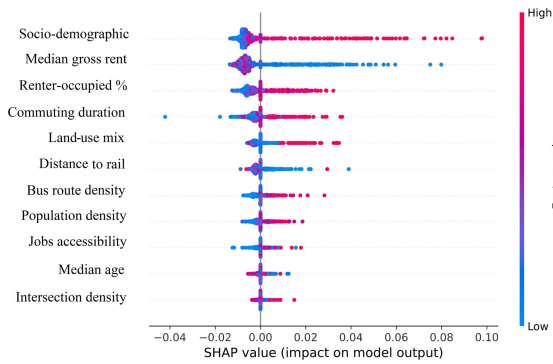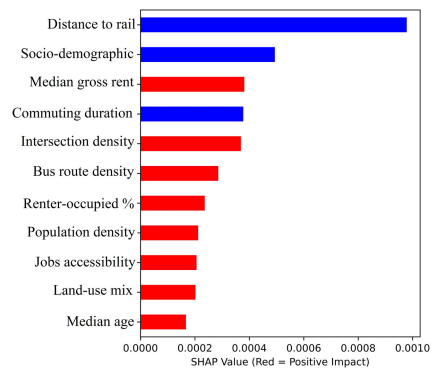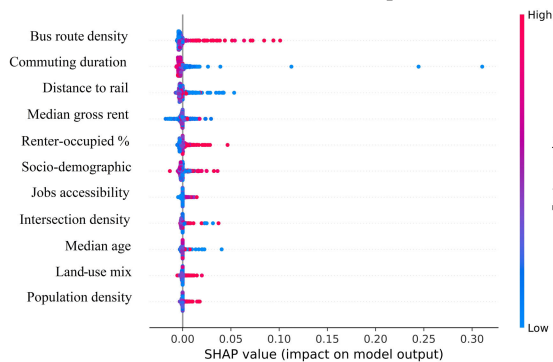
(a) Bus transit mode local explanations.



(b) Rail transit mode local explanations.
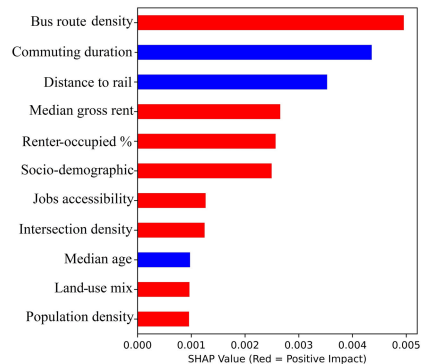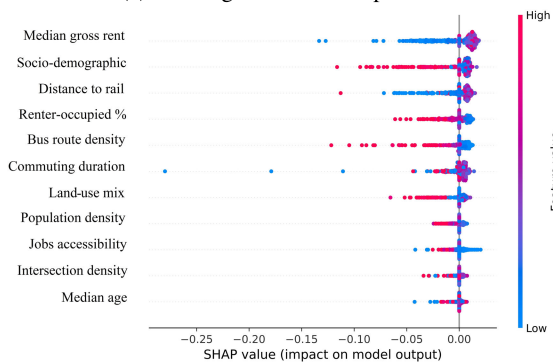


(c) Walking mode local explanations.



(d) Driving mode local explanations.

**FIGURE 4.** Importance of factors to influence commuting mode shares.



(a) Bus transit mode feature importance.



(b) Rail transit mode feature importance.



(c) Walking mode feature importance.
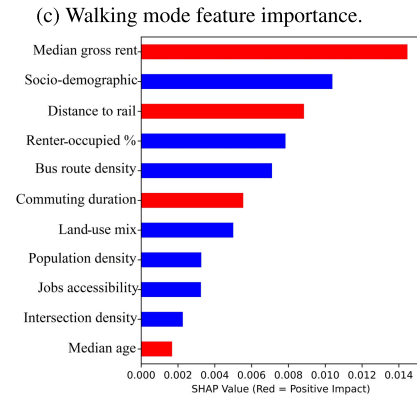


(d) Driving mode feature importance.

**FIGURE 5.** Feature importance to determine commuting mode shares.

lower median gross rent is associated with a higher share of bus mode. Additionally, block groups with a higher proportion of renter-occupied homes have a higher share of the

bus mode. The graph also demonstrates that commuters who choose the bus have longer commutes. Furthermore, among

the 5D variables, the land-use mix, rail transit proximity, and bus route density play the most important roles in bus ridership. However, socio-demographic and economic factors are more influential on the bus ridership share than the built environment factors.

Figure 5b illustrates how the built environment affects the share of commuters' use of the rail transit option. Block groups with greater proximity to light rail, street intersection density, population density, bus route density, job accessibility, and land-use diversity have higher shares of light rail mode choice. Block groups with higher proportions of renter-occupied housing and higher median gross rents demonstrate larger shares of rail transport mode choice, according to the average SHAP values. Also, a higher share of light rail transit is linked to lower commute times. The distance to rail variable is by far the most influential factor on the rail commuters share, followed by socio-demographic factor, median gross rent, and commuting duration.

The average SHAP values for the walk commuting mode (Figure 5c) demonstrate that walking is more prevalent in locations with higher densities of bus routes, higher degrees of jobs accessibility, smaller street blocks, higher land-use mixtures, higher population densities, and closer proximity to light rail. Additionally, just like with the light rail mode, more people walk to work in locations with higher proportions of renter-occupied housing and higher median gross rents. Again, similar to the light rail option, commuting duration is negatively associated with the walking mode share. The density of bus routes has the largest impact on the share of walkers, followed by the commuting duration and rail transit proximity.

Driving (Figure 5d), on the other hand, exhibits contrasting patterns from the three forms of transportation that have been discussed so far—bus, rail, and walking. Block groups located further from light rail stations, with lower densities of bus routes, less diverse mix of land uses, lower population densities, lower jobs accessibility, and lower intersection densities, and showing larger block sizes have larger shares of driving mode choice. Again, contrary to the previous sustainable modes of public transit and walking, areas with a lower percentage of renter-occupied housing are associated with a higher driving mode choice. The median gross rent variable has a positive relationship with driving mode share, which shows that areas with higher rents have more driving share. Lastly, longer commutes are associated with a higher driving mode share. Among all the variables, the socio-demographic factors and median gross rent have the highest impact on the prevalence of driving.

## IV. DISCUSSION AND CONCLUSION

Given the persistent and entrenched car dependency in the United States and its ensuing problems, such as social exclusion and disparities, high cost of urban infrastructures, and environmental issues, it is critical to investigate the travel mode choice behavior of urban residents. The built environment is expected to impact travel behavior. Specifically,

built environments with higher degrees of the 5Ds including density, diversity, design, destination accessibility, and short distance to transit are hypothesized to encourage people to choose more sustainable modes (i.e., public transit, especially rail transit, and walking). Given the importance of these relationships to policy making, in this study, we reassessed the impact of the built environment variables on commuting mode shares apprehended via the four modes of bus, rail, walking and driving. The impact was evaluated in Mecklenburg county, NC, in 2015, at the geographic granularity of block groups. While the majority of the literature has studied a subset of the 5Ds only, we investigated the relationships between all the 5D properties and commuting behavior in four transportation modes, all in one unified model.

ML models are capable of detecting complexities embedded in the data drawn from real-world events and decisions, contrary to linear econometric models. Given the complexities of the relationships between the built environment, travel behavior, and socio-demographic characteristics, such as non-linearities, a multi-output RF model was employed to model the relationships. Following the fitting of the RF model, feature importance was implemented to investigate the relationships between built environment variables and mode shares. The SHAP (SHapley Additive exPlanations) method was used to assess these relationships. Contrary to other feature importance methods that only provide information on the ranking or the score of importance of each feature, SHAP values give insights into both the magnitude and the direction of relationships.

Our main results can be summarized as follows. First, as hypothesized in this study, areas with more diverse land use have higher shares of commuting by bus, rail, and walking. On the contrary, the negative sign of the land-use mix variable for the driving mode indicates that areas with a lower land-use mix have higher shares of driving to work. Findings of this study are in line with previous studies where researchers have demonstrated that mixed land uses increase trips by active and public transportation, while decreasing trips by automobiles [25], [28]. Land-use diversity encourages non-motorized travel compared to motorized travel by placing activity locations close to each other and reducing travel distance and time [34], [71]. Second, similar to the diversity dimension, the urban feature of population density shows positive impacts on choices of bus, rail, and walking, while more compact areas have a lower share of driving to work. Thus, a higher population density increases travel by active and public transportation, which aligns well with the findings reported in the extant literature [27], [72], [73].

Third, as a proxy for the design dimension, the network connectivity variable shows the impact in the same direction. This study reiterates the findings from the extant literature where researchers found that high transport network connectivity due to smaller block size significantly increases walking, cycling, and public transport and decreases car use by proving a better connection between origin and destinations [38], [41], [74], [75].

As for the fourth dimension –destination accessibility, we can see that areas with better job accessibility have higher shares of bus, rail, and walking and lower shares of driving commute to work, which complies with previous studies [34]. Finally, the last built environment variable –distance to transit, indicates that as we get closer to the light rail transit corridor and their stations, people are more likely to take the bus, rail, and to walk more, and also drive less. Also, areas with a higher bus route density have larger shares of bus, rail, and walking, and have smaller shares of driving to work. Thus, expanding the supply of public transit throughout the city syphons demand away from driving and towards public transit and other modes used as a first mile/last mile solution.

Moreover, as the feature rankings indicate, rail transit proximity and design (street intersection density) are the most important built environment characteristics for rail mode choice. Density and rail transit proximity are the most influential variables for the walk mode. Similarly, these two variables of density and rail transit proximity have large effects on the driving mode choice. However, for the bus mode choice, socioeconomic features such as median gross rent, renter-occupied housing percentage, and the socio-demographic component (white percentage, educational attainment, and car ownership) used as a control in the RF model are more influential than any of the built environment factors.

Our findings provide robust evidence to suggest urban planners and urban designers to invest in more compact, mixed-use developments in the proximity of transit stations to encourage travel behaviors in line with sustainability objectives. As shown in Figure 5d, among the 5 Ds factors, transit proximity has the greatest impact on reducing driving. Therefore, investing in a sustained fashion in public transit systems to bring these modes closer to the urban populations of transit riders (both in the form of bus services and rail stations) will be effective at bringing people to drive less and take more sustainable modes. Given the anticipated population [76] and urbanization growth [77] and the greater needs for sustainable transportation, policy makers and decision makers can take into account this consideration and develop action plans to this end. Urban planning and design strategies in line with the other Ds will reinforce the effectiveness of such intervention in the mid-to-long term.

With regard to the complexities of the topic treated here and to the limitations of our work, we propose several considerations for extending our work in the future. The availability of disaggregated data including individuals' travel preferences and attitudes may help us investigate the heterogeneity of relationships and the selection bias issue that often permeate the urban transportation and land use nexus. Second, since work travel and non-work travel are known to exhibit very different patterns, it would be useful to also study non-work travel, in addition to work travel. Moreover, considering other modes of sustainable and active transportation such as bicycling and micromobility services would be informative to sustainable urban planning and design, provided that data availability can be secured. Multimodal options should also

be weaved through the design of the modal choice sets of residents as they represent a meaningful share of trips in larger and complex urban environments.

## REFERENCES

[1] S. Zhong and D. Sun, "Taxi hailing choice behavior and economic benefit analysis of emission reduction based on multi-mode travel big data," in *Logic-Driven Traffic Big Data Analytics*. Singapore: Springer, 2022, doi: 10.1007/978-981-16-8016-8_11.

[2] H.-J. Huang, T. Xia, Q. Tian, T.-L. Liu, C. Wang, and D. Li, "Transportation issues in developing China's urban agglomerations," *Transp. Policy*, vol. 85, pp. A1–A22, Jan. 2020.

[3] (2018). U.S. Department of Transportation. *2017 National Household Travel Survey*. [Online]. Available: https://nhts.ornl.gov/person-trips

[4] G. Saiyad, M. Srivastava, and D. Rathwa, "Exploring determinants of feeder mode choice behavior using artificial neural network: Evidences from Delhi metro," *Phys. A, Stat. Mech. Appl.*, vol. 598, Jul. 2022, Art. no. 127363.

[5] (2022). U. S. Department of Transportation. *Strategic Plan FY 2022–2026*. [Online]. Available: https://www.transportation.gov/mission/us-dot-strategic-plan-fy-2022-2026

[6] R. Cervero and K. Kockelman, "Travel demand and the 3Ds: Density, diversity, and design," *Transp. Res. Part D: Transp. Environ.*, vol. 2, no. 3, pp. 199–219, Sep. 1997.

[7] R. Ewing and R. Cervero, "Travel and the built environment: A synthesis," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1780, no. 1, pp. 87–114, Jan. 2001.

[8] S. Handy, "Regional versus local accessibility: Implications for nonwork travel," Univ. California Transp. Center, Univ. California Berkeley, Berkeley, CA, USA, UCTC Rep. 234, 1993.

[9] M. R. Stevens, "Does compact development make people drive less?" *J. Amer. Planning Assoc.*, vol. 83, no. 1, pp. 7–18, Jan. 2017.

[10] F. Hatami and J.-C. Thill, "Spatiotemporal evaluation of the built Environment's impact on commuting duration," *Sustainability*, vol. 14, no. 12, p. 7179, Jun. 2022.

[11] M. M. Rahman, S. Upaul, J.-C. Thill, and M. Rahman, "Active transportation and the built environment of a mid-size global south city," *Sustain. Cities Soc.*, vol. 89, Feb. 2023, Art. no. 104329.

[12] C. Ding, Y. Wang, T. Tang, S. Mishra, and C. Liu, "Joint analysis of the spatial impacts of built environment on car ownership and travel mode choice," *Transp. Res. D, Transp. Environ.*, vol. 60, pp. 28–40, May 2018.

[13] Y. Hu, A. Sobhani, and D. Ettema, "Exploring commute mode choice in dual-earner households in a small Chinese city," *Transp. Res. D, Transp. Environ.*, vol. 102, Jan. 2022, Art. no. 103148.

[14] T.-H.-T. Gim, "How does the length of residence in a neighborhood vary the effects of neighborhood land use on commuting trip time and mode choice?" *Ann. Regional Sci.*, vol. 68, no. 1, pp. 95–123, Feb. 2022.

[15] Y. Lu, G. Sun, Z. Gou, Y. Liu, and X. Zhang, "A dose–response effect between built environment characteristics and transport walking for youths," *J. Transp. Health*, vol. 14, Sep. 2019, Art. no. 100616.

[16] M. M. Rahman, K. C. Paul, M. A. Hossain, G. G. M. N. Ali, M. S. Rahman, and J.-C. Thill, "Machine learning on the COVID-19 pandemic, human mobility and air quality: A review," *IEEE Access*, vol. 9, pp. 72420–72450, 2021.

[17] A. Y. Sun and B. R. Scanlon, "How can big data and machine learning benefit environment and water management: A survey of methods, applications, and future directions," *Environ. Res. Lett.*, vol. 14, no. 7, Jul. 2019, Art. no. 073001.

[18] W. Wang, Y. Zhang, C. Zhao, X. Liu, X. Chen, C. Li, T. Wang, J. Wu, and L. Wang, "Nonlinear associations of the built environment with cycling frequency among older adults in Zhongshan, China," *Int. J. Environ. Res. Public Health*, vol. 18, no. 20, p. 10723, Oct. 2021.

[19] J. Liu, B. Wang, and L. Xiao, "Non-linear associations between built environment and active travel for working and shopping: An extreme gradient boosting approach," *J. Transp. Geography*, vol. 92, Apr. 2021, Art. no. 103034.

[20] A. N. Koushik, M. Manoj, and N. Nezamuddin, "Machine learning applications in activity-travel behaviour research: A review," *Transp. Rev.*, vol. 40, no. 3, pp. 288–311, May 2020.

[21] J. Hagenauer and M. Helbich, "A comparative study of machine learning classifiers for modeling travel mode choice," *Expert Syst. Appl.*, vol. 78, pp. 273–282, Dec. 2017.

[22] X. Chang, J. Wu, H. Liu, X. Yan, H. Sun, and Y. Qu, "Travel mode choice: A data fusion model using machine learning methods and evidence from travel diary survey data," *Transportmetrica A, Transp. Sci.*, vol. 15, no. 2, pp. 1587–1612, Nov. 2019.

[23] R. Cervero, "Built environments and mode choice: Toward a normative framework," *Transp. Res. D, Transp. Environ.*, vol. 7, no. 4, pp. 265–284, Jun. 2002.

[24] Y. Chen, Y. Chen, W. Tu, and X. Zeng, "Is eye-level greening associated with the use of dockless shared bicycles?" *Urban Forestry Urban Greening*, vol. 51, May 2020, Art. no. 126690.

[25] K. Park, R. Ewing, B. C. Scheer, and G. Tian, "The impacts of built environment characteristics of rail station areas on household travel behavior," *Cities*, vol. 74, pp. 277–283, Apr. 2018.

[26] R. Cervero and C. Radisch, "Travel choices in pedestrian versus automobile oriented neighborhoods," *Transp. Policy*, vol. 3, no. 3, pp. 127–141, Jul. 1996.

[27] E. L. Glaeser and M. E. Kahn, "The greenness of cities: Carbon dioxide emissions and urban development," *J. Urban Econ.*, vol. 67, no. 3, pp. 404–418, May 2010.

[28] R. Wang, Y. Lu, X. Wu, Y. Liu, and Y. Yao, "Relationship between eye-level greenness and cycling frequency around metro stations in Shenzhen, China: A big data approach," *Sustain. Cities Soc.*, vol. 59, Aug. 2020, Art. no. 102201.

[29] S. Ahmad and J. A. Puppim de Oliveira, "Determinants of urban mobility in India: Lessons for promoting sustainable and inclusive urban transportation in developing countries," *Transp. Policy*, vol. 50, pp. 106–114, Aug. 2016.

[30] M. McKibbin, "The influence of the built environment on mode choice–evidence from the journey to work in Sydney," in *Proc. 34th Australas. Transp. Res. Forum (ATRF)*, Adelaide, SA, Australia, 2011, pp. 1–12.

[31] L. D. Frank and G. Pivo, "Impacts of mixed use and density on utilization of three modes of travel: Single-occupant vehicle, transit, and walking," *Transp. Res. Rec.*, vol. 1466, pp. 44–52, Jan. 1994.

[32] R. Crane and R. Crepeau, "Does neighborhood design influence travel?: A behavioral analysis of travel diary and GIS data," *Transp. Res. D, Transp. Environ.*, vol. 3, no. 4, pp. 225–238, Jul. 1998.

[33] R. Ewing and R. Cervero, "Travel and the built environment: A meta-analysis," *J. Amer. Planning Assoc.*, vol. 76, no. 3, pp. 265–294, Jun. 2010.

[34] P. Zhao, "The impact of the built environment on individual workers' commuting behavior in Beijing," *Int. J. Sustain. Transp.*, vol. 7, no. 5, pp. 389–415, Sep. 2013.

[35] Y. Lu, Y. Yang, G. Sun, and Z. Gou, "Associations between overhead-view and eye-level urban greenness and cycling behaviors," *Cities*, vol. 88, pp. 10–18, May 2019.

[36] E. Spissu, A. R. Pinjari, R. M. Pendyala, and C. R. Bhat, "A copula-based joint multinomial discrete–continuous model of vehicle type choice and miles of travel," *Transportation*, vol. 36, no. 4, pp. 403–422, Jul. 2009.

[37] T. E. Mcmillan, "The relative influence of urban form on a child's travel mode to school," *Transp. Res. A, Policy Pract.*, vol. 41, no. 1, pp. 69–79, Jan. 2007.

[38] C. Zhao, T. A. S. Nielsen, A. S. Olafsson, T. A. Carstensen, and X. Meng, "Urban form, demographic and socio-economic correlates of walking, cycling, and e-biking: Evidence from eight neighborhoods in Beijing," *Transp. Policy*, vol. 64, pp. 102–112, May 2018.

[39] K. Larsen, J. Gilliland, P. Hess, P. Tucker, J. Irwin, and M. He, "The influence of the physical environment and sociodemographic characteristics on children's mode of travel to and from school," *Amer. J. Public Health*, vol. 99, no. 3, pp. 520–526, Mar. 2009.

[40] N. R. den Braver, J. G. Kok, J. D. Mackenbach, H. Rutter, J.-M. Oppert, S. Compernolle, J. W. R. Twisk, J. Brug, J. W. J. Beulens, and J. Lakerveld, "Neighbourhood drivability: Environmental and individual characteristics associated with car use across Europe," *Int. J. Behav. Nutrition Phys. Activity*, vol. 17, no. 1, Dec. 2020.

[41] J. Ko, S. Lee, and M. Byun, "Exploring factors associated with commute mode choice: An application of city-level general social survey data," *Transp. Policy*, vol. 75, pp. 36–46, Mar. 2019.

[42] M. Nawrath, I. Kowarik, and L. K. Fischer, "The influence of green streets on cycling behavior in European cities," *Landscape Urban Planning*, vol. 190, Oct. 2019, Art. no. 103598.

[43] J. De Vos and F. Witlox, "Do people live in urban neighbourhoods because they do not like to travel? Analysing an alternative residential self-selection hypothesis," *Travel Behaviour Soc.*, vol. 4, pp. 29–39, May 2016.

[44] E. Mirzaei, R. Kheyroddin, M. Behzadfar, and D. Mignot, "Utilitarian and hedonic walking: Examining the impact of the built environment on walking behavior," *Eur. Transp. Res. Rev.*, vol. 10, no. 2, pp. 1–14, Jun. 2018.

[45] A. M. Larranaga, J. Arellana, L. I. Rizzi, O. Strambi, and H. B. B. Cybis, "Using best–worst scaling to identify barriers to walkability: A study of porto alegre, Brazil," *Transportation*, vol. 46, no. 6, pp. 2347–2379, Dec. 2019.

[46] P. P. Sarkar and C. Mallikarjuna, "Effect of perception and attitudinal variables on mode choice behavior: A case study of Indian city, Agartala," *Travel Behaviour Soc.*, vol. 12, pp. 108–114, Jul. 2018.

[47] C.-F. Chen and W.-T. Lai, "The effects of rational and habitual factors on mode choice behaviors in a motorcycle-dependent region: Evidence from Taiwan," *Transp. Policy*, vol. 18, no. 5, pp. 711–718, Sep. 2011.

[48] N. Eluru, V. Chakour, and A. M. El-Geneidy, "Travel mode choice and transit route choice behavior in montreal: Insights from McGill university members commute patterns," *Public Transp.*, vol. 4, no. 2, pp. 129–149, Nov. 2012.

[49] F. Wang and C. L. Ross, "Machine learning travel mode choices: Comparing the performance of an extreme gradient boosting model with a multinomial logit model," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2672, no. 47, pp. 35–45, Dec. 2018.

[50] R. Cervero, "Transit-oriented development's ridership bonus: A product of self-selection and public policies," *Environ. Planning A, Economy Space*, vol. 39, no. 9, pp. 2068–2085, Sep. 2007.

[51] W. Li and M. Kamargianni, "Providing quantified evidence to policy makers for promoting bike-sharing in heavily air-polluted cities: A mode choice model and policy simulation for Taiyuan-China," *Transp. Res. A, Policy Pract.*, vol. 111, pp. 277–291, May 2018.

[52] J. M. Becerra, R. S. Reis, L. D. Frank, F. A. Ramirez-Marrero, B. Welle, E. A. Cordero, F. M. Paz, C. Crespo, V. Dujon, J. Jacoby, J. Dill, L. Weigand, and C. M. Padin, "Transport and health: A look at three Latin American cities," *Cadernos de Saude Publica*, vol. 29, no. 4, pp. 654–666, Apr. 2013.

[53] Y. Han, W. Li, S. Wei, and T. Zhang, "Research on passenger's travel mode choice behavior waiting at bus station based on SEM-logit integration model," *Sustainability*, vol. 10, no. 6, p. 1996, Jun. 2018.

[54] (2019). U.S. Census Bureau. *Mean of Transportation to Work*. [Online]. Available: https://data.census.gov/cedsci/

[55] (2020). Mecklenburg County GIS Center. *GIS Data Center*. [Online]. Available: https://www.mecknc.gov/LUESA/GIS/Pages/GIS-Data-Center.aspx

[56] (2020). LEHD. *LEHD Origin-Destination Employment Statistics Data*. [Online]. Available: https://lehd.ces.census.gov/data/

[57] (2020). American Community Survey. *American Community Survey Data*. [Online]. Available: https://data.census.gov/cedsci/

[58] (2021). United States Census Bureau. *Annual Resident Population Estimates for States and Counties: April 1, 2010 to July 1, 2019; April 1, 2020; and July 1, 2020 (CO-EST2020)*. [Online]. Available: https://www.census.gov/programs-surveys/popest/technical-documentation/research/evaluation-estimates/2020-evaluation-estimates/2010s-counties-total.html

[59] B. Wilson and Y. Song, "Comparing apples with apples: How different are recent residential development patterns in Portland and charlotte?" *J. Urbanism, Int. Res. Placemaking Urban Sustainability*, vol. 2, no. 1, pp. 51–74, Mar. 2009.

[60] E. Delmelle, Y. Zhou, and J.-C. Thill, "Densification without growth management? Evidence from local land development and housing trends in Charlotte, North Carolina, USA," *Sustainability*, vol. 6, no. 6, pp. 3975–3990, Jun. 2014.

[61] (2019). City of Charlotte. *2030 Transit Corridor System Plan*. [Online]. Available: https://charlottenc.gov/cats/transit-planning/2030-plan/Pages/default.aspx

[62] (2010). Charlotte Department of Transportation. *Centers Corridors and Wedges*. [Online]. Available: https://shorturl.at/fiPR0

[63] H. Omrani, "Predicting travel mode of individuals by machine learning," *Transp. Res. Proc.*, vol. 10, pp. 840–849, Jan. 2015.

[64] X. Zhao, X. Yan, A. Yu, and P. Van Hentenryck, "Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models," *Travel Behav. Soc.*, vol. 20, pp. 22–35, Jul. 2020.

[65] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[66] L. Cheng, X. Chen, J. D. Vos, X. Lai, and F. Witlox, "Applying a random forest method approach to model travel mode choice behavior," *Travel Behav. Soc.*, vol. 14, pp. 1–10, Jan. 2019.

[67] C. R. Sekhar, Minal, and E. Madhu, "Mode choice analysis using random Forrest decision trees," *Transp. Res. Proc.*, vol. 17, pp. 644–652, Jan. 2016.

[68] N. Huang, G. Lu, and D. Xu, "A permutation importance-based feature selection method for short-term electricity load forecasting using random forest," *Energies*, vol. 9, no. 10, p. 767, Sep. 2016.

[69] J. Ou, J. Xia, Y.-J. Wu, and W. Rao, "Short-term traffic flow forecasting for urban roads using data-driven feature selection strategy and bias-corrected random forests," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2645, no. 1, pp. 157–167, Jan. 2017.

[70] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–10.

[71] A. R. Pinjari, N. Eluru, C. R. Bhat, R. M. Pendyala, and E. Spissu, "Joint model of choice of residential neighborhood and bicycle ownership: Accounting for self-selection and unobserved heterogeneity," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 2082, no. 1, pp. 17–26, Jan. 2008.

[72] J. Humphreys and A. Ahern, "Is travel based residential self-selection a significant influence in modal choice and household location decisions?" *Transp. Policy*, vol. 75, pp. 150–160, Mar. 2019.

[73] M. M. Rahman, P. Najaf, M. G. Fields, and J.-C. Thill, "Traffic congestion and its urban scale factors: Empirical evidence from American urban areas," *Int. J. Sustain. Transp.*, vol. 16, no. 5, pp. 1–16, 2021.

[74] M. Woldeamanuel and A. Kent, "Measuring walk access to transit in terms of sidewalk availability, quality, and connectivity," *J. Urban Planning Develop.*, vol. 142, no. 2, Jun. 2016, Art. no. 04015019.

[75] A. K. Porter, H. W. Kohl, A. Pérez, B. Reininger, K. P. Gabriel, and D. Salvo, "Bikeability: Assessing the objectively measured environment in relation to recreation and transportation bicycling," *Environ. Behav.*, vol. 52, no. 8, pp. 861–894, Oct. 2020.

[76] (2018). U. S. Census Bureau. *Demographic Turning Points for the United States*. [Online]. Available: https://www.shorturl.at/NRUWX

[77] (2018). United Nations. *2018 Revision of World Urbanization Prospects*. [Online]. Available: https://www.un.org/development/desa/publications/2018-revision-of-world-urbanization-prospects.html

**MD. MOKHLESUR RAHMAN** received the Bachelor of Urban and Regional Planning (B.URP) degree from the Department of Urban and Regional Planning (URP), Bangladesh University of Engineering and Technology (BUET), in 2009, the M.Sc. degree in urban planning from the Department of Urban Planning and Design (DUPAD), The University of Hong Kong (HKU), Hong Kong, in 2013, to compete in the professional arena and acquire advanced knowledge, and the Ph.D. degree in infrastructure and environmental systems (INES) from the University of North Carolina at Charlotte, in 2022. After graduation, he started working with the Khulna University of Engineering and Technology (KUET) as a Lecturer, in December 2014, and was promoted to an Assistant Professor, in February 2017. Currently, he is a Postdoctoral Research Scholar with the School of Information Studies, Syracuse University, NY, USA. His research interests include transportation planning and engineering, travel behaviors and demand modeling, land use modeling, climate change and disaster management, application of GIS in transportation and environment, econometric modeling, and data analytics.

**BEHNAM NIKPARVAR** is currently a Geospatial Data Scientist. He leverages remotely sensed data, location-based data, and spatiotemporal modeling and analyses to understand the mobility and displacement of human populations. He has worked on population movement following natural and man-made disasters. His Ph.D. research focused on the spatiotemporal modeling of disease spread through micromobility transportation systems. He is also interested in machine learning of data with geospatial properties. He is also working on the links between population mobility and the transmission of respiratory viruses.

**JEAN-CLAUDE THILL** received the Graduate degrees from the Catholic University of Louvain, Belgium. He is currently a Knight Foundation Distinguished Professor of Public Policy with the Department of Geography and Earth Sciences, University of North Carolina at Charlotte. He is also an Affiliate Faculty of the School of Data Science. He is also an Urban Economic Geographer, whose research focuses on urban and regional transportation and mobility issues, processes of urbanization, regional science, and geospatial data science. His recent research interests include modeling past urbanization trajectories of the charlotte metropolitan area (land use transformation and local and regional drifts in neighborhood quality of life), livability, innovation, and mobility in Charlotte, other U.S. cities, and Chinese cities, and accessibility to public services. He has more than 30 years of teaching and research experience in transportation planning and spatial modeling. His current and recent grants have been awarded from the National Science Foundation, Department of Defense, Department of Energy, North Carolina Department of Transportation, World Bank, and more. He was a 2012 recipient of the Edward L. Ullman Award for Significant Contributions to Transportation Geography, Awarded by the American Association of Geographers. He also received the Hirotada Kohno Award for Outstanding Service from the Regional Science Association International, in 2012. Since 2013, he has been a fellow of the Regional Science Association International. He is honored to have been appointed as a High-End Foreign Expert with the School of Public Administration, Xi'an Jiaotong University (Ministry of Science and Technology), Xi'an, China.

**FAIZEH HATAMI** is currently pursuing the Ph.D. degree in geography with the University of North Carolina at Charlotte. She is also a Graduate Research Assistant with the University of North Carolina at Charlotte. Integrating new data-oriented approaches with spatial analysis and modeling, her research is focused on causal inference and predictions in a variety of fields. In addition to the spatial properties, she also incorporates temporal dimension in her research to reveal patterns and processes and provide forecasts over time and space. She mainly conducts research in transportation-land use, transportation-public health, biostatistics, and economic development research areas.

• • •