Journal of Transport Geography 116 (2024) 103828

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Built environment influences commute mode choice in a global south megacity context: Insights from explainable machine learning approach

F.R. Ashik [a,f], A.I.Z. Sreezon [b], M.H. Rahman [c], N.M. Zafri [d], S.M. Labib [e,*]

[a] Department of Geography, McGill University, 805 Sherbrooke St W, Montreal, Quebec H3A 0B9, Canada
[b] School of Computer Science, Queensland University of Technology (QUT), Australia
[c] Community and Regional Planning, The University of Texas at Austin, 310 Inner Campus Drive B7500, Austin, TX 78705, United States
[d] Department of Urban and Regional Planning, Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh
[e] Department of Human Geography and Spatial Planning, Faculty of Geosciences, Utrecht University, 3584, CB, Utrecht, the Netherlands
[f] Accident Research Institute, Bangladesh University of Engineering and Technology (BUET), Dhaka 1000, Bangladesh

ARTICLE INFO

Keywords:
Travel behavior
Urban form
Nonlinear effect
Interaction effect
Machine learning
Sustainable transportation systems

ABSTRACT

In this study, we aimed to investigate the influence of the built environment (BE) on commuter mode choice using machine learning models in a dense megacity context. We collected 10,150 home-based commuting trips data from Dhaka, Bangladesh. We then utilized three machine learning classifiers to determine the most accurate prediction model for predicting the mode of transportation chosen for commuting in Dhaka. Based on the predictive performance of the classifiers, we identified that the Random Forest (RF) algorithm performed the best. Using the RF model, this study also explored the relative importance of BE factors in predicting commute mode choice, identified nonlinear relationships between the BE factors and mode choice, and examined the interaction effects of these factors on mode selection. Our results reveal that, compared to socio-demographic factors, the BE substantially influence commuter travel behavior. The BE characteristics have a specific nonlinear threshold limit at which they can have a notable impact on lowering private car use, and private car use does not display a constant return of scale with BE. Their interaction effects illustrate the potential optimal combination of BE interventions to lower private car use for commuting. These findings hold substantial implications for urban environmental policy, emphasizing the need for transit-oriented development, travel demand management, and integrated land-use transportation planning to foster low-carbon transportation systems in cities like Dhaka.

1. Introduction

Despite concerted efforts to advance sustainable development, the prevalence of private car usage in urban areas continues to grow, exacerbating greenhouse gas emissions and increasing energy consumption. Consequently, this trend undermines the pursuit of sustainable mobility objectives. Developed nations engage in urban planning strategies that involve modifying the built environment of cities (e.g., compact city development, smart growth, and Transit-Oriented Development) in order to reduce private car use, thereby mitigating urban challenges and ensuring sustainable mobility (Ewing and Cervero, 2010). The built environment (BE) is widely acknowledged as a significant contextual determinant in promoting the adoption of public transportation, walking, and cycling over private car, hence offering a potential remedy for reducing emission (Ashik et al., 2023; Wu et al., 2019) and enhancing physical and mental well-being (Handy et al., 2002; Panter et al., 2019).

In the field of travel behavior research, it is commonly hypothesized that individuals tend to make travel mode decisions that optimize their utility. The BE exerts influence on the range of options available to individuals. At its core, the spatial attributes of the BE, including density, land use diversity, and street connectivity, play a crucial role in determining the proximity of humans to various destinations. Consequently, the associated costs of traversing these distances significantly impact individuals' choices about their destinations, transportation modes, and frequency of travel (Handy, 2017). The compact development hypothesis posits that in communities characterized by higher densities, land use mix, and street connectivity, residents tend to engage in reduced

* Corresponding author.
E-mail addresses: fajle.ashik@mail.mcgill.ca (F.R. Ashik), atif.sreezon@coact.org.au (A.I.Z. Sreezon), hamidur.rahman@utexas.edu (M.H. Rahman), zafri@urp.buet.ac.bd (N.M. Zafri), s.m.labib@uu.nl (S.M. Labib).

https://doi.org/10.1016/j.jtrangeo.2024.103828
Received 9 June 2023; Received in revised form 20 December 2023; Accepted 27 February 2024
Available online 4 March 2024

driving behavior. This is primarily attributed to the shorter distances required to reach nearby establishments or other destinations, as well as the availability of alternative modes of transportation such as walking, cycling, or utilizing public transit (Kamruzzaman et al., 2016; Ewing and Cervero, 2010). The existing body of literature corroborates the concept that individuals residing in more densely populated places exhibit a tendency to reduce their reliance on private vehicles and instead choose for alternative forms of transportation, assuming all other factors remain constant (Aston et al., 2020; Ewing and Cervero, 2010).

The current literature on this idea is primarily supported by empirical evidence from the Global North. However, there is limited knowledge regarding the situation in the Global South, except for a recent increase in studies focusing on China. This colonization of transport knowledge prevents knowledge from becoming more diverse and deters essential interaction with the varied geographies of transport knowledge (Castañeda, 2021). However, studying the variations in travel mode choosing behavior in different contexts can offer valuable information on how the BE can impact the choice of transportation modes (Wang and Zhou, 2017). As an example, Munshi (2016) posed the question of density: "does increasing density in an already dense environment make sense?" in relation to Rajkot, India's travel patterns. It recalls Hanson's discourse, which states, "knowledge does not mean much if it is not contextualized, if we do not know where it came from in terms of social and geographic context" (Hanson, 2010, p. 17).

Dhaka, the capital city of Bangladesh, has a population density of 55,169 individuals per square kilometer, making it one of the most densely populated cities in the world. The city is distinguished by its primacy, meaning it has a high concentration of jobs, services, and population compared to other cities in the country. It has a non-inclusive urban structure, meaning there is a lack of integration among the many components of urban design. The city is characterized by a lack of hierarchy and subpar design on its streets. The city's transportation system exhibits a diverse mix of travel modes, with 8.5% of trips being made by car, 31.2% by public transport, and 60.3% by active transport (Ashik et al., 2022). However, there is a lack of sufficient public and active transport infrastructure. Additionally, there is a growing trend of private car and motorcycle ownership, with 78% of newly registered motor vehicles in 2019 being private automobiles (such as cars, jeeps, and motorcycles), while only 2% were public buses (Nakshi and Debnath, 2020). The fast expansion of the metropolitan area, coupled with economic advancement, suggests the likelihood of a swift increase in motor vehicle usage in the near future. Due to the inadequate infrastructure and limited options for active and public transportation in Dhaka, along with the escalating rate of motorization, Dhaka presents an ideal setting for studying commuting mode choice behavior. This research focuses to investigate the whole extent of these shortcomings and identify particular chances for mobility improvements by studying the effects of the BE on commute mode choice.

In this unique setting of Dhaka, we opted for Machine Learning (ML) to ascertain the effects of the BE. Utilizing ML can assist in detecting any existing nonlinear patterns of association. Failure to consider nonlinear patterns might lead to biased modeling outcomes, hence providing false implications for urban and transport policies (Ding et al., 2018a; Hatami et al., 2023). Given the spatial interaction of BE, it is crucial to comprehend the interplay between BE measures in order to predict commuting mode selection and gain a deeper understanding of the intricate relationships and planning implications. Concerning the modeling approach, it is important to note that car ownership and commute duration are likely to be influenced by BE (Ding et al., 2017). Thus, applying a traditional regression model to estimate commute mode choice may result in the endogeneity problem, although such modeling approaches have been the dominant analytical approach used in existing literature. Machine learning relaxes the rigid assumptions of traditional statistical models and has the ability to handle high dimensional transport and environment related information, thus lowering the problem of endogeneity (Yuan et al., 2022; Mannering et al., 2020). Hence, conducting research that focuses on investigating the nonlinearity and multiplicity of elements in comprehending the BE and commuting behavior pattern in the South Asian context can offer valuable policy recommendations for local planners to reduce the usage of private cars. In this study, we will, therefore:

- Assess the direct impacts of the BE on commuting mode choice by estimating the relative importance of the BE at residence and workplace.
- Investigate the nonlinear relations between BE factors and commuting mode selection behavior, focusing on private car mode choice.
- Identify the simultaneous interactions of multiple BE indicators and their respective impact on commuting mode choice behavior.
- Identify how BE interacts with commuting time and car ownership to predict commuting mode choice behavior.

2. Background

2.1. BE and mode choice

Commuters' mode choice is influenced by BE and socio-demographic factors (Kamruzzaman et al., 2013; Munshi, 2016; Shirgaokar, 2015; Tran et al., 2016). There is a wealth of literature on whether the BE influences mode-use, yet little emphasis is given to the magnitude of the BE's influence (Axhausen, 2015; Wang and Zhou, 2017). Additionally, Stead (2001) and Singh et al. (2018) indicated that socio-demographic factors (e.g., income, gender), rather than land-use features, explain greater variability in travel behavior. These studies made this claim solely based on parametric estimation, although a more recent study by Ding et al. (2018a) argued that examining the relative importance of explanatory variables is more applicable to estimating collective predicting power. As a result, there is a debate in the research on which factor is more critical in predicting mode use: the BE or socio-demographics.

Furthermore, although land use and transport at job locations influence commuting behavior significantly (Ding et al., 2018b; Islam and Saphores, 2022; Sun et al., 2017; Wang et al., 2020), the influence of the BE at residence is mainly examined in the literature to explain commuting (Aston et al., 2021; Etminani-Ghasrodashti and Ardeshiri, 2016; Ewing and Cervero, 2010; Stevens, 2017; Wang et al., 2021). Most of the previous studies did not consider the workplace BE and might have faced biases in results due to the omission of potential independent variables. In addition, they rarely provided guidance on what kind of land use and transportation options should be established for the workplace to facilitate sustainable commuting. However, recent research that considered the BE of employment hubs in modeling found conflicting results regarding whether home or workplace has a more influential role in commuting. Studies conducted in the USA concluded that the BE of employment centers has greater potential to explain commuting mode use than residence-based BE (Chen et al., 2008; Ding et al., 2018b).

In contrast, Sun et al. (2017) confirmed that the residential environment is more important in China and stressed the need for empirical research locally. The authors argued that the relative importance of the built vs. social environment might vary based on local contexts. Only a limited number of research have examined the behavior of travel mode choice in the varied density and land-use settings of South Asian cities, including India and Bangladesh (Enam and Choudhury, 2011; Munshi, 2016; Nakshi and Debnath, 2020; Shirgaokar, 2015; Rahman et al., 2023; Goel and Mohan, 2020). Utilizing conventional statistical modeling methods, these studies have determined that the relationship between BE and mode choice varies from what has been observed in the global north. Moreover, the impact of population density on mode choice varies within the global south context: Munshi (2016) found a minimal effect, whereas Goel and Mohan (2020) found no effect.

Shirgaokar (2015) demonstrated a positive correlation between improved employment accessibility and increased car usage. Nakshi and Debnath (2020) claimed that the incorporation of BE factors has greatly enhanced the performance of the mode choice model. Furthermore, Rahman et al. (2023) investigated that the BE has a more pronounced impact on walking compared to the personal characteristics of travelers in Khulna, Bangladesh. Nevertheless, none of these studies examined the influence of workplace BE on the selection of transportation mode. Enam and Choudhury (2011) identified and explored the challenges and complexities associated with developing a comprehensive mode choice model for Dhaka. Therefore, there is less information regarding the influence of the BE on the selection of transportation modes for commuting, specifically in the context of the Global South, notably in Bangladesh. Research undertaken in the megacity of Dhaka, with its distinct land use and transportation settings, will contribute to the existing literature by addressing the current lack of understanding.

2.2. Machine learning: Linking between the BE and travel mode use

Traditional statistical regression models are extensively employed in the literature to investigate the function of the BE in predicting commuter mode use (Kamruzzaman et al., 2013; Munshi, 2016; Nguyen et al., 2020; Shirgaokar, 2015; Tran et al., 2016). However, these models often assume a linear (or linearity in parameters) relationship between the BE and travel behavior, which may not accurately reflect the complex and nonlinear nature of this relationship (van Wee and Handy, 2016). For instance, an increase in residential density from 20 to 40 people per acre may have a greater impact on walk mode share than the same increase from 200 to 400 people per acre. The latter may have unfavorable effects on walking if it causes congestion and reduces the quality of the walking environment. Due to the fixed coefficients specification in conventional models, they cannot determine such threshold or ceiling effects, which occur when an intervention has a minimum or maximum level of impact, respectively. In addition, multicollinearity is often common in classical statistical models. In the recent review article, Aston et al. (2020) reviewed the impact of research design on modeling the BE and travel mode use relationship; the authors concluded that collinearity could explain many differences in results among 140 studies (their reviewed papers). Machine learning techniques have been employed as a promising solution to address some of these challenges in recent transportation studies. To handle multicollinearity better, missing observations, predicting capability and to measure the relative importance of explanatory factors, machine learning algorithms usually outperform classical models (Cheng et al., 2019; Ding et al., 2018a; Ding et al., 2018b).

Although ML algorithms usually provide higher accuracy in prediction than traditional modeling, some researchers also argued that the prediction capability and sensitivity of these algorithms should be investigated in various contexts. For example, these models have different levels of uncertainty (number of design parameters, amount of training data) and sample bias (Aston et al., 2020; Cheng et al., 2019), which requires better training and testing in different spatial contexts. Such evaluations among various machine learning algorithms in the BE and mode choice modeling are still limited among existing studies. Furthermore, the hyper-parameter values should also be determined with caution, as the tuning optimal hyper-parameters vary depending on data and circumstances (Hillel et al., 2021). Ding et al. (2018b) used hyper-parameter values from another work (Ding et al., 2018a), which may have resulted in sub-optimal hyper-parameter selection, and the resulting model might have erroneous predictions due to sub-optimal selection of hyper-parameters.

As ML algorithms do not rely solely on the distribution of input data, and most of the common algorithms, such as decision trees, random forests, and support vector machine, can investigate relationships within the data without strict distributional assumptions as needed in the classical statistical models (Hastie et al., 2009; Somvanshi et al., 2016). Consequently, in contrast to standard statistical models (e.g., linear regression), they possess greater potential for comprehending nonlinear relationships between Built Environment (BE) factors and commuting behavior (Ding et al., 2018b; Ding et al., 2018a). Due to the potential presence of the nonlinear effects, Ding et al. (2018b) and Wu et al. (2019) argued that the BE variables might have a considerable impact on commuting behavior but only within a specific range which may not be reflected in linear relations. For example, Ding et al. (2018a) discovered that population density affected travel distance within specific population density values. Such nonlinear associations may help determine a variable's most effective range that has considerable influence on commuting mode choice. However, Ding et al. (2018b) only investigated the nonlinear effects of the BE on commute car usage, and the study did not include some critical factors (i.e., car ownership) in the modeling process. Further investigations are warranted to better understand these relations using varying machine learning algorithms in such a context.

Interactions among BE indicators could also be caused by their spatial correlation. Consequently, the impact of a BE indicator on mode choice can be amplified or diminished due to the interactions with other BE indicators. For example, residents in a well-connected community with high transit accessibility would be more likely to employ active and transit modes of transportation (Kamruzzaman et al., 2014; Kumar et al., 2018). As a result, it is also vital to investigate the simultaneous interaction effects among BE indicators to design a low carbon oriented urban structure.

Usually, commuting duration and car ownership are two key variables that influence commuting mode choices (Ashik et al., 2023; Ding et al., 2017). Simultaneously, the BE is associated with commuting time and car ownership (Rahman and Antipova, 2024). Therefore, in order to accurately assess the impacts of the BE, it is crucial to consider the issue of endogeneity associated with car ownership and commuting time (Ding et al., 2017). Utilizing explainable machine learning techniques to identify interaction effects is a valuable addition to traditional models for uncovering the heterogeneity within data, especially in cases where there may be potential endogeneity (Yuan et al., 2022). As a result, understanding how the BE interacts with commuting time and car ownership in predicting commute mode choice is critical; however, previous studies have often overlooked how these features interact with the BE or vice versa (Church et al., 2012). Ignoring the interaction effects between commute time and BE, as well as car ownership and BE, raises the following questions: a) how much variance can these interactions explain in predicting mode choice behavior? b) what functionalities do these interactions play in reducing private car use? Answering these questions would provide a better understanding of the importance of these interactions and may aid policymakers in promoting a low-carbon oriented sustainable metropolis.

3. Methods and materials

3.1. Data

3.1.1. BE indicators

BE variables were chosen based on Ewing and Cervero (2010) widely used 5D norms. Moreover, we chose several *D* variables based on previous research in Dhaka (Ashik et al., 2023; Ashik et al., 2022; Rahman et al., 2022). Street connectivity, which impacts the proximity of destinations and the directness of travel, was included as it may encourage walking to various destinations, including transit stations, thereby increasing transit ridership (Etminani-Ghasrodashti and Ardeshiri, 2016; Kamruzzaman et al., 2014; Stevens, 2017). Land use diversity and the job-household ratio promote active mode and transit choice by clustering a wide variety of destinations close together (Aston et al., 2021; Ewing and Cervero, 2010; Munshi, 2016). Since higher densities are what make regular mass transit service possible (Kamruzzaman et al., 2014), it stands to reason that higher densities would also enhance transit use. Higher densities are believed to be advantageous because more individuals will choose to walk since they will observe more people walking and will perceive themselves to be safer as a result. It also seems to reason that those who live near a transit station will make more use of transit services.

The BE indicators and measurement units and descriptions are presented in **Table-1**. For BE data, we used Draft Dhaka Structure Plan (DSP) 2016–2035 spatial data set collected from Rajdhani Unnayan Kartripakkha (RAJUK), a development authority for the capital city of Dhaka of Bangladesh. The spatial dataset included primary road network, building, land use-related information for the entire Dhaka Metropolitan Region (DMR) area. Inside DMR, two City Corporation Dhaka North and Dhaka South, with 92 Wards (the lowest administrative units), are located. These 92 Wards (average size of 150 Hectare), termed as Traffic Analysis Zones (TAZs) in the Revised Strategic Transport Plan (RSTP)-2015 document, are study units where BE data were measured (Fig. 1).

3.1.2. Socio-demographic and commuting indicators

For socio-demographic and commuting data, we used a secondary database from the RSTP-2015, the latest and comprehensive study on the transportation for Dhaka city conducted under the Dhaka Transport Coordination Authority (DTCA). RSTP-2015 included a Household Interview Survey (HIS), conducted from August 2014 to November 2014, dividing the whole DMR area into TAZs. RSTP-2015 delineated Traffic Analysis Zones (TAZs) following the Ward boundaries for ensuring consistency with DSP 2016–2035 and other government plans. A proportionate sampling strategy was used to estimate the sample sizes for each TAZ. The very first respondent in each TAZ was randomly selected, followed by a systematic random sampling procedure. HIS survey data included the trip maker's socio-demographic information and weekday household travel records. This study considered 10,150

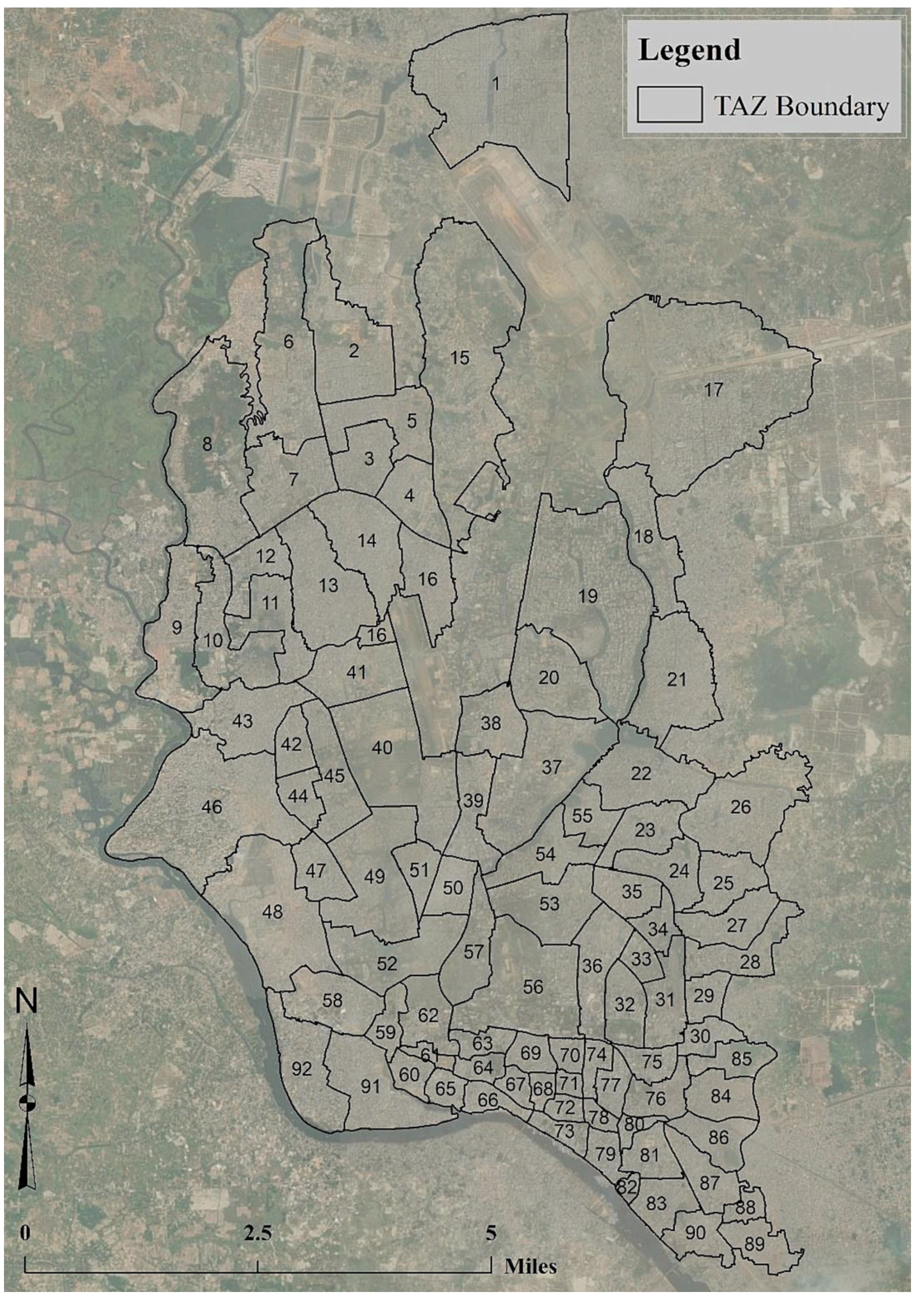

Fig. 1. Study Area Boundary showing the traffic analysis zones (TAZs).

home-based commute travels (from 7632 families) inside the defined TAZs. However, it is important to mention that the socio-demographic variables are assessed at the person level, whereas the BE variables are aggregated at the TAZ level.

Commuting is a major contributor to urban traffic, making it a focal point of research on how the BE impacts people's commuting behavior (Ashik et al., 2023; Ding et al., 2018b; Ding et al., 2021; Munshi, 2016; Wang and Zhou, 2017). With urbanization and increased reliance on cars, the carbon footprint of commuting is likely to increase, exacerbating the effects of climate change (Wang et al., 2017). As a result, the scope of this study is limited to commuting trips.

3.1.3. Descriptive statistics of the variables

The BE, socio-demographic and commuting characteristics for all the TAZs are presented in **Table 1**. In our case study area, the link to node ratio value (mean = 0.77, SD = 0.11) indicates the road network is not sufficiently well-connected. Researchers advocated for a street network with a link node ratio >1.4 (Dill, 2004). The average distance to a transit stop is 0.90 km, which is about 10 min' walk, indicating a reasonable walking distance to avail public transit (see 20-min neighborhood concept by Thornton et al. (2022); TOD typology by Rahman et al. (2022)). The land-use diversity score indicated a moderate level of diversity in land use mixes within each TAZs (Ashik et al., 2022; Rahman et al., 2022). Additionally, the average population density is considerably high because of the massive population of Dhaka city, and the mean Job-household ratio value is 1.35, which can be characterized as a balance of job availability and demand within the city (Rahman and Ashik, 2020).

Table 1
Definition of model variables and descriptive statistics.

Variables		Variables Description	Outcome	N (%)	Mean (SD)	Wider population[7] %/Mean
BE Factors	Population Density	Household per hectare of residence and workplace	Households/ha		265 (244.62)	
	Employment Density	Jobs per hectare of residence and workplace	Jobs/ha		304 (301.38)	
	Link Node Ratio	Network analysis of road network data is used to figure out the number of links to nodes in each TAZ.	Ratio ranging from 1 to 2.5		0.77 (0.11)	
	Diversity	Land use diversity is measured by entropy index considering residential, commercial, industrial, institutional, and open space land use.	Ranging from 0 to 1		0.471 (0.14)	
	Job-household Ratio	How many jobs there are per household in each TAZ	Ratio		1.35 (0.95)	
	Distance to the transit stop[1]	Distance in kilometers between the center of each TAZ and the closest bus stop on the network	Kilometer		0.90 (0.44)	
Socio-demographic Factors	Gender	Gender of the respondent	Male	7826 (77.1)		55.61[2]
			Female	2324 (22.9)		44.38[2,3]
	Age	Respondent age in year	Year		35.13 (12.18)	25.7[4]
	Household Size	How many people live in the respondent's home	Number		4.34 (1.33)	4.4[2]
	Income Group	Respondent's income group based on how much money his or her household makes every month.	Low-income (<15,000 BDT)	1866 (18.4)		22,565[5]
			Medium income (15000–50,000 BDT)	3828 (37.7)		
			High-income (>50,000 BDT)	4456 (43.9)		
	Driving license	The access of a driver's license	Yes	928 (9.1)		
			No	9222 (90.9)		
	Private Car Ownership	How many private cars including two-wheeler each household has.	Number		0.120 (0.37)	0.07[4]
	Household Bicycle	How many bicycles each household has.	Number		0.023 (0.15)	0.01[4]
Commuting factors	Commuting time	The time traveled a respondent to reach his/her destination from home	Minute		33.94 (27.33)	33.27[6]
	Modal Share	Use of vehicle as a commuting mode	Transit (Bus)	3168 (31.2)		27.7[6]
			Private (car, microbus, and motorcycle)	860 (8.5)		7.7[6]
			Active (walking, bicycle, and rickshaw)	6122 (60.3)		64.6[6]
Number of observations	10,150					

[1] In our study, rather than using the centroid of a TAZ as a point to calculate the distance to the nearest bus stop, this study divided the TAZs into 200 m*200 m squares. As part of a network analyst dataset, the centroid of each grid cell was utilized to compute the distance to the next bus stop. The average distance for each TAZ was estimated once all of the origin-to-bus-stop distances had been obtained.
[2] (BBS, 2011).
[3] According to World Bank, women contributed only around 28% of work force in 2014 (Trading Economics, 2021).
[4] (Ashik et al., 2022).
[5] (HIES, 2016).
[6] (DTCA, 2015).
[7] Wider population refers to all population not only commuters.

Socio-demographic characteristics of the sample households suggest males as a dominant section of home-based work trips. Regarding the respondents' total household income, 43.9% fall in high-income groups, 37.7% in medium-income groups, and the rest (18.4%) fall in the low-income group. Here, the sample population has a mean age of 35.13 years old with an average household size of 4.34, and median household income is 25,000 BDT (an equivalent of US $300). Household automobile ownership information data shows that most of them have no private mode for household use. Bicycle availability of households also showed a majority have no bicycle in their house. Driving license information shows that most of the respondents have no driving license (90.9%).

Regarding the commuting characteristics, we observed that 8.5% of trips were made by private car (motorized four-wheelers: car and microbus and motorized two-wheelers: motorcycle), while 60.3% by active mode, and the rest of 31.2% by transit mode. This reflects the dominance of active (i.e., walking, bicycle, and rickshaw) and transit (e.g., bus) as the main commuting modes in the city, which is supported by the policy documents (RSTP-2015); DSP 2016–2035) of the city (DTCA, 2015; RAJUK, 2015). The low share of private cars is usually common in other developing countries. For example, in Mumbai, India (Shirgaokar, 2015) and Hanoi, Vietnam (Nguyen et al., 2020), private cars account for only 6% and 5.91% of trips, respectively. In Bangalore, India, <10% of non-work trips are made by private car (Manoj and Verma, 2016).

3.2. Machine learning modeling approach

3.2.1. Data processing

The next step after combining the built, social, and travel data for all TAZs in Dhaka was to prepare the dataset for model training. Several pre-processing steps were performed to curate the data properly for training. First, the data was split into a 70:30 ratio for the training and testing dataset. Out of the total samples, 70% (7107) were used for training, and 30% (3047) were used to test the classifier performance. It was ensured that each split represented the class distribution of the original dataset. However, class imbalance (transit class: 3168 samples, private class: 860 samples, active class: 6122 samples) (Table 1) in the training dataset would impose a bias on the classifier performance. During the training iterations, the model would have learned the characteristics of the transit and active classes better than the private class, which was not desirable. Although the ML models are generally designed to improve prediction accuracy, in our case, we also used explainable ML techniques along with these predictive models to better understand the influence of BE on mode choice. Therefore, to obtain an accurate representation, we needed the classifier model to train on a balanced dataset. So, we resolved the issue by over-sampling the minority class, known as the Synthetic Minority Oversampling Technique (SMOTE) (Chawla et al., 2002). During training, this process will randomly duplicate the private class samples, which, in this case, is the minority class. Class imbalance can negatively affect the classifier's performance (Guo et al., 2008). Thus, it is crucial to address it in the early stage. Oversampling was done only for the training set to ensure training does not introduce bias. Finally, to account for the potential presence of multicollinearity for each independent variable, a variance inflation factor (VIF) test was performed. Explanatory variables with VIF values exceeding 10 need to be removed, as Salmeron et al. (2018) and Li et al. (2024) recommend. Notably, the independent variables we considered in this study did not exceed the value of 10, and all have a VIF that falls within 6.5. Therefore, we assumed multicollinearity would not considerably influence our models' performance and may not induce bias in measures of importance.

3.2.2. Selected algorithms and parameters

In this study, we chose three algorithms for their past success in classification tasks in various fields of research, including transportation studies (Hagenauer and Helbich, 2017; Xu et al., 2014): GBDT, RF, and Support Vector Machine (SVM). Both GBDT and RF are ensemble methods that use decision trees (Lindner et al., 2017) as their base algorithm. SVM is a statistical approach to drawing decision boundaries within high-dimensional data in hyperplanes (Noble, 2006).

The GBDT algorithm is a nonlinear model based on a decision tree (Friedman, 2001). This algorithm can model nonlinear and complex relationships of explanatory variables (Ding et al., 2018b). The main goal of such boosting is to identify optimal trees for predicting complex cases by generating multiple iterations of trees ensembled sequentially (Ma et al., 2017). Compared to the boosting approach, other ensemble machine learning algorithms based on decision trees (e.g., RF) apply bootstrap aggregation or bagging methods to predict better. The RF algorithm proposed by Breiman (2001) reduces the overfitting of classification models by growing an ensemble of decision trees in parallel. The trees would be built on identical and independent random vectors selected from the bootstrapped sample of the original dataset. Then each tree is trained on the bootstrapped data, and each casts a unit vote for the final decision on class prediction. Usually, bagging and boosting provide better predictions than the single decision tree models by reducing overfitting issues (Breiman, 2001; Friedman, 2001).

In contrast to the ensemble methods based on decision trees, the SVM is a supervised machine learning algorithm that uses hyperplanes to classify or separate data points in different classes from high-dimensional data (Noble, 2006). For a linearly separable dataset, the algorithm tries to find the maximum and minimum distance between classes by fitting several hyperplanes. Then, based on a loss function (e.g., convex loss), it optimizes the distances for all the hyperplanes and selects a hyperplane that maximizes the distance between the support vectors (Tong and Chang, 2001). Here, the points along the hyperplane are called support vectors. For nonlinearly separable data space, SVM can be used with the kernel trick approach, where different kernels such as polynomial and radial basis functions (RBF) are applied to enlarge the feature space into higher dimensional planes, where it would be possible to fit hyperplanes that easily separate data points (Jakkula, 2006).

3.2.3. Model calibration and evaluation of algorithms

The three algorithms mentioned in the previous section have multiple hyper-parameters, and their performance depends on a range of values of those hyper-parameters. To select the best classifier for our purpose, we needed to compare the parameter-optimized algorithms and use the best performer (Table 2). A grid search was performed based on a range of values for each classifier to find the best combination of hyper-parameters (Hsu et al., 2003). During the grid-search training process, we used cross-fold validation (Burman, 1989) to further randomize and validate the models. Cross-fold validation splits the data into k-number of groups and uses $k-1$ number of them for training, and the rest were used to validate the classifier. This process was repeated until all the groups were used as the validation set. For our case, we selected $k=10$, as noted by Wong and Yeh (2020). After the grid search, the best combinations of the models were evaluated to select the

Table 2
Grid values of the hyper-parameters of the classifiers.

Classifier	Hyper-parameter	Hyper-parameter Grid
RF	Number of estimators/Trees	100, 500, 1000
	Criterion	Gini, Entropy
	Maximum Depth	1–10 (steps of 2)
GBDT	Number of estimators/Trees	100, 500, 1000
	Learning Rate	1e-5, 1e-2, 0.1
	Maximum Depth	1–10 (steps of 2)
	Maximum Features	Sqrt, Log2
SVM	C (Regularization parameter)	1e-2, 0.1, 1
	Kernel	Linear, Poly, RBF, Sigmoid
	Decision Function Shape	One-vs-Rest ('OVR'), One-vs-One ('OVO')

appropriate one.

To evaluate the algorithms, we used accuracy and F1-score metrics. Accuracy refers to the percentage of correct number predictions among all predictions made Junker et al. (1999). $F - measure$ was introduced by Rijsbergen (1979) as the harmonic mean of precision and recall (Hripcsak and Rothschild, 2005). Recall, also known as Sensitivity, refers to the proportion of accurate class detection (true positives) among all the samples of such classes (true positives and false negatives). Simply, it measures the number of correctly identified classes from that class and which ones it missed (Powers, 2020). Precision, also known as Confidence, refers to the proportion of predictions for a class that are truly correct, meaning of all the predictions for a single class by the model (true positives and false positives), how many are truly correct (true positives) (Powers, 2020).

3.3. Research design

This study measures the direct effect of the BE by measuring the relative importance of independent variables derived using impurity measures. Friedman's H-statistic is applied to measure whether and to what extent each individual BE indicator interacts with others (Friedman and Popescu, 2008). The same statistic is used to determine how commute time and car ownership interact with BE indicators in predicting commute mode choice. In this paper, we only considered bivariate interaction. We employed Accumulated Local Effects (ALE) methodology to determine the main effects of the BE. ALE is capable of isolating the effect of a specific feature, even in cases when the features exhibit some degree of correlation (Apley and Zhu, 2020; Molnar, 2020). Furthermore, we used partial dependence plots (PDP) to investigate the simultaneous interaction effects between multiple BE indicators, BE and commute time, and BE and car ownership. PDP makes it possible to display the overall effect by combining the main effect with the interaction effect (Apley and Zhu, 2020; Molnar, 2020).

4. Results

4.1. Evaluation of algorithms in modeling mode choice

Before exploring the impact of BE on commuter mode choice, we first evaluated the accuracy of the modeling algorithms in predicting their relationships. Regarding accuracy, among all the algorithms tested along with their optimum hyper-parameters, the GBDT model was the best, with an 89% overall accuracy score, and the RF model was the second best, with an 87% score (Table 3). Table 4 indicates that the F-1 score (harmonic mean of sensitivity and precision) of these models generally follows the same pattern: all the models predict active class better and private car class worse. This implies that, while these models are good predictors of mode choice regarding overall accuracy and F1-score, neither has distinguishing characteristics that would make one algorithm superior to another in predicting specific travel modes. However, as GBDT and RF both showed promising results, it is also necessary to examine their training performance to choose the best of them. In our case, we compared the grid search training and test accuracy between RF and GBDT model over several iterations to identify when and which algorithm might be overfitted with increased number of training iterations. We found that the training and test accuracy was most consistent and closely aligned for the RF model. But we observed a large difference in training and testing accuracy or divergence of accuracy for the GBDT model, which indicates overfitting of the GBDT model (details of the training and test accuracy in ***Supplementary Note 1***). It should be noted that overfitting is a critical problem when using machine learning algorithms for predictions (Ying, 2019; Raihan et al., 2023). Considering all these evaluation criteria and overfitting principles, we concluded the RF is the algorithm for our dataset and we selected the RF model with its best parameters to further continue our investigation.

Table 3
Grid search hyper-parameter best result for the selected algorithms and accuracy evaluation.

Classifier	Hyper-parameter	Best Hyper-parameter	Accuracy
RF	Number of estimators/Trees	1000	87%
	Criterion	Entropy	
	Maximum Depth	9	
GBDT	Number of estimators/Trees	500	89%
	Learning Rate	0.1	
	Maximum Depth	10	
	Maximum Features	Sqrt	
SVM	C (Regularization parameter)	1	79%
	Kernel	Linear	
	Decision Function Shape	One-vs-One ('OVO')	

Table 4
F1-Score comparison of the algorithms in predicting commuter mode choice.

	GBDT	RF	SVM
Active Mode	0.90	0.89	0.87
Private Car	0.66	0.67	0.65
Transit Mode	0.85	0.81	0.73

4.2. Relative importance of social and BE on commuting mode use

Our modeled results identified the relative importance of the explanatory variables in explaining the commuting mode choice using the interpretable RF. RF is an interpretable classifier that further investigates the importance of commuting mode choice prediction (Table 5). We observed that commuting time is the most crucial feature in selecting a mode. Additionally, among socio-demographic characteristics, automobile ownership has the most important direct influence

Table 5
Relative importance of the explanatory variables to explain commute mode choice.

Explanatory Variables		Direct relative importance (%)	Cumulative relative importance (%)
Trip feature	Commuting time	18.6	18.6
Socio-demographic features	Car ownership	13.8	36.1
	Driving License	6.30	
	Age	6.20	
	Household income	5.60	
	Household size	2.90	
	Gender	1.10	
	Bicycle ownership	0.20	
BE at residential location	Diversity	4.50	23.5
	Link Node Ratio	4.20	
	Job Household Ratio	4.00	
	Distance to transit stop	3.70	
	Employment density	3.60	
	Population density	3.50	
BE at workplace	Diversity	4.00	21.8
	Job Household Ratio	3.70	
	Distance to transit stop	3.70	
	Population density	3.60	
	Link Node Ratio	3.50	
	Employment density	3.30	
Total		100	100

on mode choice. Other than these dominant variables, the model identified driving license and age are the following two critical variables with higher total relative importance than other socio-demographic characteristics (e.g., household size, gender). For BE features at home and work, diversity of land use, distance to transit stops, job household ratio, and link node ratio usually has high relative importance through direct effects on the mode choice (Table 5). At residence, diversity of land use contributes 4.5% of the direct effect in predicting mode use.

Our results also indicate that the direct cumulative effect of the BE (home: 23.5%, work: 21.8%; total: 45.3%) is more pronounced than the direct contribution of socio-demographic attributes (36.1% of the direct effect) in explaining the mode choice for commuting trips (Table 5). It should be highlighted that the BE at residence has greater explanatory power than the BE at work. The BE at residence location accounts for 23.5% of the direct effect in explaining commuting mode utilization, while the BE at work accounts for 21.8% of the direct effect.

4.3. Nonlinear influence of BE on commuter mode choice

The nonlinear effects of the BE on active use at residence and workplace are presented in ***Supplementary Note 2***. Overall, the residents living in a neighborhood characterized by a well-connected road network, diversified land use pattern, and higher job accessibility are more likely to choose an active mode for commuting. Similarly, for transit mode, **Fig. S3** and **Fig. S4** illustrate that the residents working in areas with diversified land use, better transit accessibility, and higher job potential are more likely to choose transit for commuting (details in ***Supplementary Note-2***). Such results generally reflected the positive influence of several BE characteristics (e.g., diversification, increased access) on active and transit mode use. However, in Dhaka, increasing numbers of people are now focusing on using private cars for commuting. This might lead to further deterioration of the traffic congestion situation and decreased air quality (Ashik et al., 2023; Ashik et al., 2022; Haider and Papri, 2021; Labib et al., 2018; Zafri et al., 2021). Considering the rise of car-based commuting trips in Dhaka, we intended to focus on the impact of the BE on private car choice in our study area.

Fig. 2 illustrates the non-linear relationship between the selection of private cars and the BE at residential locations based on ALE plots. Specifically, as the distance between a residence and a transit station expands from 250 m to 750 m, there is an observed rise in the probability of individuals opting for private automobile usage. However, there is a decline in the likelihood of individuals opting for private cars when the distance between their place of residence and transit stops is between 750 m and 1.25 km (see Fig. 2-**ii**). The unexpected occurrence can be elucidated by considering the local context: low-income households residing at greater distances from the transit station (Supplementary note-3) have a higher propensity for using active transportation (Supplementary note-2). Nevertheless, it has been observed that when the distance between a workplace and a transit stop is within 1.25 km, the probability of individuals choosing to drive to work tends to increase. However, once the distance surpasses 1.25 km, the likelihood of opting for a car as the mode of transportation begins to decline (see Fig. 3-**ii**).

The presence of a homogeneous land use pattern (as indicated by an entropy index of <0.35) in residential areas has been found to have a positive or little impact on the choice of private cars, as shown in Fig. 2-**iii**. This positive impact may be attributed to the potential existence of a singular dominant land use type. However, it was observed that there was a significant decrease in the likelihood of selecting a private car as the level of land use diversity grew from 0.35 to 0.65, indicating the presence of more than two distinct land-use types. In the context of residential settings, the marginal effects of increasing diversity become trivial after this range. As a result, land use diversification to reduce private car use may have a critical threshold. Moreover, the presence of diverse land use patterns in the workplace setting encourages individuals to opt for private car transportation (see Fig. 3-**iii**). Diverse land use patterns within workplaces have the potential to incentivize commuters to engage in intricate trip chains that may prioritize the choice of automobiles.

Lower job-household ratio (<1) in residential areas increases the likelihood of using private cars for commuting, but the probability starts to decline when the ratio reaches 1 (Fig. 2-**iv**). This implies having more jobs within the residential areas can lower private car trips by allowing people to walk or use other transportation modes to get to work. For workplace areas, Fig. 3-**iv** indicates, in the range of [0.5, 1.5], the job household ratio is associated with higher private car use, but once it reaches 1.5, car use starts to decline. This suggests that housing and employment should be located together in a city to minimize the use of private cars.

The negative association between link node ratio and private car use is observed both at residence and workplace from Figs. 2-**i** and Fig. 3-**i**. Population density at residence and employment density at work positively affect car choice probability. Employment density at residence and population density at work inversely influence private car use; however, after a specific range (i.e., about 300 jobs per hectare at the residence Fig. 2-**vi**, about 200 households per hectare at the workplace Fig. 3-**v**), these parameters have a positive relationship with private car

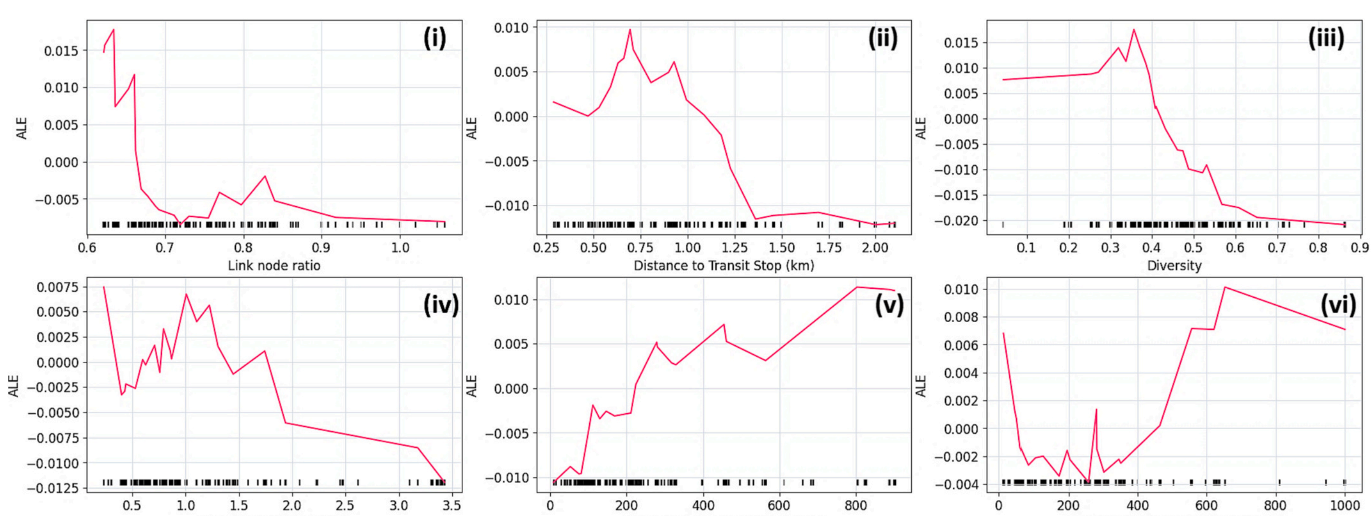

Fig. 2. Nonlinear effects of BE at residence based on ALE plots.

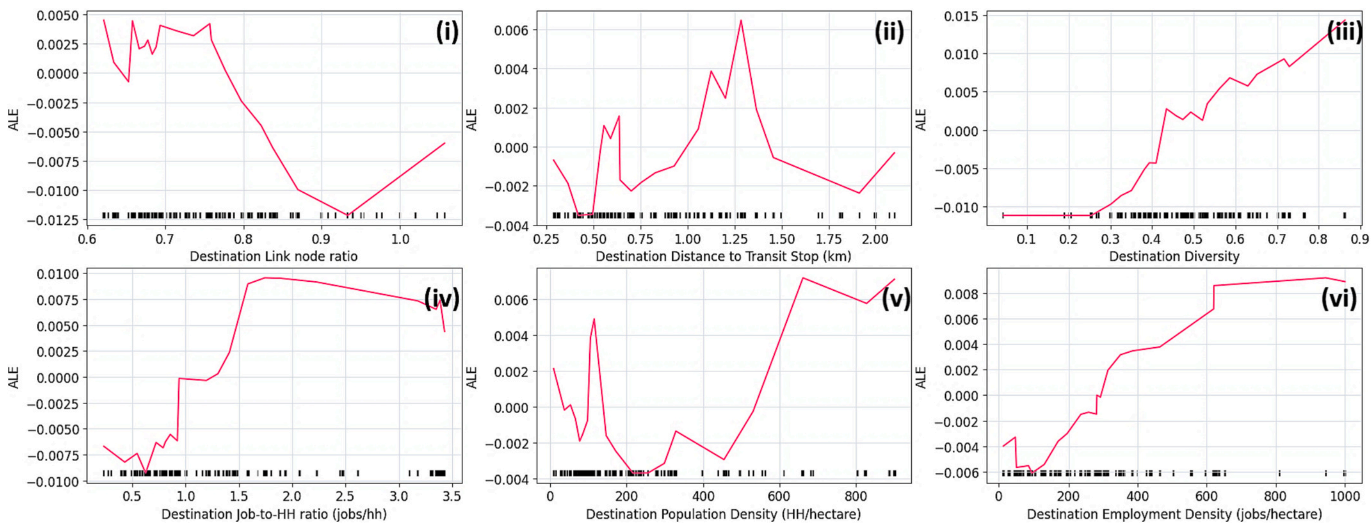

Fig. 3. Nonlinear effects of BE at workplace based on ALE plots.

use. These density findings also indirectly suggest the need to have jobs and housing located close together in order to reduce the use of private cars. In the context, the highest density brings about at least two contextual issues. The city's population has risen over the previous few decades, and with it, the city's economy. However, it is evident that the quality of the public transportation system is substandard. Consequently, individuals with higher incomes are inclined to acquire personal vehicles and rely on them for mobility (Ashik et al., 2023; Barua et al., 2013). Additionally, when cities grow in population and area, the commuting distance increases, reducing the number of trips taken by walk and increasing the number of journeys taken by car (Ashik et al., 2023; Goel and Mohan, 2020).

4.4. Interaction effects of the BE

4.4.1. Interaction effects between multiple BE indicators

Friedman's H-statistic, shown in Fig. 4, reflects whether and to what extent individual BE indicators interact with each other in the RF model by measuring the interaction strength between them. Fig. 4a demonstrates, for example, that the interaction between population density and employment density at residence (at work) accounts for 11.5% (around 7%) of the explained variation of commute mode choice, which is the strongest degree of interaction strength. In general, the interaction effects of individual BE measures at home are greater than those at work in explaining mode use (Fig. 4). In order to better understand the interaction effects, we plotted six interactions, shown in Fig. 5, with the highest level of interaction intensity (At home: population density-employment density- private car use; distance to transit stop-job household ratio- private car use; and diversity-population density- private car use. At work: population density-employment density- private car use; link node ratio-population density- private car use; and job household ratio-employment density- private car use). Although Friedman's H-statistic was calculated considering all sorts of modes, for specific reasons, as we mentioned in the previous section 4.3, we solely focus on private car trips for a detailed discussion of the interaction effects.

The interaction impact between population density, land use diversity, and private car use at residence is illustrated in Fig. 5a. It shows that land use diversity is generally associated with lower private car use and the opposite pattern for population density, where higher population density is correlated with higher private car use. Their interaction reveals that a diverse land use pattern (entropy index >0.4) counteracts the favorable effect of population density on private car use. When a person chooses to live in a neighborhood with a diverse land use pattern (entropy index = 0.6) and greater population density (500 households/ha) over a neighborhood with lower land use diversity (entropy index = 0.3) and the same population density, the likelihood of using a private car drops by around 18%. Their interaction also demonstrates that increasing land-use diversity is more beneficial in lowering private car use when population density is <500 households per hectare. This means that increased population density weakens the effect of land use diversification in lowering private car use. Similarly, the interaction between link node ratio, population density, and private car use at work shows that a well-connected road network offsets the favorable effect of population density, and that the efficiency of a well-connected road network is achieved in employment centers with densities of <500 households per hectare (Fig. 5d). Fig. 5d illustrates that having a modest population density boosts the effect of the connectedness of the road network in lowering private car use.

The interaction effect of population density, employment density, and private car use at home demonstrates that moderate density and balance between these two density measures are helpful in lowering private car use (Fig. 5b). When a resident lives in a moderately densified neighborhood (population density 500 households/ha and employment density 500 jobs/ha) compared to a highly densified neighborhood (population density > 700 households/ha and employment density > 500 jobs/ha), the likelihood of driving decreases by about 10–21%. Similarly, Fig. 5e depicts a somewhat moderate population and job density, and equilibrium between these two density measures at work is favorable for reducing private car use. At work, moderate employment density and balance in job household distribution are also necessary to lower private car use (Fig. 5f).

According to Fig. 5c at residence, even if a neighborhood is closer to a transit stop, it will not be able to reduce rather than maximize automobile use unless the neighborhood has sufficient job accessibility. For example, a resident's likelihood of driving is 9.8% when he lives in a neighborhood with a shorter distance (600-800 m) to a transit stop but lower job accessibility (job household ratio = 1), but the probability reduces when he lives in a neighborhood with higher job accessibility (job household ratio > 2).

4.4.2. Interaction effects between BE and commute time and BE and car ownership

Friedman's H-statistic illustrated in **Fig. 6** shows that commute time and car ownership have considerable interactions with BE both at residence and workplace locations. Cumulatively, commuting time-BE interaction and car ownership-BE interaction account for 30.35% and 26.21% of the explained variations in commute mode choice behavior,

Fig. 4. The interaction strength between individual BE indicators at residence and workplace, (4a) and (4b), respectively at the RF model.

respectively. However, we observed that the residential BE interactions better explained how respondents chose commute mode. Considering such results, we further examined the effects of the top three interactions on car-based commuting trips by plotting two-way interactions as illustrated in Fig. 7. These figures further explain how these interactions influence private car use for commuting trips in our study area.

Fig. 7(a) shows higher employment density (at residence) and longer commute time result in increased private car use probability. Their interaction demonstrates that when commute duration exceeds 40 min, the effect of higher employment density (> 500 jobs/ha) on private car choice is amplified. In contrast, when a commuter lives in a neighborhood with an employment density of <500 jobs/ha, the effect of a longer commute time (>40 min) is offset. Consecutively, only when there are fewer than 500 jobs/ha and commute time is <20 min, commuting time synergizes with employment density in reducing private car usage probability. Similarly, Fig. 7(b) indicates that job household ratio (at residence) above 1.5 and commuting time of <20 min have a synergistic influence on decreasing private car use. Furthermore, an important observation is confirmed in Fig. 7(c), which shows that even though residents live closer to a transit stop, their likelihood of choosing a private car increases dramatically if their commute is >40 min. But shorter commuting time (<20 min) results in lower car choice probability regardless of the distance to a transit stop. Overall, interactions among commuting time, BE at residence, and car choice probabilities indicate that they have synergistic or counteractive influences on car choice probabilities for respondents in our case study area.

The interaction effects of car ownership and BE measures at residence on car choice probabilities are presented in Fig. 7. Figs. 7 (d) and (f) show that even though residents of a car-owning household are close to a public transportation station (e.g., within 1.2 km) or reside in a well-connected neighborhood, they are more likely to use their cars. When car-owning commuters dwell in an area with a job-to-household ratio smaller than 1.5, the effect of car ownership on private car usage is considerably amplified, Fig. 7 (e). This suggests that car ownership strongly affects the use of private vehicles for commuting, even though car owners' residential BE characteristics are friendly to walking and

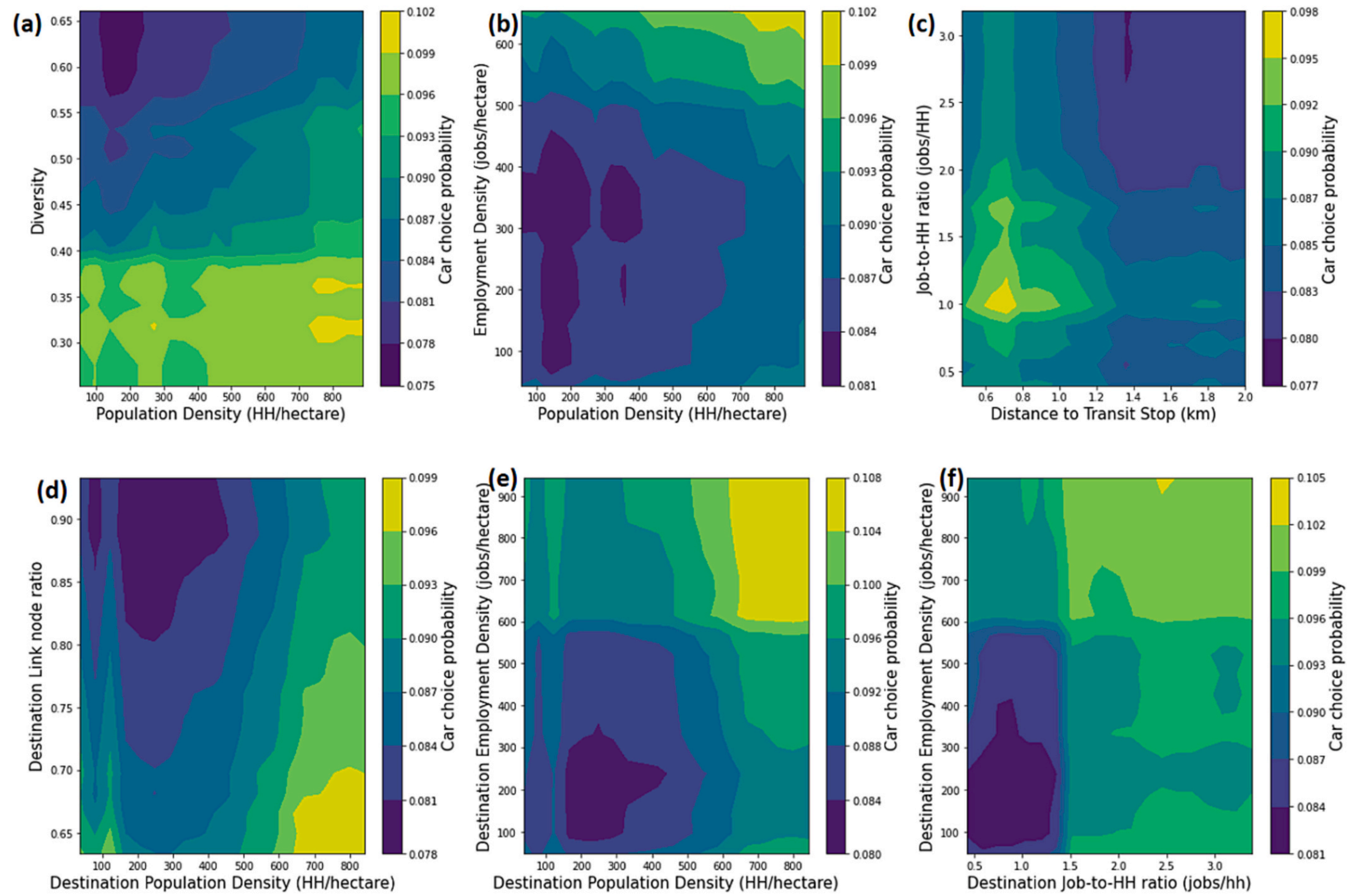

Fig. 5. Two-way interaction effects of BE measures at residential (5a, 5b, and 5c) and workplace (5d, 5e, and 5f) locations for private car choice.

public transit.

5. Discussion

5.1. Relative importance of factors influencing commuting mode choice

This study applied several explainable machine learning techniques and tools to demystify the complex relationships between BE and computer mode choice in a global south mega city. The most important explanatory variable to explain mode choice behavior in our research area is *commute time*; this conclusion is consistent with Cheng et al. (2019) and Hagenauer and Helbich (2017). Car ownership and a driver's license availability are the next two major determinants of commuting mode. While the dominant impact of commuting time on commuter mode choice is well known, the dominant role of car ownership in explaining commuter behavior is rarely discussed in the existing machine learning research (Cheng et al., 2019; Ding et al., 2018a; Wu et al., 2019). In our instance, the majority of Dhaka residents who own cars utilize them almost exclusively for commuting. This suggests that increasing automobile ownership in a megacity like Dhaka may lead to people becoming more car-oriented in the near future if effective steps to minimize car ownership or boost the attractiveness of other modes for a modal shift are not adopted. This is critical as the RSTP-2015 indicated that as the economy expands in Bangladesh, automobile ownership will also increase in major urban centers in the longer term.

Our findings suggest that built environment (BE) indicators considerably impact predicting commute mode choice in the dense urban context of a mega city located in the global south, accounting for approximately 45% of the direct effect when controlling for socio-demographic and trip factors. Further, commuting duration and car ownership are two of the most influential single variables determining mode choice; their interaction with the BE also plays key role in predicting mode choice. This observation is in line with the findings of Ding et al. (2018b); they also analyzed commuting behavior using machine learning methods. Our findings, however, differ from those of several other studies that employed classic parametric modeling to determine the relative importance of built and socio-demographic factors in commuting mode choice (Singh et al., 2018; Stead, 2001). Such distinctions provide new insight into model selection. Modeling relationships between BE indicators and mode choice with different modeling approaches (machine learning vs. parametric statistical modeling) might yield different results. This is significant because it implies that studies should critically evaluate modeling approaches in order to fully comprehend the cumulative effects of BE on mode choice.

Furthermore, we observed that the residential BE is slightly more important than the workplace BE to explain commuting mode choice in the case of Dhaka. This conclusion differs from previous research in Western and European cities (Chen et al., 2008; Ding et al., 2018b), but is similar to research in East Asia cities such as in China (Sun et al., 2017). We speculate that commuters of Western countries prefer to reside in the periphery, and their workplace location is generally in the city center. In contrast, in developing countries, like Dhaka, commuters are more likely to live near the city centers (Dhaka has a polycentric city structure) though they work in the city's outskirts (Ashik et al., 2020, 2024; Choudhury and Ayaz, 2015). While in both scenarios, the BE in the city center is frequently designed to limit car use (Ashik et al., 2022; Kamruzzaman et al., 2014; Kumar et al., 2018). Based on these distinct spatial contexts, we argue that the observed disparity in the BE's role between origin and destination is logical in our study area. We noted that, even though the BE of the workplace has a considerable impact in

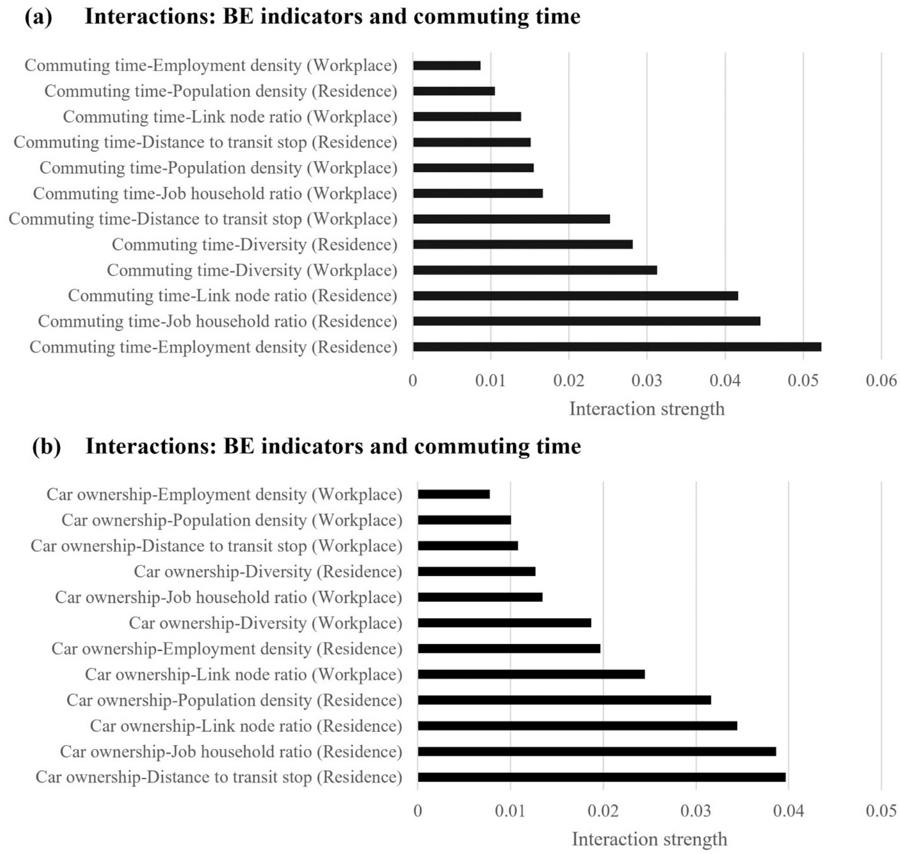

Fig. 6. (a) the interaction strength between BE indicators and commuting time; (b) BE indicators and car ownership interaction strength in predicting commute mode choice.

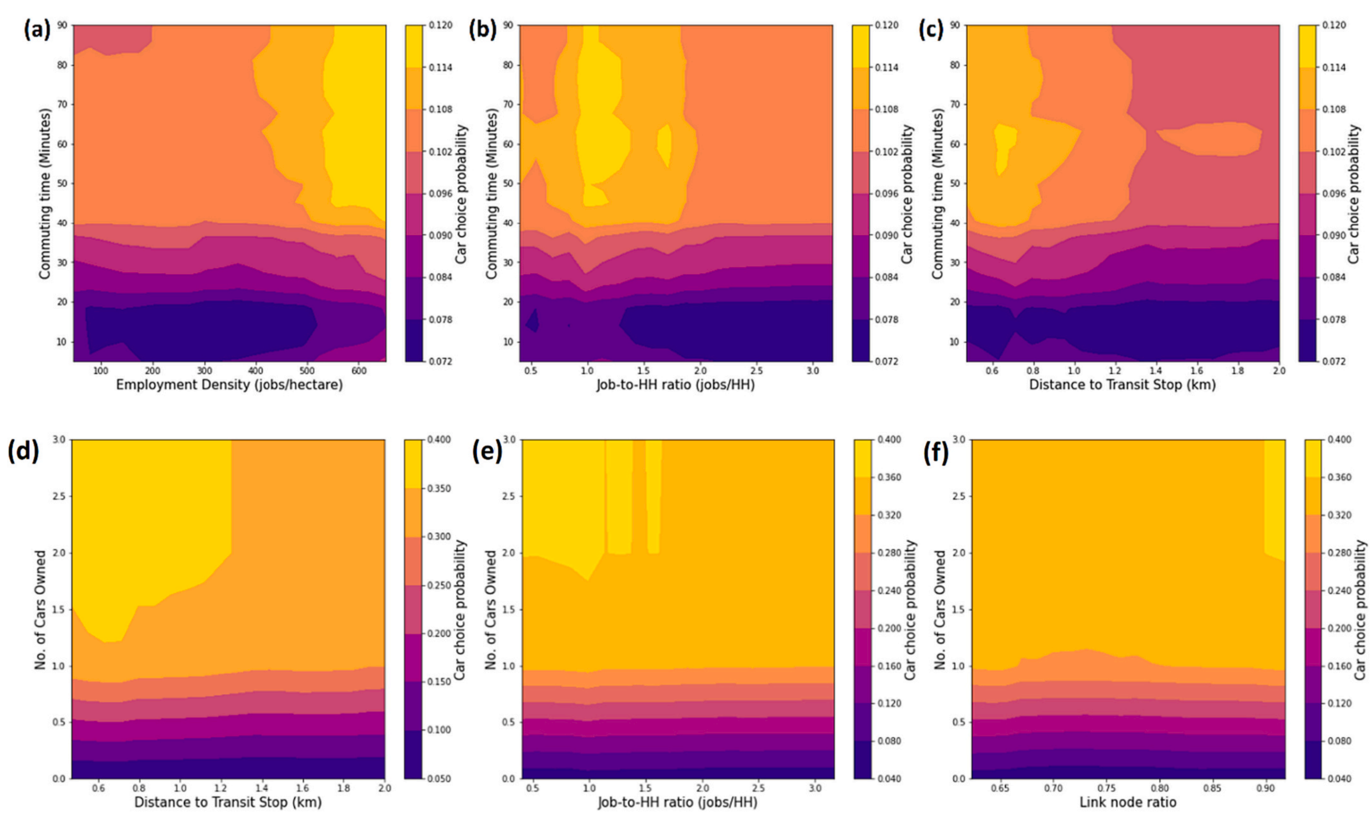

Fig. 7. Two-way interaction effects of BE measures, commute time and car ownership for private car choice.

predicting mode choice, the bulk of existing research largely ignores workplace attributes in their modeling framework. Based on our observations in this study, it should be noted that disregarding workplace BE attributes in explaining commuter mode choice may result in omitted variable bias in several existing modeling frameworks that only consider residential BE attributes.

5.2. Nonlinearity and interaction of the BE on sustainable mode use

The findings of this study have several key findings, such as identifying the nonlinear associations and interaction links between the BE and commuting mode choice. We observed that all the BE indicators had nonlinear relations with different modes used for commuting. Private car use does not display a constant return of scale with land-use diversity, distance to the transit stops, street connectivity, job household ratio, population density, and employment density for different ranges of values for each indicator. Notably, we observed for residential land use diversity, the effect of diversity on lowering car use may only be observed after a threshold limit of 0.35 (which is at least two land use types). This indicates that to lower the probability of car choice, residential areas should have more than two land use types. Similarly, we also observed that other BE characteristics also have a precise threshold limit at which they can have the greatest impact on reducing private car use (see Section 4.3). These observations are also consistent with recent studies by Ding et al. (2021) and Wu et al. (2019).

Our study clearly indicated that the relationships between the BE and mode choice are primarily nonlinear. The assumption of travel behavior has constant returns to scale with BE measures in the conventional modeling framework, which is usually unable to capture such nonlinear relationships (Ribeiro et al., 2019). Such a phenomenon may explain the inconsistent outcomes of BE effects on commute mode choice in the literature (Aston et al., 2021; Ewing and Cervero, 2010; Stevens, 2017). Therefore, we argue that assuming linearity in understanding the impact of the BE on travel behavior modeling might not always reflect the reality as usually understood among existing transport planning studies.

Furthermore, our results illustrated that the interaction effects in the residential environment have greater explanatory power in explaining the variability of commuting mode use. Although population and employment density are the two least important individual BE indicators (Table 5), their interactions with one another as well as with other BE indicators are critical in explaining commute mode choice heterogeneity (Fig. 4). That means density measures should be planned together, connecting multiple BE indicators when designing or retrofitting an area to promote a sustainable mode for commuting. Our evidence suggests that the effect of one BE measure may be strengthened or weakened or counteracted by the presence of another BE measure, which may provide critical insights into mode choice behavior and neighborhood design. Synergistic or counteractive effects also occur when BE measures interact with commute time or car ownership. A commuter's commute time is interactive with the BE in which the commuter lives or works.

5.3. Machine learning algorithm compared to the traditional modeling

We compared the model performance of RF, GBDT, and SVM and discovered that RF performed better in modeling commuting mode selection. Similar results were also observed by Cheng et al. (2019) and Hagenauer and Helbich (2017). Our paper contributes by comparing two ensemble classifiers, RF and GBDT, in predicting mode choice, whereas previous studies compared RF or GBDT with only other individual classifiers (Koushik et al., 2020). Following the evaluation of model performance, we used the RF model to investigate and predict the relative importance, complicated nonlinear effect, and interaction effect of the BE on commuting mode usage. Conventional models, on the other hand, would not be able to identify such key findings because they are based on rigid assumptions of linearity among the features. For data with high complexity, the traditional statistical models also require inherent knowledge about the qualities of the features as well as human intervention, and these models mimic individual mechanisms rather than aggregated ones (Li et al., 2021). However, with increasing data availability from multiple sources, machine learning algorithms provide greater insights in revealing the complicated and delicate relationships between the BE and commute mode choice, therefore having more significant potential for future studies.

Although our machine learning algorithms have adequate predictive capacities and the ability to identify relative importance and nonlinearity, the algorithms cannot identify the statistical significance of the identified association, as found in the traditional regression modeling. Furthermore, the outcomes of relative importance, nonlinearity, and interactions are challenging to interpret without contextual knowledge (Cheng et al., 2019). Despite these challenges, machine learning provides an advanced framework to understand the complexities of predicting commute mode choice.

5.4. Practice and policy implications

The findings of our study have several potential policy and practice implications. We provided several practice and policy recommendations for Dhaka based on our findings and the policy aims of the Revised Strategic Transport Plan-2015. The study's results can be used to inform the development of transit-oriented development (TOD) in Dhaka. The effects of private car use probability identified in the nonlinearity analysis (section 4.3) can guide TOD development plans by providing information on the optimal radius from a transit stop (750 m) and densities (maximum population density of 500 households/ha and employment density of 500 jobs/ha) to decrease private car usage.

Several mass-rapid transits and bus-rapid transit initiatives are currently in the works. Our findings could help guide the development of TODs around these new transit facilities in the future. The relative importance of socio-demographic and built-environment elements we identified in this study can be used to determine where TODs should be located, considering the combination of socio-demographic and built-environment features that are necessary for successful TODs along transit lines. Consideration should be given to locating TODs in areas that provide low and middle-income groups with greater access to public transit, making it easy and safe for female commuters to access transit modes. Additionally, car ownership plays a crucial role in determining an individual's likelihood of using private cars; TOD policies should be primarily earmarked for individuals owning a car. For instance, in the case of Dhaka, Travel Demand Management (TDM) measures such as higher parking charges, congestion pricing, and low emission zoning should be introduced and integrated into TOD development framework so that people having a car become more responsible for cutting down their car use. Previously Labib et al. (2018) noted for Dhaka, in certain central business areas, low emission zoning can be implemented to reduce car traffic and related air pollutants. The RSTP-2015 also noted congestion charging could be an effective solution in reducing care usage in central business districts. Furthermore, based on our findings, we argue that planners should focus more on developing a low-carbon oriented BE in TOD neighborhoods, as BE has significant effects on reducing car ownership (Ding and Cao, 2019) and car use. The threshold effects of BE on car use probability we identified can offer nuanced guidance.

Furthermore, our study has important policy implications for urban density in cities like Dhaka. Our findings suggest that increasing densities may be ineffective in decreasing private car use in previously congested environments. This is supported by previous research that found that density has minimal effects on travel behavior in Indian cities (Goel and Mohan, 2020). Furthermore, growing density will increase commute distance in a comparatively low-density city, Toulouse, in 2100 (Masson et al., 2014), like the problems faced by current high-density cities. As a result, planners and policymakers should create specific strategies for long-term implications while formulating land use plans, given that cities are rapidly urbanizing, to comprehend the irreversible impacts of land use policies, such as whether the policies we are pursuing today will produce beneficial results in the future (Gonçalves and Portugal, 2020; Kuang et al., 2019). It should be noted that our study does not imply that density should not be promoted now, but it does unequivocally suggest that if a city continues to become more appealing and expanded, this policy should be adjusted by maintaining a balance between population density and job density based on their interaction and threshold point.

5.5. Limitations of this study

There are several areas in which this study could be further developed. Firstly, we were unable to fully capture residential self-selection since we were unable to account for commuting mode preferences. Self-selection may affect commute mode choice (Ettema and Nieuwenhuis, 2017; Guan et al., 2020), and a detailed exploration of the impact of self-selection on commuter travel behavior requires further investigation. Secondly, despite measuring BE variables at the TAZ level, we did not account for spatial dependencies within different travel zones, which may have inflated our observed effects. We believe future studies on spatially explicit machine learning modeling would account for such issues. Thirdly, although our data has thousands of observations, these were collected in a cross-sectional design. Therefore, we cannot confirm the causal effect of the BE on commuting mode selection. This limitation indicates the need for further studies focusing on exploring longitudinal study design and advanced spatial modeling to better understand the effects of the BE on commuting mode choice.

6. Conclusion

In this study, we investigated the influence of the BE on commuter travel behavior using a data-driven machine learning approach for a dense mega city context located in the global south. We focused on exploring automobile-based commuting patterns in Dhaka and the potential of BE modifications in lowering such commuter trips. Using the RF algorithm, we identified the relative importance of the BE in predicting commuting mode choice. We observed that BE indicators could explain the mode choice probability to a more considerable extent (around 45%) than socio-demographic factors. We also observe that BE at residence is slightly more influential than the workplace on choice of commute mode. Our nonlinearity and interaction analyses also identified the nonlinear relations between BEs and commuting mode choice in our case study context. Furthermore, they showed complex interactions between BEs and commuting travel patterns: the effect of one BE measure is strengthened or, weakened or counteracted by the presence of another BE measure. Our study illustrated the importance, directions, and multidimensional influences of the BE on commuter travel behavior. Modifying BE variables (e.g., land use diversity, street connectivity) within specific ranges has the potential to lower private car-based commuting trips and promote the use of active and transit modes in Dhaka, although the city is currently shifting more towards car-centric mobility pattern. Their interaction impact helps us determine the ideal mix of built-environment interventions to reduce private car-based commuting trips in a dense city context. These findings could be utilized to develop sustainable transportation plans and policies that emphasize transit-oriented development, travel demand management, and integrated land-use transportation planning in cities global south, which are observing increased car usage.

CRediT authorship contribution statement

F.R. Ashik: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **A.I.Z. Sreezon:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **M.H. Rahman:** Conceptualization, Data curation, Formal analysis, Methodology, Writing – original draft, Writing – review & editing. **N.M. Zafri:** Conceptualization, Data curation, Methodology, Writing – original draft, Writing – review & editing. **S.M. Labib:** Conceptualization, Investigation, Methodology, Supervision, Validation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

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

Data availability

The code can be accessed at https://github.com/Spatial-Data-Science-and-GEO-AI-Lab/BuiltEnvModeChoice. We will make the data available upon request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jtrangeo.2024.103828.

References

Apley, D.W., Zhu, J., 2020. Visualizing the effects of predictor variables in black box supervised learning models. Journal of the Royal Statistical Society Series B: Statistical Methodology 82 (4), 1059–1086.
Ashik, F.R., Islam, M.S., Alam, M.S., Tabassum, N.J., Manaugh, K., 2024. Dynamic equity in urban amenities distribution: An accessibility-driven assessment. Applied Geography 164, 103199.
Ashik, F.R., Mim, S.A., Neema, M.N., 2020. Towards vertical spatial equity of urban facilities: an integration of spatial and aspatial accessibility. J. Urban Manag. 9 (1), 77–92. https://doi.org/10.1016/j.jum.2019.11.004.
Ashik, F.R., Rahman, M.H., Kamruzzaman, M., 2022. Investigating the impacts of transit-oriented development on transport-related CO2 emissions. Transp. Res. Part D: Transp. Environ. 105, 103227 https://doi.org/10.1016/j.trd.2022.103227.
Ashik, F.R., Rahman, M.H., Antipova, A., Zafri, N.M., 2023. Analyzing the impact of the built environment on commuting-related carbon dioxide emissions. Int. J. Sustain. Transp. 17 (3), 258–272. https://doi.org/10.1080/15568318.2022.2031356.
Aston, L., Currie, G., Kamruzzaman, M., Delbosc, A., Teller, D., 2020. Study design impacts on built environment and transit use research. J. Transp. Geogr. 82, 102625 https://doi.org/10.1016/j.jtrangeo.2019.102625.
Aston, L., Currie, G., Delbosc, A., Kamruzzaman, M., Teller, D., 2021. Exploring built environment impacts on transit use – an updated meta-analysis. Transp. Rev. 41 (1), 73–96. https://doi.org/10.1080/01441647.2020.1806941.
Axhausen, K.W., 2015. Agent-based or agent based modelling: reflections on choices, constraints and commitments. In: 14th International Conference on Travel Behavior Research (IATBR 2015), Windsor, United Kingdom, July 19-23, 2015. http://hdl.handle.net/20.500.11850/102482.
Bangladesh Bureau of Statistics (BBS), 2011. District Statistics 2011 Dhaka (2011). Available at http://203.112.218.65:8008/WebTestApplication/userfiles/Image/District%20Statistics/Dhaka.pdf.
Barua, S., Alam, D., Roy, A., 2013. Modal integration for improving urban mobility in Dhaka. In: Urban Public Transportation Systems 2013, pp. 179–192.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.
Burman, P., 1989. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. Biometrika 76 (3), 503–514. https://doi.org/10.1093/biomet/76.3.503.
Castañeda, P., 2021. Cycling case closed? A situated response to Samuel Nello-Deakin's "environmental determinants of cycling: not seeing the forest for the trees?". J. Transp. Geogr. 90, 102947 https://doi.org/10.1016/j.jtrangeo.2020.102947.
Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. 16, 321–357.
Chen, C., Gong, H., Paaswell, R., 2008. Role of the built environment on mode choice decisions: additional evidence on the impact of density. Transportation 35 (3), 285–299. https://doi.org/10.1007/s11116-007-9153-5.
Cheng, L., Chen, X., De Vos, J., Lai, X., Witlox, F., 2019. Applying a random forest method approach to model travel mode choice behavior. Travel Behav. Soc. 14, 1–10. https://doi.org/10.1016/j.tbs.2018.09.002.
Choudhury, C.F., Ayaz, S.B., 2015. Why live far? — insights from modeling residential location choice in Bangladesh. J. Transp. Geogr. 48, 1–9. https://doi.org/10.1016/j.jtrangeo.2015.08.001.

Church, E., Szibbo, N., Elliott, R., Cranz, G., 2012. Cities and People Project: A White Paper on Human Interaction with the Built Environment. Institute for Environmental Entrepreneurship, Berkeley, CA, pp. 1–67.

Dill, J., 2004. Measuring network connectivity for bicycling and walking. In: 83rd Annual Meeting of the Transportation Research Board, Washington, DC, USA.

Ding, C., Cao, X., 2019. How does the built environment at residential and work locations affect car ownership? An application of cross-classified multilevel model. J. Transp. Geogr. 75, 37–45. https://doi.org/10.1016/j.jtrangeo.2019.01.012.

Ding, C., Wang, D., Liu, C., Zhang, Y., Yang, J., 2017. Exploring the influence of built environment on travel mode choice considering the mediating effects of car ownership and travel distance. Transp. Res. A Policy Pract. 100, 65–80. https://doi.org/10.1016/j.tra.2017.04.008.

Ding, C., Cao, X., Næss, P., 2018a. Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. Transp. Res. A Policy Pract. 110, 107–117. https://doi.org/10.1016/j.tra.2018.02.009.

Ding, C., Cao, X., Wang, Y., 2018b. Synergistic effects of the built environment and commuting programs on commute mode choice. Transp. Res. A Policy Pract. 118, 104–118. https://doi.org/10.1016/j.tra.2018.08.041.

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. Transp. Res. A Policy Pract. 148, 22–35. https://doi.org/10.1016/j.tra.2021.03.021.

DTCA, 2015. The Project on the Revision and Updating of the Strategic Transport Plan for Dhaka.

Enam, A., Choudhury, C.F., 2011. Methodological issues in developing mode choice models for Dhaka, Bangladesh. Transp. Res. Rec. 2239 (1), 84–92. https://doi.org/10.3141/2239-10.

Etminani-Ghasrodashti, R., Ardeshiri, M., 2016. The impacts of built environment on home-based work and non-work trips: an empirical study from Iran. Transp. Res. A Policy Pract. 85, 196–207. https://doi.org/10.1016/j.tra.2016.01.013.

Ettema, D., Nieuwenhuis, R., 2017. Residential self-selection and travel behaviour: what are the effects of attitudes, reasons for location choice and the built environment? J. Transp. Geogr. 59, 146–155. https://doi.org/10.1016/j.jtrangeo.2017.01.009.

Ewing, R., Cervero, R., 2010. Travel and the built environment. J. Am. Plan. Assoc. 76 (3), 265–294. https://doi.org/10.1080/01944361003766766.

Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29 (5), 1189–1232. https://doi.org/10.1214/aos/1013203451.

Friedman, J.H., Popescu, B.E., 2008. Predictive learning via rule ensembles. Ann. Appl. Stat. 2 (3), 916–954. https://doi.org/10.1214/07-AOAS148.

Goel, R., Mohan, D., 2020. Investigating the association between population density and travel patterns in Indian cities—an analysis of 2011 census data. Cities 100, 102656. https://doi.org/10.1016/j.cities.2020.102656.

Gonçalves, F.D.S., Portugal, L.D.S., 2020. Traffic impact studies committed to sustainability: the case of Rio de Janeiro. J. Environ. Manag. 253, 109573 https://doi.org/10.1016/j.jenvman.2019.109573.

Guan, X., Wang, D., Jason Cao, X., 2020. The role of residential self-selection in land use-travel research: a review of recent findings. Transp. Rev. 40 (3), 267–287. https://doi.org/10.1080/01441647.2019.1692965.

Guo, X., Yin, Y., Dong, C., Yang, G., Zhou, G., 2008. On the class imbalance problem. In: 2008 Fourth International Conference on Natural Computation (18-20 Oct. 2008).

Hagenauer, J., Helbich, M., 2017. A comparative study of machine learning classifiers for modeling travel mode choice. Expert Syst. Appl. 78, 273–282. https://doi.org/10.1016/j.eswa.2017.01.057.

Haider, M.Z., Papri, R.S., 2021. Cost of traffic congestion in Dhaka Metropolitan City. Public Transp. 13 (2), 287–299. https://doi.org/10.1007/s12469-021-00270-4.

Handy, S., 2017. Thoughts on the meaning of mark Stevens's Meta-analysis. J. Am. Plan. Assoc. 83 (1), 26–28. https://doi.org/10.1080/01944363.2016.1246379.

Handy, S.L., Boarnet, M.G., Ewing, R., Killingsworth, R.E., 2002. How the built environment affects physical activity: views from urban planning. Am. J. Prev. Med. 23 (2, Supplement 1), 64–73. https://doi.org/10.1016/S0749-3797(02)00475-0.

Hanson, S., 2010. Gender and mobility: new approaches for informing sustainability. Gend. Place Cult. 17 (1), 5–23. https://doi.org/10.1080/09663690903498225.

Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H., 2009. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, vol. 2. springer, New York, pp. 1–758.

Hatami, F., Rahman, M.M., Nikparvar, B., Thill, J.C., 2023. Non-linear associations between the urban built environment and commuting modal split: a random forest approach and SHAP evaluation. IEEE Access 11, 12649–12662.

Hillel, T., Bierlaire, M., Elshafie, M.Z.E.B., Jin, Y., 2021. A systematic review of machine learning classification methodologies for modelling passenger mode choice. J. Choice Model. 38, 100221 https://doi.org/10.1016/j.jocm.2020.100221.

Household Income and Expenditure Survey (HIES), 2016. Preliminary Report on Household Income and Expenditure Survey 2016. Available at: https://catalog.ihsn.org/index.php/catalog/7399.

Hripcsak, G., Rothschild, A.S., 2005. Agreement, the F-measure, and reliability in information retrieval. J. Am. Med. Inform. Assoc. 12 (3), 296–298. https://doi.org/10.1197/jamia.M1733.

Hsu, C.-W., Chang, C.-C., Lin, C.-J., 2003. A practical guide to support vector classification. In: Taipei, Taiwan.

Islam, M.R., Saphores, J.-D.M., 2022. An L.a. story: the impact of housing costs on commuting. J. Transp. Geogr. 98, 103266 https://doi.org/10.1016/j.jtrangeo.2021.103266.

Jakkula, V., 2006. Tutorial on support vector machine (svm). School of EECS, Washington State University, p. 37. In.

Junker, M., Hoch, R., Dengel, A., 1999. On the evaluation of document analysis components by recall, precision, and accuracy. In: Proceedings of the Fifth International Conference on Document Analysis and Recognition. ICDAR '99 (Cat. No.PR00318) (22-22 Sept. 1999).

Kamruzzaman, M., Baker, D., Washington, S., Turrell, G., 2013. Residential dissonance and mode choice. J. Transp. Geogr. 33, 12–28. https://doi.org/10.1016/j.jtrangeo.2013.09.004.

Kamruzzaman, M., Baker, D., Washington, S., Turrell, G., 2014. Advance transit oriented development typology: case study in Brisbane, Australia. J. Transp. Geogr. 34, 54–70. https://doi.org/10.1016/j.jtrangeo.2013.11.002.

Kamruzzaman, M., Washington, S., Baker, D., Brown, W., Giles-Corti, B., Turrell, G., 2016. Built environment impacts on walking for transport in Brisbane, Australia. Transportation 43 (1), 53–77. https://doi.org/10.1007/s11116-014-9563-0.

Koushik, A.N.P., Manoj, M., Nezamuddin, N., 2020. Machine learning applications in activity-travel behaviour research: a review. Transp. Rev. 40 (3), 288–311. https://doi.org/10.1080/01441647.2019.1704307.

Kuang, Y., Yen, B.T.H., Suprun, E., Sahin, O., 2019. A soft traffic management approach for achieving environmentally sustainable and economically viable outcomes: an Australian case study. J. Environ. Manag. 237, 379–386. https://doi.org/10.1016/j.jenvman.2019.02.087.

Kumar, P.P., Sekhar, C.R., Parida, M., 2018. Residential dissonance in TOD neighborhoods. J. Transp. Geogr. 72, 166–177. https://doi.org/10.1016/j.jtrangeo.2018.09.005.

Labib, S.M., Neema, M.N., Rahaman, Z., Patwary, S.H., Shakil, S.H., 2018. Carbon dioxide emission and bio-capacity indexing for transportation activities: a methodological development in determining the sustainability of vehicular transportation systems. J. Environ. Manag. 223, 57–73. https://doi.org/10.1016/j.jenvman.2018.06.010.

Li, L., Rong, S., Wang, R., Yu, S., 2021. Recent advances in artificial intelligence and machine learning for nonlinear relationship analysis and process control in drinking water treatment: a review. Chem. Eng. J. 405, 126673 https://doi.org/10.1016/j.cej.2020.126673.

Li, Y., Yao, E., Liu, S., Yang, Y., 2024. Spatiotemporal influence of built environment on intercity commuting trips considering nonlinear effects. J. Transp. Geogr. 114, 103744.

Lindner, A., Pitombo, C.S., Cunha, A.L., 2017. Estimating motorized travel mode choice using classifiers: an application for high-dimensional multicollinear data. Travel Behav. Soc. 6, 100–109. https://doi.org/10.1016/j.tbs.2016.08.003.

Ma, X., Ding, C., Luan, S., Wang, Y., Wang, Y., 2017. Prioritizing influential factors for freeway incident clearance time prediction using the gradient boosting decision trees method. IEEE Trans. Intell. Transp. Syst. 18 (9), 2303–2310. https://doi.org/10.1109/TITS.2016.2635719.

Mannering, F., Bhat, C.R., Shankar, V., Abdel-Aty, M., 2020. Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis. Anal. Methods Accident Res. 25, 100113.

Manoj, M., Verma, A., 2016. Effect of built environment measures on trip distance and mode choice decision of non-workers from a city of a developing country, India. Transp. Res. Part D: Transp. Environ. 46, 351–364. https://doi.org/10.1016/j.trd.2016.04.013.

Masson, V., Marchadier, C., Adolphe, L., Aguejdad, R., Avner, P., Bonhomme, M., Bretagne, G., Briottet, X., Bueno, B., de Munck, C., Doukari, O., Hallegatte, S., Hidalgo, J., Houet, T., Le Bras, J., Lemonsu, A., Long, N., Moine, M.P., Morel, T., Zibouche, K., 2014. Adapting cities to climate change: a systemic modelling approach. Urban Clim. 10, 407–429. https://doi.org/10.1016/j.uclim.2014.03.004.

Molnar, C., 2020. Interpretable Machine Learning. Lulu. com.

Munshi, T., 2016. Built environment and mode choice relationship for commute travel in the city of Rajkot, India. Transp. Res. Part D: Transp. Environ. 44, 239–253. https://doi.org/10.1016/j.trd.2015.12.005.

Nakshi, P., Debnath, A.K., 2020. Impact of built environment on mode choice to major destinations in Dhaka. Transp. Res. Rec. 0361198120978418 https://doi.org/10.1177/0361198120978418.

Nguyen, T.M., Kato, H., Phan, L.B., 2020. Is built environment associated with travel mode choice in developing cities? Evidence from Hanoi. Sustainability 12 (14).

Noble, W.S., 2006. What is a support vector machine? Nat. Biotechnol. 24 (12), 1565–1567. https://doi.org/10.1038/nbt1206-1565.

Panter, J., Guell, C., Humphreys, D., Ogilvie, D., 2019. Title: can changing the physical environment promote walking and cycling? A systematic review of what works and how. Health Place 58, 102161. https://doi.org/10.1016/j.healthplace.2019.102161.

Powers, D.M.W., 2020. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.

Rahman, M.H., Antipova, A., 2024. Structural equation model in exploring urban sprawl and its impact on commuting time in 162 US urbanized areas. Cities 148, 104855.

Rahman, M.H., Ashik, F.R., 2020. Is neighborhood level jobs-housing balance associated with travel behavior of commuters?: a case study on Dhaka City, Bangladesh. GeoScape 14 (2), 122–133. https://doi.org/10.2478/geosc-2020-0011.

Rahman, M.H., Ashik, F.R., Mouli, M.J., 2022. Investigating spatial accessibility to urban facility outcome of transit-oriented development in Dhaka. Transp. Res. Interdiscip. Perspect. 14, 100607 https://doi.org/10.1016/j.trip.2022.100607.

Rahman, M.M., Upaul, S., Thill, J.C., Rahman, M., 2023. Active transportation and the built environment of a mid-size global south city. Sustain. Cities Soc. 89, 104329.

RAJUK, 2015. Dhaka Structure Plan 2016–2035.

Ribeiro, H.V., Rybski, D., Kropp, J.P., 2019. Effects of changing population or density on urban carbon dioxide emissions. Nat. Commun. 10 (1), 3204. https://doi.org/10.1038/s41467-019-11184-y.

Rijsbergen, C.J.V., 1979. Information Retrieval. Butterworth-Heinemann.

Salmeron, R., García, C.B., García, J., 2018. Variance inflation factor and condition number in multiple linear regression. J. Stat. Comput. Simul. 88 (12), 2365–2384.

Shirgaokar, M., 2015. Expanding cities and vehicle use in India: differing impacts of built environment factors on scooter and car use in Mumbai. Urban Stud. 53 (15), 3296–3316. https://doi.org/10.1177/0042098015608050.

Singh, A.C., Astroza, S., Garikapati, V.M., Pendyala, R.M., Bhat, C.R., Mokhtarian, P.L., 2018. Quantifying the relative contribution of factors to household vehicle miles of travel. Transp. Res. Part D: Transp. Environ. 63, 23–36. https://doi.org/10.1016/j.trd.2018.04.004.

Somvanshi, M., Chavan, P., Tambade, S., Shinde, S.V., 2016. A review of machine learning techniques using decision tree and support vector machine. In: In 2016 International Conference on Computing Communication Control and Automation (ICCUBEA). IEEE, pp. 1–7.

Stead, D., 2001. Relationships between land use, socioeconomic factors, and travel patterns in Britain. Environ. Plan. B: Plan. Design 28 (4), 499–528. https://doi.org/10.1068/b2677.

Stevens, M.R., 2017. Does compact development make people drive less? J. Am. Plan. Assoc. 83 (1), 7–18. https://doi.org/10.1080/01944363.2016.1240044.

Sun, B., Ermagun, A., Dan, B., 2017. Built environmental impacts on commuting mode choice and distance: evidence from Shanghai. Transp. Res. Part D: Transp. Environ. 52, 441–453. https://doi.org/10.1016/j.trd.2016.06.001.

Thornton, L.E., Schroers, R.-D., Lamb, K.E., Daniel, M., Ball, K., Chaix, B., Kestens, Y., Best, K., Oostenbach, L., Coffee, N.T., 2022. Operationalising the 20-minute neighbourhood. Int. J. Behav. Nutr. Phys. Act. 19 (1), 15. https://doi.org/10.1186/s12966-021-01243-3.

Tong, S., Chang, E., 2001. Support vector machine active learning for image retrieval. In: Proceedings of the Ninth ACM International Conference On Multimedia, Ottawa, Canada. https://doi.org/10.1145/500141.500159.

Trading Economics, 2021. Bangladesh - Labor Force, Female. Retrieved October 21, 20201, from https://tradingeconomics.com/bangladesh/labor-force-female-percent-of-total-labor-force-wb-data.html.

Tran, M.T., Zhang, J., Chikaraishi, M., Fujiwara, A., 2016. A joint analysis of residential location, work location and commuting mode choices in Hanoi, Vietnam. J. Transp. Geogr. 54, 181–193. https://doi.org/10.1016/j.jtrangeo.2016.06.003.

van Wee, B., Handy, S., 2016. Key research themes on urban space, scale, and sustainable urban mobility. Int. J. Sustain. Transp. 10 (1), 18–24. https://doi.org/10.1080/15568318.2013.820998.

Wang, D., Zhou, M., 2017. The built environment and travel behavior in urban China: a literature review. Transp. Res. Part D: Transp. Environ. 52, 574–585. https://doi.org/10.1016/j.trd.2016.10.031.

Wang, Y., Yang, L., Han, S., Li, C., Ramachandra, T.V., 2017. Urban CO 2 emissions in Xi'an and Bangalore by commuters: implications for controlling urban transportation carbon dioxide emissions in developing countries. Mitig. Adapt. Strateg. Glob. Chang. 22, 993–1019.

Wang, X., Shao, C., Yin, C., Guan, L., 2020. Exploring the relationships of the residential and workplace built environment with commuting mode choice: a hierarchical cross-classified structural equation model. Transp. Lett. 14 (3), 274–281. https://doi.org/10.1080/19427867.2020.1857010.

Wang, X., Shao, C., Yin, C., Dong, C., 2021. Exploring the effects of the built environment on commuting mode choice in neighborhoods near public transit stations: evidence from China. Transp. Plan. Technol. 44 (1), 111–127. https://doi.org/10.1080/03081060.2020.1851453.

Wong, T.T., Yeh, P.Y., 2020. Reliable accuracy estimates from k-fold cross validation. IEEE Trans. Knowl. Data Eng. 32 (8), 1586–1594. https://doi.org/10.1109/TKDE.2019.2912815.

Wu, X., Tao, T., Cao, J., Fan, Y., Ramaswami, A., 2019. Examining threshold effects of built environment elements on travel-related carbon-dioxide emissions. Transp. Res. Part D: Transp. Environ. 75, 1–12. https://doi.org/10.1016/j.trd.2019.08.018.

Xu, L., Li, J., Brenning, A., 2014. A comparative study of different classification techniques for marine oil spill identification using RADARSAT-1 imagery. Remote Sens. Environ. 141, 14–23. https://doi.org/10.1016/j.rse.2013.10.012.

Ying, X., 2019. An overview of overfitting and its solutions. J. Phys. Conf. Ser. 1168 (2), 022022 https://doi.org/10.1088/1742-6596/1168/2/022022.

Yuan, C., Li, Y., Huang, H., Wang, S., Sun, Z., Wang, H., 2022. Application of explainable machine learning for real-time safety analysis toward a connected vehicle environment. Accid. Anal. Prev. 171, 106681 https://doi.org/10.1016/j.aap.2022.106681.

Zafri, N.M., Khan, A., Jamal, S., Alam, B.M., 2021. Impact of COVID-19 pandemic on motorcycle purchase in Dhaka, Bangladesh [original research]. Front. Future Transp. 2 https://www.frontiersin.org/articles/10.3389/ffutr.2021.646664.

Raihan, M. A., Anik, B. M., Ashik, F. R., Hasan, M. M., & Mahmud, S. M. (2023). *Motorcycle Helmet Use Behavior: What Does the Data Tell Us?* (No. TRBAM-23-03573).