

Integrated impact of urban mixed land use on TOD ridership: A multi-radius comparative analysis

Xinyue Gu

The Hong Kong Polytechnic University
xinyue.gu@connect.polyu.hk

Siyan Lin

The Hong Kong Polytechnic University
siyan.lin@connect.polyu.hk

Chengfang Wang (corresponding author)

South China University of Technology; State Key
Laboratory of Subtropical Building Science
cfwang@scut.edu.cn

Abstract: The global trend toward urbanization has spurred the widespread adoption of transit-oriented development (TOD). While previous research has extensively explored the relationship between land use and TOD ridership, much of it has focused on linear associations at a singular scale. Leveraging recent advancements in nonlinear modeling and the accessibility of open-source data, this study employs a comprehensive two-step methodology. Firstly, K-means clustering algorithm categorizes TOD sites in Shenzhen into three distinct clusters, providing a site-based understanding of their characteristics. Subsequently, a Light Gradient Boosting Machine (LightGBM) classification model, complemented by SHapley Additive exPlanations (SHAP) values for interpretation, quantitatively evaluates the influence of mixed land use on TOD ridership across various catchment areas. As for the findings, we discover that land-use factors have different effects on TOD site patronage at different buffer radii and delve into the intricacies of these effects. Further results reveal non-linear relationships with varying degrees of positivity and negativity. For instance, residents and health sites positively impact patronage across all buffer radii, while certain commercial land uses exhibit a negative influence. The study demonstrates how the importance of different land-use structures varies across these clusters, shedding light on the nuanced impacts of land use on TOD catchment areas. Our research optimizes land-use mixes based on predominant cluster characteristics by offering actionable recommendations for urban managers.

Keywords: Mixed land use, transit-oriented development, nonlinear decoupling model, modifiable areal unit problem

Article history:

Received: November 25, 2023

Received in revised form:
January 28, 2024

Accepted: April 24, 2024

Available online: May 28, 2024

1 Introduction

Transit-oriented development (TOD) is becoming more popular worldwide in urban planning (Hasibuan et al., 2014). For instance, 55 cities in mainland China have opened urban rail projects with a total operational mileage of 10,291.95 km by 2022, comprising 8,012.85 km or 77.85% of subways, most of which are TOD-oriented, large in number,

Copyright 2024 Xinyue Gu, Siyan Lin & Chengfang Wang

<https://doi.org/10.5198/jtlu.2024.2462>

ISSN: 1938-7849 | Licensed under the [Creative Commons Attribution – Noncommercial License 4.0](https://creativecommons.org/licenses/by-nc/4.0/)

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

and costly. Research shows that TOD has effectively responded to numerous countries' social and environmental issues stemming from car-centric transportation planning. It addresses challenges such as the unsustainable expansion of roads, traffic congestion, and air pollution. Besides, TOD also promotes the development of urban areas that are well-connected and accessible through efficient public transportation systems, fostering sustainable mobility and livable communities (Ao et al., 2019; Gan et al., 2020; Shao et al., 2020).

One of the most important issues in constructing the TOD project is considering the planning of surrounding mixed land use. This requires careful consideration of various factors, including land availability, transit accessibility, urban design principles, and stakeholder engagement, to ensure the successful and sustainable implementation of TOD initiatives (Cao et al., 2020; Taki et al., 2017). A few studies and planning projects strongly emphasize "mixed-use" around TOD stations, considering how best to match the mix of urban land use to maximize TOD's ridership and efficiency (Marilee A, 2016; Niu et al., 2019). Hence, a better understanding of TODs can be achieved by characterizing them through mixed land use.

Another important issue in the TOD project is determining the proper service radius, also known as a modifiable areal unit problem (MAUP) in relevant fields. At the meso-micro level, previous studies found that TOD projects typically conduct catchment areas with a radius of 500-1000m (5-10 minutes walking distance) for land use intensification. While these core areas are important, the land use effects of TOD areas at the meso-macro scale are equally important. When Cervero and Kockelman (1997) proposed the concept of TOD in the 1990s, he constructed a structure of coordinated development of two circles, the direct impact area of TOD and the peripheral neighboring sub-areas, which could have up to a radius of 1600m. In Korea, the planning of TOD further considers a radius of 2,000m to achieve better control of the land surrounding the TOD (Lee et al., 2005; Sung & Oh, 2011). Also, in Taiwan, Yen et al. (2023) applied the 2km as the radius of the catchment area to explore the land value around the TODs. These studies considered that a continuous, safe, and comfortable form of pedestrianized space could attract more residents to use metro stations (Wang et al., 2022). All these indicate that intensity control and functional organization should not be targeted only at the interior of the TOD but should be integrated and coordinated with the larger surroundings to enhance the stations' efficiency.

While several studies have explored the relationship between transit-oriented development (TOD) and ridership (Su et al., 2022; Yang et al., 2022), there are still some limitations within the existing literature. Firstly, the current nonlinear studies predominantly utilize early machine learning models, which may need to be better suited for multiclassification, imbalanced datasets, and lack of interpretability. Secondly, the discussions on the MAUP mainly focus on small-scale catchment areas. At the same time, the increasing popularity of shared bicycles or electric vehicles for TOD necessitates a broader examination. Lastly, despite similarities in the built environment, TODs constructed in downtown and suburban areas can exhibit substantial variations in ridership. Consequently, there is a critical need to categorize TODs, but only some studies have focused on this.

In reality, planners and government agencies have faced challenges in identifying suitable operational zones and developing effective mixed-use site plans to manage, facilitate, or regulate TOD projects. The complexities lie in determining the optimal locations for TOD implementation and designing comprehensive plans that accommodate mixed land uses, transportation infrastructure, and community needs (Hasibuan et al., 2014; Lin & Gau, 2006; Taki et al., 2017).

To address the above questions, this study takes Shenzhen in China as a testbed. In Shenzhen, one of the most developed and compact cities, the TOD planning strategy has also emerged as a new, widely accepted idea for urban construction and urban rehabilitation (Niu et al., 2019; Ramlan et al., 2021; Renne & Appleyard, 2019). Therefore, a study in Shenzhen can help us understand the relationship between urban land and TOD construction to evaluate sustainable transportation planning for developed cities accurately.

Regarding the methods, this study uses interpretable machine learning to decipher the impact of various land uses on TOD ridership in different catchment areas. It demonstrates the following contributions: (1) Revealing the specific effect of mixed land use on TOD ridership using the state-of-the-art machine learning models, namely Light Gradient Boosting Machine (LightGBM) and SHAP (SHapley Additive exPlanations) values. (2) Comparing the non-linear and heterogeneous effects of land uses across three catchment areas with radii of 500, 1000, and 2000 meters. (3) Discussing the effects of land uses based on different categories of TODs. The remainder of the paper is organized as follows. The paper is structured as follows: Section 2 reviews the relevant literature. Section 3 introduces the study area, data sources, and the modeling approach. Section 4 presents the visualizations and interpretations of the model results. Finally, the paper concludes with insights for planning practice and future directions for further research.

2 Literature review

2.1 The relationship between mixed land use and TOD catchment areas

The relationship between the built environment and ridership in the catchment areas, a key aspect of TODs, plays a critical role in reinforcing the scientific basis of urban policies and the efficiency of planning solutions (Shao et al., 2020). With the TOD catchment areas as study units in earlier research, direct ridership models, a pioneering and widely adopted technique for transport planning, involve regressing ridership in TOD's environment. Through this process, it has been identified that demographic characteristics and land use factors significantly influence TOD ridership, thus playing a crucial role in predicting travel demand (Cervero, 2006). To gain deeper insights into this phenomenon, researchers in urban-rural planning and transport geography have investigated the land use characteristics of catchment areas and their correlation with ridership patterns.

Furthermore, researchers have explored these relationships across various areas to address the Modifiable Areal Unit Problem, ensuring a more comparative analysis (Ma et al., 2018; Su et al., 2022). The Modifiable Areal Unit Problem, commonly called MAUP, is a challenge frequently encountered in geographic analysis. It arises when the choice of spatial units (such as neighborhoods or districts) influences the statistical analysis results, potentially introducing biases. With the growing availability of geospatial big data, encompassing points of interest (POI), metro swipe data, social media data, and mobile phone signaling data, researchers can integrate multiple data sources. This integration facilitates a more comprehensive characterization of the intricate details surrounding Transit-Oriented Developments (TODs). By harnessing these diverse data sources, researchers can better understand the complex dynamics and interactions within TOD environments (Iseki et al., 2018; Li, Lyu et al., 2020).

Recently, studies on the relationship between mixed land use and TOD catchment areas have increasingly emerged all over the world, including the United States (Ding et al., 2019), the UK and Europe (Ingvardson & Nielsen, 2018), South Korea (Choi et al., 2012; Sohn & Shim, 2010), and China (An et al., 2019; Huang et al., 2020). Extensive

research has consistently shown that built environment and land use influence TOD ridership along three principal dimensions: density, diversity, and design (Cervero & Kockelman, 1997; Harirchian et al., 2021; Lin & Gau, 2006). Density refers to the number of people or jobs within a given area. Higher-density areas are more likely to support transit use because they provide a larger pool of potential riders within a walkable distance of transit stops (Cervero, 1989). Design refers to the physical layout and organization of buildings, streets, and public spaces. Good design can support transit use by making it easy and pleasant to walk or bike to transit (Cervero, 1989; Nyunt & Wongchavalidkul, 2020). Diversity refers to the mix of different land uses in an area, such as residential, commercial, and institutional uses. A diverse mix of land uses can support transit use by providing a variety of destinations within a walkable distance of transit stops (Nyunt & Wongchavalidkul, 2020). Furthermore, a diverse mix of land uses can also support a more balanced demand for transit throughout the day, as people travel for different purposes at different times. These findings of diversity research emphasize the importance of achieving an optimal land use mix. For instance, a 1-hectare increase in Hong Kong in commercial/residential floor areas is associated with a 100 average weekday rail ridership rise. Similarly, a change in business floor space leads to an increase of 20 average weekday rail ridership per hectare in New York (Loo et al., 2010; Shao et al., 2020). That is to say, the land use of the surrounding sites plays a crucial role in sustaining rail ridership, particularly in high-density and developed cities, as it reflects the spatial function and intensity of human activity. Given its status as a mass transit vehicle, maintaining an appropriate and supportive land use environment is essential for ensuring the continued success of rail systems planning (Li et al., 2024; Loo et al., 2010).

Despite numerous studies focusing on the relationship between the built environment and TOD ridership (Li, Lyu et al., 2020), a common limitation is the narrow scope of analysis within a specific radius catchment, typically around 500 or 800 meters (approximately a five to ten-minute walk) (Su et al., 2022; Yang et al., 2022). However, this perspective overlooks the potential effects at larger scales. After all, commuters arriving at the TOD have more ways to travel than just walking at a constant speed. One such mode is the increasingly popular bike-sharing system, which allows users to cover distances of up to approximately two kilometers in just five minutes, significantly expanding the range of activities for TOD commuters. This highlights the importance of considering not only pedestrian accessibility but also the availability and integration of other transportation options when assessing the effectiveness and reach of TOD developments.

It is critical to direct the building of sites throughout numerous surrounding circles. Instead of concentrating just on the local area, a more global and comprehensive understanding of the influence of TOD on neighboring sites should be reached. As a result, there is a need for research that explores the impact of the built environment on TOD ridership across broader geographic areas, encompassing larger scales. Different catchment areas of multiple radii in 500m, 1000m, and 2000m are constructed in this study to reveal the impact of land use on TOD ridership.

2.2 The relationship shifts from linear to nonlinear

Before machine learning was widely available, the majority of existing empirical studies adopted simple regression models such as least squares, structural equation modeling, and geographically weighted regression to fit the direct or indirect effects of land variables on TOD ridership (Estupiñán & Rodríguez, 2008; Sohn & Shim, 2010). Many previous studies explored the relationship between population density, residential/employment areas, land use, and TOD ridership. However, these studies only

assume a linear or log-linear correlation between these factors and TOD ridership, which may oversimplify the complex dynamics (Shao et al., 2020).

More advanced modeling techniques are needed to capture and accurately represent the multifaceted factors influencing TOD ridership. Researchers can uncover and effectively model the intricate relationships and underlying factors that impact TOD ridership by employing sophisticated modeling approaches. For instance, research in Washington, USA, unveiled a positive correlation between job density and metro riding up to a specific threshold, while no further relationship once the threshold was exceeded. The marginal effect of land-use mixing is zero when the mixed-use index is too little or too large (Ding et al., 2019). These studies have revealed that certain variables may only significantly correlate with ridership once a specific threshold is surpassed. Furthermore, the interpretation of this non-linearity can vary depending on the specific independent variable being examined (Tao et al., 2020; Yang et al., 2021).

Even more and more research on the impact of land use on TOD ridership shifting to more sophisticated non-linear models (Cheng et al., 2020; Wu et al., 2022), findings could not avoid the diverse effects of mixed land use across the various catchment areas, causing the MAUP. The effect of a specifically built environment variable on ridership may saturate or reduce up to a certain scale, and figuring out its effective range might offer more precise guidelines for land use planning for TOD (Ding et al., 2019; van Wee & Handy, 2016).

Hence, when planners utilize land use policies to encourage transit usage, it is imperative to understand how different scopes of citizen activities contribute to TODs. The scopes of activities are not only limited to the different radiuses of catchment areas but also indicate different urban locations at the TOD station. Sites in well-developed regions and suburban areas will have different activity effects (Song et al., 2022). Therefore, in addition to delineating the radius of different catchment areas, it is equally necessary to delineate the categories of TODs.

However, there needs to be more research examining the heterogeneous effects of mixed land use on different types of TODs and their corresponding catchment areas. Hence, our study first clustered the TOD stations using the K-means algorithm to address the above challenges. Subsequently, a more effective and interpretable non-linear model, LightGBM classification with SHAP values, is employed to construct accurate models capable of handling multiple classifications. The results of this study will add a more comprehensive and rigorous perspective to related studies.

3 Material and methods

3.1 Research area

Shenzhen, positioned as one of the four central cities of the Guangdong-Hong Kong-Macao Greater Bay Area, is the first and most significant special economic zone in China (Gu, Tang et al., 2024; Shao et al., 2020). It also secured the third position in the Walden Economics Institute's 2021 ranking of the top 100 cities in China.

Functioning as a national logistics hub and an international integrated transport hub, Shenzhen spans an area of 1,997.47 km², encompassing ten districts. It benefits from a well-developed public transportation system (Yang et al., 2021). Notably, on December 28, 2004, Shenzhen became the fifth city in mainland China to inaugurate a metro system by opening its first metro line. With a total length of 265 km, the Shenzhen metro system ranked fifth in China in terms of overall mileage in 2018. The network of metro lines accommodates a daily average passenger flow exceeding 5.1 million individuals,

contributing to a public transportation share of 48%. Consequently, the metro system has become a fundamental urban public transportation network component.

Furthermore, the Shenzhen metro system incorporates substantial TOD principles, employing a mixed and intensive approach to station planning that fosters connectivity with surrounding neighborhoods and facilitates the city's rapid and mature metro system expansion (Zhou & Yang, 2021). Thus, this study focuses on the Shenzhen Metro as of 2018 and offers crucial insights for TOD construction, particularly in developing countries with medium-density urban environments.

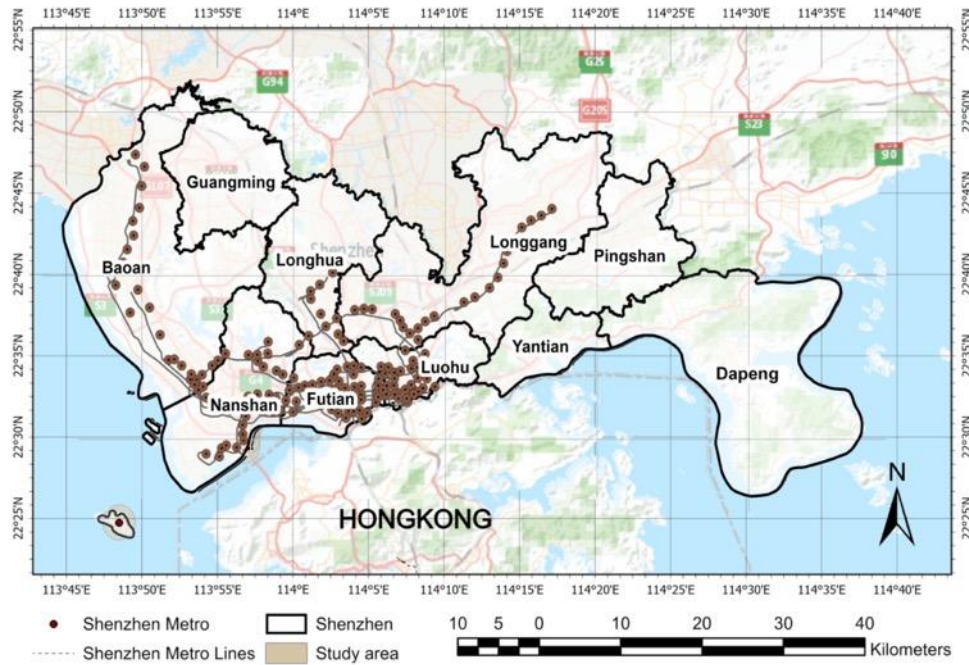


Figure 1. The location of Shenzhen and TOD metro stations

Note: The circle of each metro station shown in this figure is a 500m radius. The actual land use is based on the map in GIS Pro.

Figure 1 shows the geographical locations of the metro stations in Shenzhen. The study measures the average distances between two adjacent stations in the whole districts, the center districts (Futian, Luohu, and Nanshan), and the suburban districts (Baoan, Longgang, and Longhua) using GIS Pro, which are around 982m, 806m, and 1306m respectively. In addition to this, the minimum distance between the two sites is 354m, and the maximum distance is 2984m. The distance between adjacent stations also aided in selecting the radius for this study.

3.2 Data collection and processing

In this study, we primarily focus on analyzing various land use proportions while incorporating additional control variables, namely floor area ratio (FAR), mix land use index (MLU), and points of interest (POI). These variables are crucial in investigating the research objectives and understanding the relationship between land use patterns and other relevant factors.

The land use data in this study is derived from the China Basic Urban Land Use Type Map, which was researched and published by the Tsinghua University team (Gong et al., 2020). This dataset provides a detailed map based on a unified urban land use standard (Table 1). It integrates various data sources, including 10m satellite images, OpenStreetMap, nighttime lights, and Tencent social big data from 2018. Machine learning techniques were employed to train the data outcomes using the input attributes (Gong et al., 2020). The dataset covers 6,808 plots (excluding roads) in Shenzhen. According to the dataset, in 2018, the impervious area in Shenzhen was 911.969 km². The land use proportions were as follows: residential land accounted for 32.5% (296.25 km²), commercial land accounted for 4.9% (44.86 km²), industrial land accounted for 31.0% (282.86 km²), transportation land accounted for 1.2% (10.91 km²), and land for public administration and services accounted for 1.8% (16.86 km²).

Table 1. Classification of urban land-use data

Land use	Descriptions
Resident	Houses and apartment buildings are places where people live.
Business	Buildings where people work, including office buildings, internet technology, e-commerce, media, etc.
Commerce	Houses and buildings for commercial retail, restaurants, lodging, and entertainment.
Industry	Land and buildings used for manufacturing, warehouse, mining, etc.
Government	Lands used for government, military, and public service agencies.
Education	Lands, including schools, universities, institutes, and ancillary facilities, are used for education and research.
Health	Lands are used for hospitals, disease prevention, and emergency services.
Culture	Lands are used for public sports and cultural services, including gyms, libraries, museums, and exhibition centers.
Green	Parks and green spaces are used for entertainment and environmental conservation.

While the overall accuracy of the original dataset ranged from 61% to 80% (Gong et al., 2020), a random calibration process was conducted to verify the accuracy of the land use data in different regions of Shenzhen by comparing it with real satellite remote sensing images from the same year with the base map in GIS Pro. The results demonstrated that the overall data accuracy was sufficiently high, with parcel boundaries aligning with actual road planning and land classifications consistent with real land use functions. These findings confirm the suitability of the dataset for conducting high-precision research in this study. Figure 2 also shows some sample mappings to illustrate the land use around the TOD stations in different locations of Shenzhen.

Then, to more accurately capture the intensity of functional activities associated with specific land use types in three dimensions, we obtained the building roof vector dataset of Shenzhen City (Zhang et al., 2022). This dataset includes building footprints (*BF*) and building layers (*BL*) and was obtained from the National Qinghai-Tibet Plateau Scientific Data Center in China. By incorporating this dataset, we aim to enhance the characterization of functional activities within different land use types, focusing on their vertical dimension.

$$SH = \sum_{k=1}^n (BF_n \times BL_n) \quad (2-1)$$

where, SH denotes the total area of high-rise buildings in each catchment radius, n represents the number of each building, BF indicates the building footprints, and BL signifies the building layers.

Based on the above land use and architecture data, FAR and MLU are calculated in this study. FAR (Floor Area Ratio) is the ratio of the total floor area of buildings to the land area, reflecting the intensity of development and building density on a given land. MLU measures the diversity of land use structures and functions within a certain area. It can be obtained by calculating the proportion of different land use types and the information entropy (Lin et al., 2022). These indicators provide important information about land use patterns and diversity, helping researchers understand the characteristics and effects of land use. The calculation formula for FAR and MLU are as follows:

$$FAR_j = SH_j / BF_j \quad (2-2)$$

where FAR_j represents the floor area ratio of the j catchment area, SH_j represents the total floor area, and BF_j is the total floor print area in the catchment area.

$$MLU_j = 1 - \sum_{i=1}^n (P_i \times \log(P_i)) \quad (2-3)$$

where MLU_j represents the mix land use index of the j catchment area, P_i represents the proportion of the i land use type.

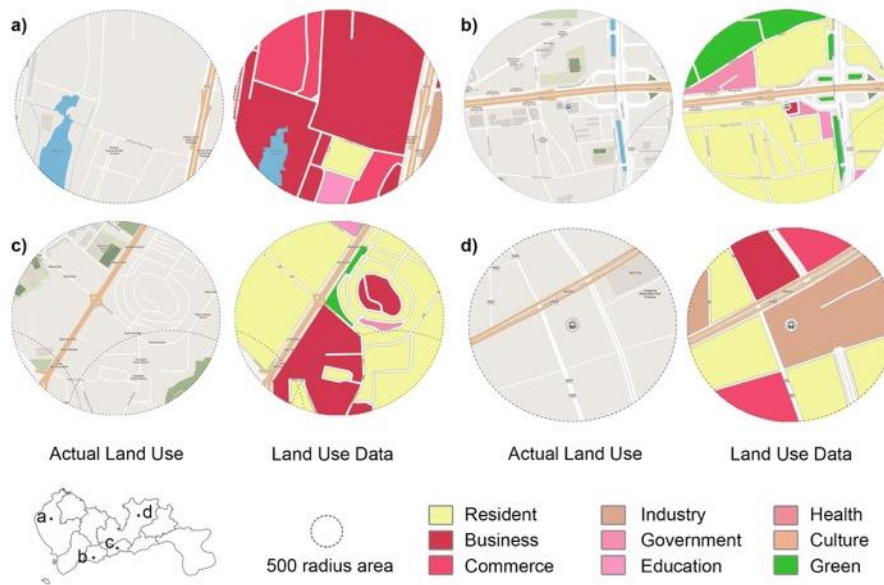


Figure 2. Sample mappings of the land use around the TOD stations

Note: The actual land use is based on the OSM base map in GIS Pro.

Lastly, POI data reflects multiple functions' overlay and mixing degrees on a specific land use category. It provides a more accurate reflection of the completeness of facilities on a particular land and is an important factor in attracting human flow. Therefore, we include it as a control variable in this study. The POI data used in this research is sourced from the OpenStreetMap website and represents the total number of POIs in Shenzhen in 2018. It includes restaurants, leisure and entertainment venues, and cultural and sports facilities. The complete list of POI categories can be queried on the OpenStreetMap website.

Although POI classifications are consistent with land use classifications, POIs are the basic units of place function and focus on reflecting the density and diversity of socio-economic activities. In comparison, land use data can be used to characterize the planned area related to building sites and plot ratios. Based on these differences, land use is the independent variable, and POI, FAR, and MLU are used as a control variable to analyze the effect of mixed land use on the TOD ridership.

3.3 Metros ridership and catchment area definition

With the advancement of urban metro transportation and the availability of various transportation big data, such as cab GPS data, shared bicycle data, metro card data, and bus data, these datasets have become essential for planning and research purposes. This study's metro ridership data is derived from the Shenzhen Pass card data, publicly available on the Shenzhen government data open platform (<https://opendata.sz.gov.cn/>). The specific data used in this study covers the dates of September 1, 2018 (Saturday) and consists of a total of 875,296 metro ride records, with both passenger on- and off-boarding records recorded. It reflects the overall ridership numbers at the metro stations, which gives a more holistic picture of the heat of activity on the metro and the differences in all-day traffic flow between metros. It is worth noting that on those specific dates, the metro system carried most of the city's public transportation ridership. The data includes passenger IDs, ride dates, transaction values, and on- and off-boarding status. By conducting data cleaning and statistical processing of the on- and off-boarding stations, the ridership for each line station can be obtained.

To assess the completeness and accuracy of the land use data, a comparison was made using the information provided by the data for the 165 Shenzhen metro stations, which serve as the focus of this study (Figure 1). The ArcGIS Pro platform created catchment areas with radii of 500m, 1000m, and 2000m around each metro station. The radius was chosen from previous studies on land use and TOD (Shao et al., 2020; Wang et al., 2022). Subsequently, the land use and building attributes were extracted and quantified within each catchment area. This process allows for a detailed analysis of the land use characteristics and patterns near the metro stations.

3.4 Clustering analysis: The K-Means approach

By cleaning and counting the metro data, we have a significant variation in ridership between different stations, and it is not ideal to directly make a regression of the relationship between ridership and land use. Therefore, it is necessary to classify the sites into different groups, which means that the stations with more similar ridership are in the same group, so the classification regression model can be used to explore the mixed land use effects on the ridership.

Cluster analysis enables the classification of sample sets into classes in an unsupervised mode and discovers the characteristics of different classes. K-Means, one of the most commonly used methods in clustering, calculates the best class attribution based on the similarity of the distance between points, and the division principle is the minimization of samples within a group (Krishna & Narasimha Murty, 1999; Sinaga & Yang, 2020), as shown in the following equations:

$$C^{(i)} = \arg \min_j \|x^{(i)} - u_j\|^2 \quad (2-4)$$

$$u_j = \frac{\sum_{i=1}^m 1\{C^{(i)} = j\}x^{(i)}}{\sum_{i=1}^m 1\{C^{(i)} = j\}} \quad (2-5)$$

where, $C^{(i)}$ represents the closest one among a sample i and k classes, and the value of $C^{(i)}$ is one of 1 to k . The center of mass u_j represents our guess of the centroids of the samples belonging to the same class, and the clustering is considered converged if we repeat the iterations until the center of mass remains the same or changes very little.

Besides, one crucial parameter that needs to be specified in K-Means is the number of clusters. The elbow method is utilized to determine the optimal number of clusters. This method involves plotting the within-cluster sum of squares (WCSS) against the number of clusters and identifying the point at which the change in WCSS starts to level off significantly (Sinaga & Yang, 2020). The number of clusters corresponding to this point is chosen as the classification criterion, ensuring a meaningful and informative classification of the metro stations.

3.5 Regression and decoupling: LightGBM classification and SHAP value

LightGBM classification, an integrated machine learning algorithm based on decision trees and gradient boosting, has demonstrated superior performance compared to artificial neural networks and traditional linear statistical models in terms of accuracy and generalization, especially when dealing with unstructured data in small and medium-sized structured datasets (Ke et al., 2017). The underlying framework of LightGBM classification has been enhanced through system optimization and algorithm improvements, surpassing other popular machine learning regression models such as gradient boosting, random forest, and logistic regression in both execution speed and model performance. As a result, it has gained considerable attention in various data science domains (Memon et al., 2019).

The development of eXplainable AI (XAI) has made machine learning and artificial intelligence models more interpretable, overcoming their black-box nature. One such XAI technique is Shapley Additive explanations (SHAP), which can model, interpret, and visualize complex processes (Gu, Wu et al., 2024). The Shapley value, a concept from game theory (Kuhn & Tucker, 2016), is a fundamental method for model interpretation. SHAP builds upon this concept and quantifies the contribution value of each feature in the model, considering the dimensions of individual observations (Štrumbelj & Kononenko, 2014). Therefore, this study applied the SHAP value to explain the LightGBM models. The equations are as follows:

$$SHAP(X_j) = \sum_{S \subseteq N \setminus \{j\}} \frac{k!(p-k-1)!}{p!} (f(S \cup \{j\}) - f(S)) \quad (2-6)$$

$$\hat{y}_i = y_{base} + \sum_{j=1}^k SHAP(x_{ji}) \quad (2-7)$$

$$SHAP_{global_j} = \frac{\sum_{i=1}^n |SHAP(x_{ji})|}{k} \quad (2-8)$$

In equations, $SHAP(X_j)$ denotes the SHAP value of the feature j , p represents the total number of features, $N \setminus \{j\}$ indicates a set of all possible combinations of features, excluding X_j , S is a feature set in $N \setminus \{j\}$, $f(S)$ signifies the model prediction with features in S , and $f(S \cup \{j\})$ stands for the model prediction with features in S plus feature X_j . Meanwhile, \hat{y}_i represents the model prediction value for the observation i , y_{base} denotes the mean value of the predictive value at other samples and $SHAP(x_{ji})$ indicates the SHAP value of the j^{th} feature for i , reflecting the marginal contribution of the feature to the prediction.

4 Analysis results

4.1 Clustering results for TOD metros

We employed the elbow method to identify the optimal number of clusters to understand better how land use influences TOD passenger flow in various geographical locations. Subsequently, we applied the k-means clustering algorithm to analyze the passenger flow data for each subway station. This approach allows us to examine the specific impacts of land use in different geographical contexts on TOD ridership.

To study the impact of land use in different geographical locations on TOD passenger flow, more specifically, after scaling the data and determining the optimal number of clusters based on the elbow plot (see Figure 3), we ran the k-means clustering algorithm with the passenger flow of each subway station.

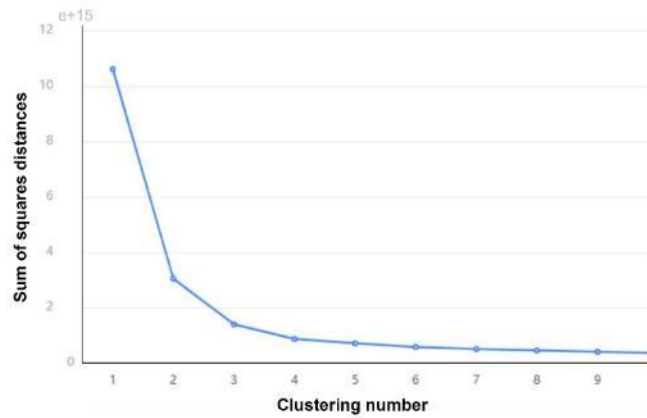


Figure 3. Elbow plot result, which was used to determine the best clustering group number

As shown in Table 2, the algorithm clustered the TODs into three distinctive clusters, designated T1 to T3. The number and the percentage share of the TOD metro station in the three clusters are described in the table. Cluster T1 has the maximum share of the total station, 52.7%, while cluster T3 has the least share, 6%.

Table 2. Classification of urban land-use data

Cluster	T1	T2	T3
Number	87	68	10
Share of TOD metro station	52.7%	41.2%	6%

Figure 4 shows the spatial distribution of three types of metro stations after clustering. We found that cluster T1 is located in the center and south of the city, which are the places where the economy is more developed. Clusters T2 and T3 are spread on the edge of the city center and north of the city. Combined with Shenzhen’s specific urban status quo, cluster 1 is the more densely populated and developed area, followed by cluster 2, and cluster 3 is the more sparsely populated and undeveloped area.

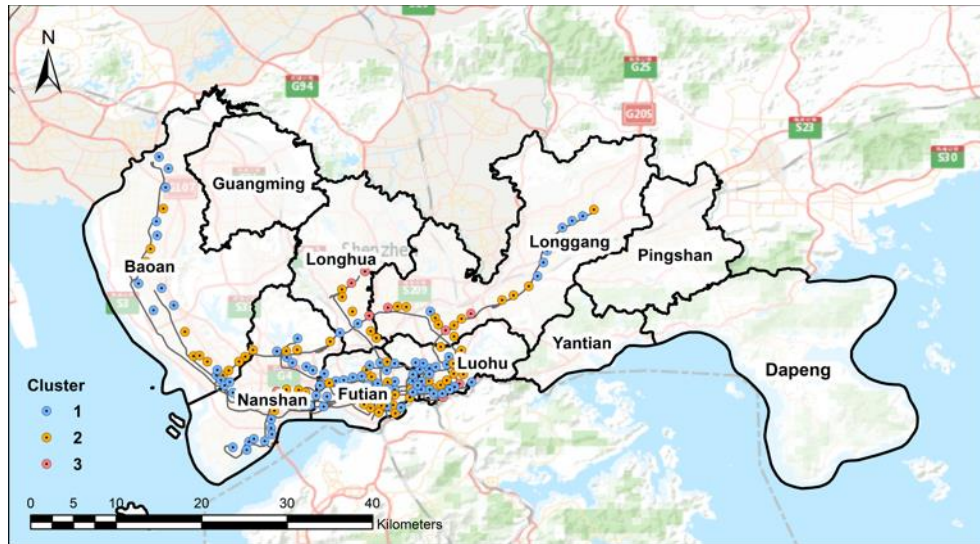


Figure 4. Spatial distribution of three types of TOD metro stations in Shenzhen

Note: The actual land use is based on the map in GIS Pro.

4.2 Multicollinearity test for independent variables

In this study, we used the Variance Inflation Factor (VIF) to assess the correlation and multicollinearity between independent variables. Typically, a VIF value greater than five or a high value (e.g., greater than 10) may indicate the presence of severe multicollinearity, signifying stronger collinearity among the independent variables. The VIF evaluation results for the three catchment areas are shown in Table 3, indicating that the 12 independent variables used in our analysis do not exhibit multicollinearity issues. This supports the subsequent regression analysis.

Table 3. The VIF result of the multicollinearity test for twelve independent variables

Variables	R=500	R=1000	R=2000
<u>MLU</u>	<u>1.604</u>	<u>1.854</u>	<u>2.025</u>
Resident	1.623	1.675	2.106
Business	1.729	1.61	2.542
Commerce	1.300	1.232	1.524
Industry	1.198	1.259	1.598
Government	1.093	1.052	1.416
Education	1.130	1.103	1.390
Health	1.074	1.096	1.609
Culture	1.182	1.228	1.329
Green	1.166	1.098	1.210
FAR	1.603	2.077	1.730
POI	2.003	2.044	2.447

4.3 Machine learning parameterization and fitting performance

This study split the data into a training set (80%) and a test set (20%). The hyperparameters of the LightGBM classification models were determined using a systematic grid search method available in the Scikit-learn library of Python, based on the model's performance evaluated on the validation set.

Furthermore, we used Accuracy, MAE (Mean Absolute Error), and RMSE (Root Mean Squared Error) to assess the predictive performance of the machine learning model. Accuracy is suitable for classification tasks, measuring the proportion of correctly classified samples in the predictions, as shown in Table 4. MSE calculates the average of the squared differences between the model's predicted values and the true values. MAE computes the average of the absolute differences between the model's predicted and true values, and it is more robust and less sensitive to outliers than MSE. RMSE calculates the square root of the mean squared error, providing a better measure of the differences between the predicted and true values while evaluating the magnitude of prediction errors. A higher accuracy and smaller values for MSE, MAE, and RMSE indicate better performance of the model.

Table 4. The metrics of the three LightGBM classification models in different radii

Radius	Accuracy		MAE		RMSE	
	Training	Testing	Training	Testing	Training	Testing
500	0.884	0.736	0.015	0.394	0.123	0.674
1000	0.885	0.615	0.015	0.515	0.123	0.759
2000	0.892	0.645	0.008	0.455	0.087	0.674

The detailed results are shown in Table 4. Due to the relatively small sample size and the relatively simple nature of the classification task in this study, considering various indicators comprehensively, there are no apparent signs of overfitting or underfitting. As a result, the fitting performance of the three models can be regarded as satisfactory.

4.4 Global relative importance of land Uses in TOD Metros

Figure 5 and Table 5 show the relative importance of the explanatory variables representing TOD metro ridership of LighGBM classification models when the catchment radius is 500, 1000, and 2000, respectively.

Table 5. Comparing the global relative importance value in the model of different variables facilities on TOD ridership

Variables	R=500	R=1000	R=2000
MLU	165	50	115
Resident	237	156	78
Business	65	73	89
Commerce	72	103	133
Industry	16	62	97
Government	46	133	102
Education	74	119	131
Health	24	96	124
Culture	106	81	102
Green	79	137	127
FAR	244	134	162
POI	183	143	104

A different trend was found in each catchment radius zone. In the R=500 catchment radius (Figure 5-a), the functional types contributing more to TOD ridership consist of FAR, resident, MLU, POI, and culture. In contrast, the factors contributing less to ridership include industry, health, and business. In the R=1000 catchment radius zone (Figure 5-b), the most important factors included the POI, residents, health, education,

and commerce, and the least important factors were culture and industry. Meanwhile, in the R=2000 catchment radius zone (Figure 5-c), TOD ridership is strongly influenced by health, MLU, commerce, business, and FAR. However, education and industry have less impact. This highlighted the essential role of different catchment radius zones in unraveling the TOD-metro ridership relationship and suggested that discussing the different catchment radius zone scenarios is necessary.

Comparing the relative importance of factors affecting ridership at TOD subway stations with different catchment radius zones, it becomes evident that MLU (Mixed Land Use), FAR (Floor Area Ratio), and POI (Points of Interest) are influential factors that must not be overlooked in the study of ridership at TOD stations, with POI being particularly noteworthy. Surprisingly, residents and health play crucial roles as influencing factors among the various land use factors. Specifically, residents have a greater impact on the passenger flow of TOD sites with R=500 and 1000, while health significantly influences the passenger flow of TOD sites with R=1000 and 2000.

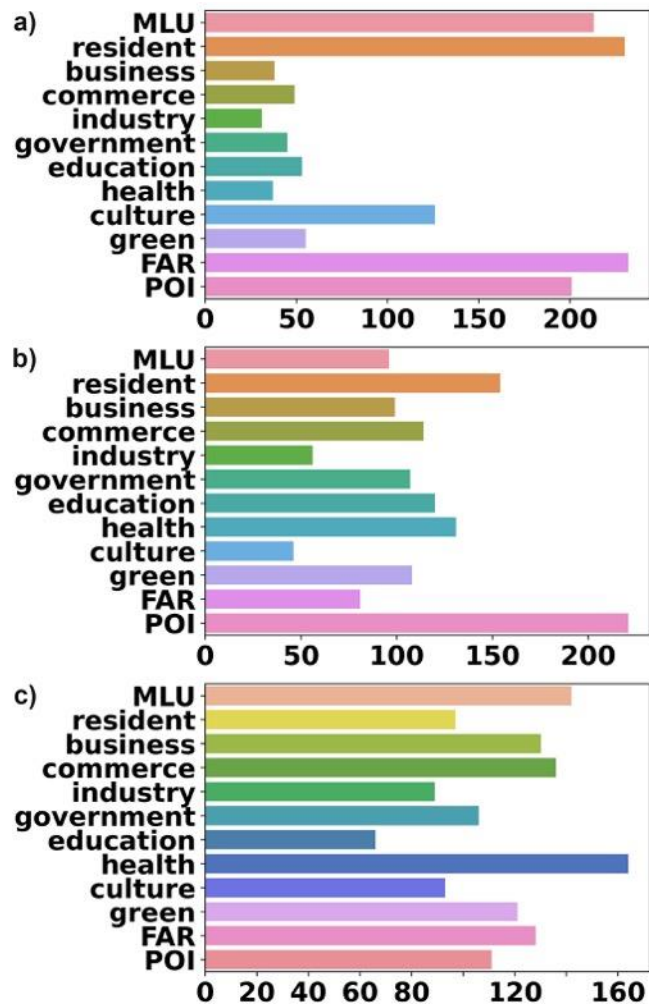


Figure 5. The global relative importance of the explanation and control variables on TOD ridership, respectively. a) R=500, b) R=1000, and c) R=2000

4.5 Decoupling the nonlinear impact of land use on various TOD metros

Although the importance of features in LightGBM can intuitively reflect the importance of features, it cannot measure how the features are related to the final prediction results. Thus, the SHAP value should be introduced to reflect the influence of the features in each sample and the positive and negative impact. The subsections analyze the SHAP values of different clusters at different radius catchment radii.

In this study, we utilized the SHAP tool implemented through the SHAP module in Python to elucidate the dissimilarities among the clusters. By employing the SHAP tool, we produced a graphical representation called a “summary plot,” which effectively showcases the pivotal features and their respective influence on the prediction. The summary plot generated by the SHAP tool is presented in Figures 6 to 8.

In the SHAP value summary plot, each feature is depicted by a horizontal bar, and the most important features will be on top. The bar’s color gradient reflects the variable’s corresponding value or magnitude. Notably, blue represents a lower value, while red means a higher value. The direction of the SHAP value indicates the effectiveness of a feature’s value in characterizing the cluster. A positive (negative) SHAP value signifies a feature more (less) likely to contribute significantly to the cluster’s characterization.

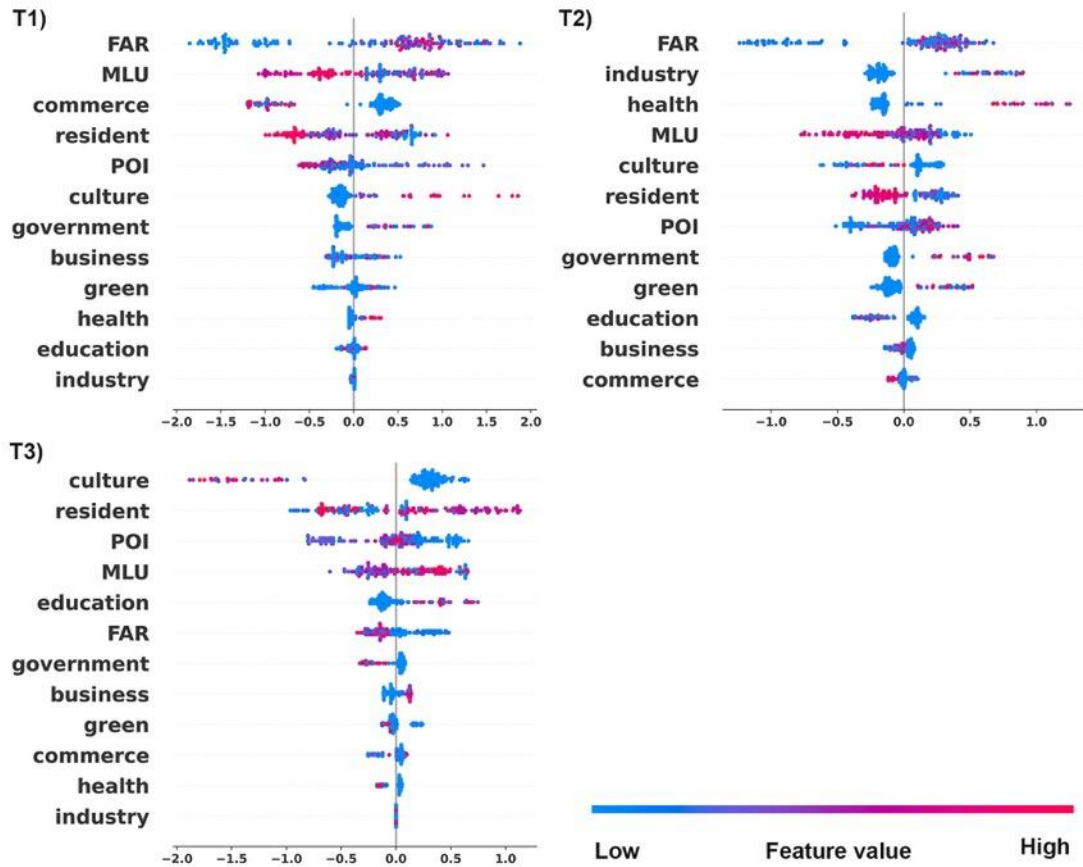


Figure 6. SHAP value for T1, T2, and T3 clusters in the R=500 catchment area

When the catchment radius is set to 500, our analysis revealed a noteworthy observation: a higher proportion of the Culture land type exhibited a positive SHAP value

for the T1 type site (Figure 6-T1). Conversely, among the other influencing factors, the proportion of Commerce and Residents displayed relatively high negative SHAP values. This implies that the passenger flow within cluster T1 is more likely to increase with an augmented proportion of Culture land. In contrast, both Commerce and Resident proportions harm the passenger flow of cluster T1.

Regarding T2 sites (Figure 6-T2), a similar pattern emerges, where the proportion of industry and health positively impacts the passenger flow within the T2 cluster. Conversely, the proportion of culture and residents negatively impacts this cluster's passenger flow. Regarding T3 sites (Figure 6-T3), an intriguing shift is observed: the proportion of culture has emerged as the most influential factor in passenger flow. Surprisingly, a high value of culture land exhibits a negative SHAP value, indicating its adverse effect on passenger flow. Conversely, the high value of resident land type demonstrates a positive SHAP value, suggesting that an increase in this land type may positively enhance the passenger flow within the T3 cluster when R=500. Furthermore, education also positively impacts passenger flow in this context.

When R=500, the control variables MLU, POI, and FAR emerge as significant factors influencing passengers. Simultaneously, the specific direction and magnitude of their influence on passenger flow will vary across different cluster types. The proportions of land uses such as culture, commerce, residential, industry, and health also play crucial roles in this variability, as their impact on passenger flow differs among the various clusters.

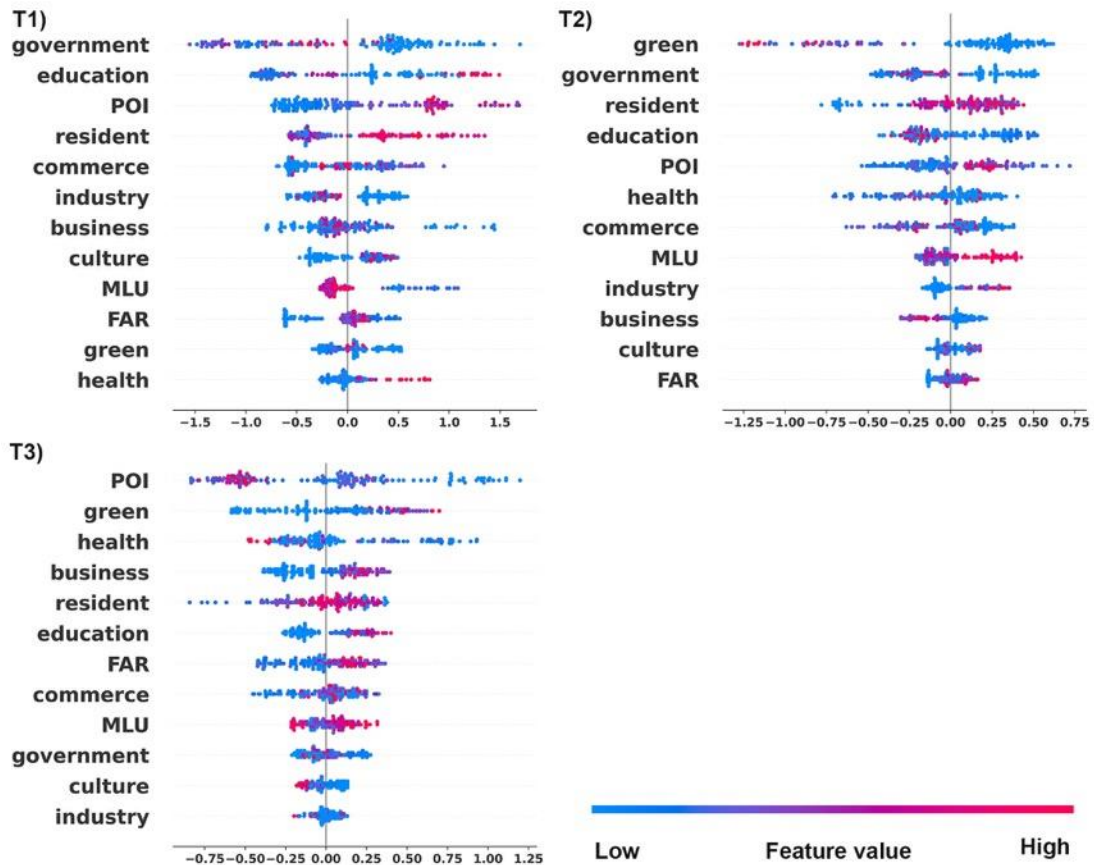


Figure 7. SHAP value for T1, T2, and T3 clusters in the R=1000 catchment area

When the catchment radius zone is set to 1000, focusing on T1 sites (Figure 7-T1), among the land use factors, the proportion of government and education exhibits the most substantial impact on the passenger flow of TOD sites. However, it is important to note that their influence directions are opposite: while government proportion negatively affects passenger flow, education proportion positively impacts passenger flow. Additionally, the proportion of residents positively influences the passenger flow in this context.

Regarding T2 sites (Figure 7-T2), the proportions of green and government have emerged as the two most significant factors. Notably, their high values are associated with negative SHAP values, indicating that the passenger flow of TOD sites is more likely to decrease with an increase in these two land types. Conversely, the high value of the Resident ratio exhibits a positive SHAP value, suggesting that its growth is more likely to enhance the passenger flow of the T2-type site.

In the case of T3 sites (Figure 7-T3), the passenger flow of TOD sites may be positively influenced by the proportions of green, business, resident, and education land types. Increasing these land types will likely increase the site's passenger flow. On the contrary, health-related land types harm the site's traffic, potentially leading to decreased passenger flow.

When $R=1000$, across the three different TOD sites, special attention should be given to the land types of government and green. Government land type has a relatively significant negative impact on the passenger flow of T1 and T2 sites. In contrast, green land type exhibits a substantial negative effect on T2 TOD sites and a noteworthy positive impact on T3 TOD sites. These findings underscore the importance of carefully considering the influences of government and green lands in the context of varying TOD sites.

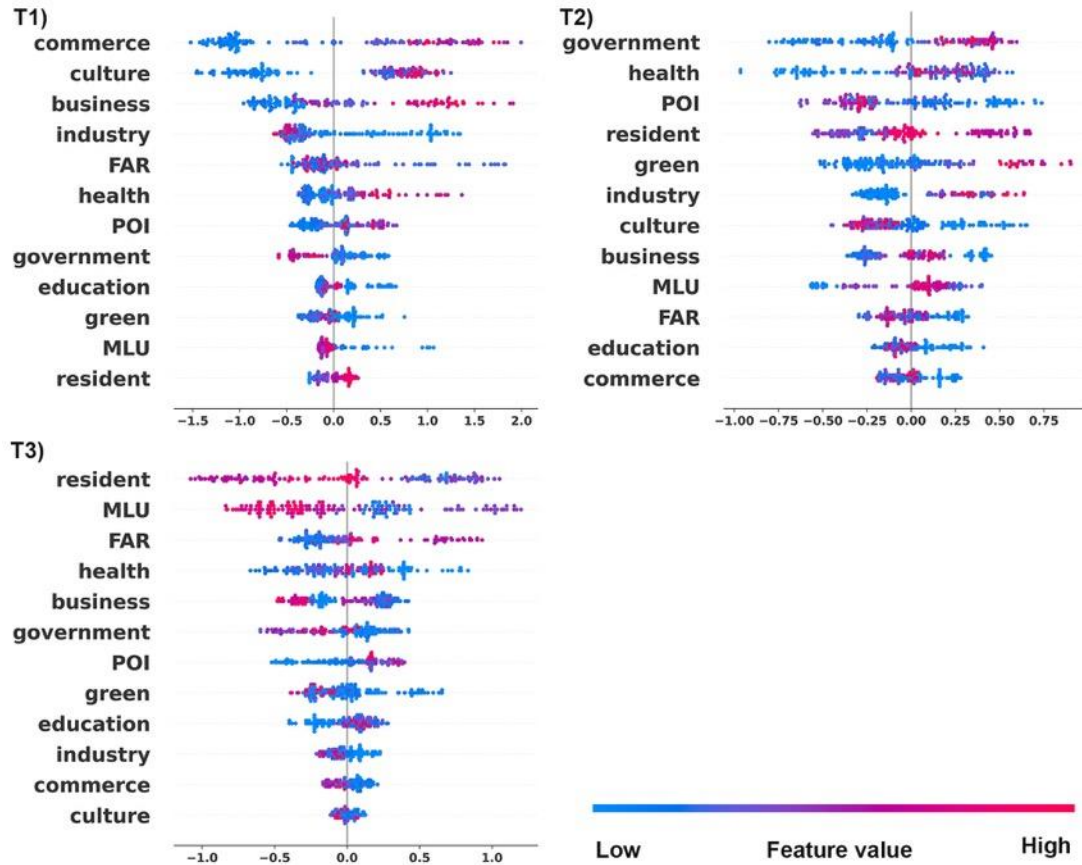


Figure 8. SHAP value for T1, T2, and T3 clusters in the R=2000 catchment area

When the catchment radius is set to 2000, focusing on T1 stations (Figure 8-T1), the proportions of commerce, culture, and business positively influence the passenger flow of TOD subway stations. Conversely, the proportion of industry hurts the passenger flow of the station. Hence, an increase in commerce, culture, and business land types is likely to enhance the passenger flow, while a higher proportion of industrial land types may lead to a decrease in passenger flow for the T1 station.

Regarding T2 sites (Figure. 8-T2), the proportions of green, government, resident, and education are the key land type factors that significantly influence their passenger flow. Among these factors, the proportions of green, government, and education hurt the passenger flow of the TOD site. In contrast, the proportion of residents has a positive effect on the passenger flow of the site. Therefore, increasing the proportions of green, government, and education land types may decrease passenger flow, whereas increasing the proportion of resident land types is likely to improve passenger flow for T2 sites.

Concerning T3 sites (Figure. 8-T3), residents, health, business, and government are the primary factors influencing passenger flow among the various land use types. Specifically, resident, business, and government land types negatively impact TOD site passenger flow, while the health land type positively affects TOD site passenger flow. Consequently, an increase in resident, business, and government land types may decrease passenger flow, whereas an increase in healthy land types will likely lead to a rise in passenger flow for T3 sites.

When $R=2000\text{m}$, the impact of land type factors on the passenger flow of different TOD stations varies significantly. Therefore, a case-by-case analysis is essential to discern the specific influential factors for each station accurately. It is crucial to assess and evaluate each TOD station individually to accurately understand the unique interplay of land type factors and their effects on passenger flow.

5 Discussions

5.1 Land management and policy planning response

Transit-Oriented Development (TOD) plays a pivotal role in urban planning and design, especially as the city experiences rapid expansion in construction land, resulting in land constraints. Achieving a well-balanced and integrated land use layout becomes a prerequisite for effective urban management and policy planning. This study delves into the impact of various land uses on ridership within three TOD catchment areas, considering real conditions and planning in the city. The results of this study are not only applicable to Shenzhen but also have great significance for other cities with similar construction backgrounds and development levels:

Firstly, the existing TOD planning needs comprehensive strategies for different catchment areas. As indicated in the results, catchment areas with different radii exhibit varying impacts on travel flow, in line with previous findings (Jun et al., 2015; Li, Zhao et al., 2020). It is suggested that many cities should adopt an approach to expanding service radii in their urban planning. Specifically, emphasis should be placed on the impacts of smaller radii land use in the city center and densely populated areas where TOD stations are concentrated. Conversely, in non-central and less populated areas, the primary focus should be on the impacts of larger radii area.

From this planning perspective, we advocate for customized strategies based on different geographical locations and station types. City managers should strive for precise alignment in TOD planning, ensuring station planning aligns with real demand. In-depth analyses of traffic flow data, population density, and employment distribution can better inform station locations and densities, preventing issues of under-planning or excessive station density. In particular, stations like the T1 cluster should facilitate more commerce and business land, which consistent with many small-scale studies using non-linear models (Shao et al., 2020; Xiao et al., 2021; Yang et al., 2021). Then, stations of the T2 cluster need more facilities related to residents' living, while the T3 stations will have more ridership by constructing more resident land near the stations.

Furthermore, TOD projects should actively encourage policies promoting mixed-use development. Results indicate that, in smaller or larger radii, Mixed Land Use (MLU) is the most significant influencing factor on passenger volume, particularly with a pronounced positive impact in smaller radii. This results also similar with relevant small-scale studies using non-linear models (Shao et al., 2020; Xiao et al., 2021). This finding aligns with the majority of TOD research, demonstrating that the organic integration of different land uses, such as industrial, residential, commercial centers, and cultural facilities, attracts diverse groups to use public transportation (Niu et al., 2019; Shao et al., 2020; Su et al., 2022). Mixed-use development enhances vibrancy around TOD stations, increases passenger flow, and optimizes the transportation network.

Moreover, city transportation planning should prioritize optimizing transport facilities and services at TOD stations to enhance the connection between stations and surrounding areas (Taki et al., 2017). This study found that residential neighborhoods remain a significant source of passenger volume in small catchment areas of TOD, followed by commercial areas and industrial parks (Li, Zhao et al., 2020). As the catchment area

radius increases, the importance of green spaces and administrative districts becomes more prominent. This suggests that the current connectivity of TOD with more distant functional spaces in the city needs to be sufficiently robust. Therefore, in addition to considering mixed land use, attention should also be given to station entrances and exits, transfer passages, bicycle parking areas, and other measures to improve station services and enhance passengers' travel experience.

Lastly, the results also show that the catchment areas of TOD should strategically plan for residents' travel needs and provide health and education facilities. Data analysis reveals that health and education facilities significantly contribute to TOD passenger flow, particularly in the city center and densely populated areas. Hence, Shenzhen should meticulously consider the travel demands of surrounding residents, ensuring adequate health services coverage to enhance the service quality of TOD stations and contribute to developing residential living circles.

The comprehensive findings of this study, analyzing the influence of land use on TOD ridership from various perspectives, provide valuable insights into the significance of land planning and mixed land use. Moreover, these results guide the city government in formulating scientifically driven transportation planning policies for the future.

5.2 Limitations and perspectives

This study delves into the impacts of urban land use on TOD ridership in various catchment areas within the highly developed city of Shenzhen. While we address the previous research limitations related to nonlinear model construction and multi-scale analysis, contributing to a more comprehensive understanding of Shenzhen's land use-TOD ridership relationship, it is crucial to acknowledge certain limitations that necessitate further investigation.

Our analysis focuses on the impacts of urban land use on TOD ridership, specifically in one city's subway station cases. The findings are highly relevant to medium-developed cities in China, offering a valuable supplement to current global TOD research. However, other researchers should conduct cross-sectional comparison studies across multiple cities for more generalized conclusions. Additionally, our study employs state-of-the-art LightGBM models for nonlinear modeling, using the model's rankings as a reference for various catchment planning. Nevertheless, there is room for improvement by incorporating additional metrics to elucidate further the effects of mixed land use on TOD ridership. For example, an in-depth discussion on how the mixed-use developments in a building can be categorized or combined can more detailly assist the design and planning at the micro-level.

Furthermore, while this research enhances understanding of the non-linear effects of land use on ridership and pedestrian flow in future TOD studies and planning, extending beyond a single small-scale study, it is essential to recognize the study's limitations. Despite the extensive use of data and interpretable machine learning methods, the findings are constrained by specific datasets and contexts. In particular, limited by data availability, this study could not use data from working days for further research analysis. Hence, comparing the differences in impacts that vary between weekdays and weekends is also meaningful.

6 Conclusions

This study seeks to identify the non-linear relationship between built environment features and ridership to optimize the TOD land-use pattern. To achieve this, three LightGBM classification models are applied to analyze three catchment areas, and the

SHAP value is used to decouple the mixed land use pattern within multiple urban TOD catchment radii. The primary conclusions drawn are as follows:

- 1) While residential and health land consistently rank highly, their significance varies slightly across buffer radius zones. As the radius increases, the land use type that contributes the least to ridership shifts from industry to education. Therefore, the land uses that significantly impact TOD ridership growth are residential, health, and industrial, while educational, green, cultural, and government land uses do not play a significant role in driving ridership growth.
- 2) Based on the outcomes of the machine learning model, it is evident that the contribution of various influencing factors to TOD passenger flow varies depending on the catchment radius. This observation highlights the significance of the catchment areas in controlling research results, emphasizing the need for classification and detailed analysis in our study.
- 3) Different types of TOD stations exhibit varying effects of land use on passenger flow at different buffer radii. For T1-type subway stations, commerce hurts TOD passenger flow in smaller buffer zones, but it positively impacts larger buffer zones. As for T2 sites, industry and health are more important when the buffer is small (500), while the significance of government increases as the buffer zone expands. Meanwhile, for T3 subway stations, residential land use consistently plays a crucial role in the passenger flow of the station, regardless of the buffer radius.

In summary, during the urbanization process, the government should carefully consider the impact of land uses on ridership at both small and large catchment radii. A comprehensive allocation of land use functions based on the varying human flow from different land areas is essential. Additionally, efforts should be made to enhance the efficiency of mixed land-use sites. These conclusions provide critical guidance for urban planners and policymakers in developing and implementing effective TOD strategies that align with the complex interplay between land use and transportation patterns. Additionally, the study sets the stage for further research in exploring the potential of nonlinear modeling and multi-scale analysis in TOD studies across different cities and regions.

Acknowledgment

This project was funded by the National Natural Science Foundation of China, General Program (52078217) and the Natural Science Foundation of Guangdong Province of China (2024A1515011998).

Author contribution

Xinyue Gu, Siyan Lin and Chengfang Wang conceived and planned the research. Xinyue Gu and Siyan Lin carried out the data processing, analysis, visualization and the interpretation of the results. Chengfang Wang provided critical feedback and helped shape the research, analysis and manuscript. All authors discussed the research, commented and revised on the manuscript.

References

- An, D., Tong, X., Liu, K., & Chan, E. H. W. (2019). Understanding the impact of built environment on metro ridership using open source in Shanghai. *Cities*, *93*, 177–187. <https://doi.org/10.1016/j.cities.2019.05.013>
- Ao, Y., Yang, D., Chen, C., & Wang, Y. (2019). Effects of rural built environment on travel-related CO2 emissions considering travel attitudes. *Transportation Research Part D: Transport and Environment*, *73*, 187–204. <https://doi.org/10.1016/j.trd.2019.07.004>
- Cao, Z., Asakura, Y., & Tan, Z. (2020). Coordination between node, place, and ridership: Comparing three transit operators in Tokyo. *Transportation Research Part D: Transport and Environment*, *87*, 102518. <https://doi.org/10.1016/j.trd.2020.102518>
- Cervero, R. (1989). Jobs-housing balancing and regional mobility. *Journal of the American Planning Association*, *55*(2), 136–150. <https://doi.org/10.1080/01944368908976014>
- Cervero, R. (2006). Alternative approaches to modeling the travel-demand impacts of smart growth. *Journal of the American Planning Association*, *72*(3), 285–295. <https://doi.org/10.1080/01944360608976751>
- Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219. [https://doi.org/10.1016/S1361-9209\(97\)00009-6](https://doi.org/10.1016/S1361-9209(97)00009-6)
- Cheng, L., De Vos, J., Zhao, P., Yang, M., & Witlox, F. (2020). Examining non-linear built environment effects on elderly's walking: A random forest approach. *Transportation Research Part D: Transport and Environment*, *88*, 102552. <https://doi.org/10.1016/j.trd.2020.102552>
- Choi, J., Lee, Y. J., Kim, T., & Sohn, K. (2012). An analysis of Metro ridership at the station-to-station level in Seoul. *Transportation*, *39*(3), 705–722. <https://doi.org/10.1007/s11116-011-9368-3>
- Ding, R., Ujang, N., Hamid, H. B., Manan, M. S. A., Li, R., Albadareen, S. S. M., ... & Wu, J. (2019). Application of complex networks theory in urban traffic network researches. *Networks and Spatial Economics*, *19*(4), 1281–1317. <https://doi.org/10.1007/s11067-019-09466-5>
- Estupiñán, N., & Rodríguez, D. A. (2008). The relationship between urban form and station boardings for Bogotá's BRT. *Transportation Research Part A: Policy and Practice*, *42*(2), 296–306. <https://doi.org/10.1016/j.tra.2007.10.006>
- Gan, Z., Yang, M., Feng, T., & Timmermans, H. J. P. (2020). Examining the relationship between built environment and metro ridership at station-to-station level. *Transportation Research Part D: Transport and Environment*, *82*, 102332. <https://doi.org/10.1016/j.trd.2020.102332>
- Gong, P., Chen, B., Li, X., Liu, H., Wang, J., Bai, Y., ... & Xu, B. (2020). Mapping essential urban land use categories in China (EULUC-China): Preliminary results for 2018. *Science Bulletin*, *65*(3), 182–187. <https://doi.org/10.1016/j.scib.2019.12.007>
- Gu, X., Tang, X., Chen, T., & Liu, X. (2024). Predicting the network shift of large urban agglomerations in China using the deep-learning gravity model: A perspective of population migration. *Cities*, *145*, 104680. <https://doi.org/10.1016/j.cities.2023.104680>
- Gu, X., Wu, Z., Liu, X., Qiao, R., & Jiang, Q. (2024). Exploring the nonlinear interplay between urban morphology and nighttime thermal environment. *Sustainable Cities and Society*, *101*, 105176. <https://doi.org/10.1016/j.scs.2024.105176>
- Harirchian, M., Azadi, M., Kermanshahi, S., Bashirinia, M., & Ghasri, M. (2021). Revisiting transit-oriented development evaluation for urban master plans in the

- context of developing countries. *Transportation Research Record*, 2675(10), 1069–1082. <https://doi.org/10.1177/03611981211014526>
- Hasibuan, H. S., Soemardi, T. P., Koestoer, R., & Moersidik, S. (2014). The role of transit oriented development in constructing urban environment sustainability, the case of Jabodetabek, Indonesia. *Procedia Environmental Sciences*, 20, 622–631. <https://doi.org/10.1016/j.proenv.2014.03.075>
- Huang, H.-J., Xia, T., Tian, Q., Liu, T.-L., Wang, C., & Li, D. (2020). Transportation issues in developing China's urban agglomerations. *Transport Policy*, 85, A1–A22. <https://doi.org/10.1016/j.tranpol.2019.09.007>
- Ingvardson, J. B., & Nielsen, O. A. (2018). How urban density, network topology and socio-economy influence public transport ridership: Empirical evidence from 48 European metropolitan areas. *Journal of Transport Geography*, 72, 50–63. <https://doi.org/10.1016/j.jtrangeo.2018.07.002>
- Iseki, H., Liu, C., & Knaap, G. (2018). The determinants of travel demand between rail stations: A direct transit demand model using multilevel analysis for the Washington D.C. metrorail system. *Transportation Research Part A: Policy and Practice*, 116, 635–649. <https://doi.org/10.1016/j.tra.2018.06.011>
- Jun, M.-J., Choi, K., Jeong, J.-E., Kwon, K.-H., & Kim, H.-J. (2015). Land use characteristics of subway catchment areas and their influence on subway ridership in Seoul. *Journal of Transport Geography*, 48, 30–40. <https://doi.org/10.1016/j.jtrangeo.2015.08.002>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems* 30. Retrieved from <https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html>
- Krishna, K., & Narasimha Murty, M. (1999). Genetic K-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29(3), 433–439. <https://doi.org/10.1109/3477.764879>
- Kuhn, H. W., & Tucker, A. W. (2016). *Contributions to the theory of games (AM-28)*, Volume II. Princeton, NJ: Princeton University Press.
- Lee, K.-I., Kim, K.-J., & Kwon, S.-J. (2005). A Study on characteristics of subway utilization and pedestrians' accessibility at new towns in Korea. *Journal of Asian Architecture and Building Engineering*, 4(1), 85–95. <https://doi.org/10.3130/jaabe.4.85>
- Li, P., Yu, Y., Wang, Z., & Zhang, F. (2024). Analysis of the external attraction of Shanghai urban functions based on the travel characteristics. *Urban Informatics*, 3(1), 11. <https://doi.org/10.1007/s44212-024-00041-z>
- Li, P., Zhao, P., & Schwanen, T. (2020). Effect of land use on shopping trips in station areas: Examining sensitivity to scale. *Transportation Research Part A: Policy and Practice*, 132, 969–985. <https://doi.org/10.1016/j.tra.2019.12.029>
- Li, S., Lyu, D., Huang, G., Zhang, X., Gao, F., Chen, Y., & Liu, X. (2020). Spatially varying impacts of built environment factors on rail transit ridership at station level: A case study in Guangzhou, China. *Journal of Transport Geography*, 82, 102631. <https://doi.org/10.1016/j.jtrangeo.2019.102631>
- Lin, J. J., & Gau, C. C. (2006). A TOD planning model to review the regulation of allowable development densities around subway stations. *Land Use Policy*, 23(3), 353–360. <https://doi.org/10.1016/j.landusepol.2004.11.003>
- Lin, Y., Huang, H., Gu, X., Xie, Y., Li, J., & Wang, C. (2022). Identification and evaluation of urban public centers from temporal and spatial perspectives: A case

- study of six districts in Guangzhou. *Proceedings of ICCREM*, 594–607. <https://doi.org/10.1061/9780784484562.065>
- Loo, B. P. Y., Chen, C., & Chan, E. T. H. (2010). Rail-based transit-oriented development: Lessons from New York City and Hong Kong. *Landscape and Urban Planning*, 97(3), 202–212. <https://doi.org/10.1016/j.landurbplan.2010.06.002>
- Ma, X., Zhang, J., Ding, C., & Wang, Y. (2018). A geographically and temporally weighted regression model to explore the spatiotemporal influence of built environment on transit ridership. *Computers, Environment and Urban Systems*, 70, 113–124. <https://doi.org/10.1016/j.compenvurbsys.2018.03.001>
- Marilee A, U. (2016). Developing TOD in America: The private sector view. In *Transit oriented development* (pp. 229–244). London: Routledge. <https://doi.org/10.4324/9781315550008-27>
- Memon, N., Patel, S. B., & Patel, D. P. (2019). Comparative analysis of artificial neural network and XGBoost algorithm for PolSAR image classification. In B. Deka, P. Maji, S. Mitra, D. K. Bhattacharyya, P. K. Bora, & S. K. Pal (Eds.), *Pattern recognition and machine intelligence* (pp. 452–460). Cham, Switzerland: Springer International Publishing. https://doi.org/10.1007/978-3-030-34869-4_49
- Niu, S., Hu, A., Shen, Z., Lau, S. S. Y., & Gan, X. (2019). Study on land use characteristics of rail transit TOD sites in new towns—Taking Singapore as an example. *Journal of Asian Architecture and Building Engineering*, 18(1), 16–27. <https://doi.org/10.1080/13467581.2019.1586712>
- Nyunt, K. T. K., & Wongchavalidkul, N. (2020). Evaluation of relationships between ridership demand and transit-oriented development (TOD) indicators focused on land use density, diversity, and accessibility: A case study of existing metro stations in Bangkok. *Urban Rail Transit*, 6(1), 56–70. <https://doi.org/10.1007/s40864-019-00122-2>
- Ramlan, N. H., Osman, M. M., Rabe, N. S., Azizan, A., Azmi, N. A., & Amiruddin, S. (2021). Assessing the current implementation of compact and mixed-use development within public rail transit stations in Malaysia. *Planning Malaysia*, 19(1), 174–185. <https://doi.org/10.21837/pm.v19i15.934>
- Renne, J. L., & Appleyard, B. (2019). Twenty-five years in the making: TOD as a new name for an enduring concept. *Journal of Planning Education and Research*, 39(4), 402–408. <https://doi.org/10.1177/0739456X19885351>
- Shao, Q., Zhang, W., Cao, X., Yang, J., & Yin, J. (2020). Threshold and moderating effects of land use on metro ridership in Shenzhen: Implications for TOD planning. *Journal of Transport Geography*, 89, 102878. <https://doi.org/10.1016/j.jtrangeo.2020.102878>
- Sinaga, K. P., & Yang, M.-S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Sohn, K., & Shim, H. (2010). Factors generating boardings at metro stations in the Seoul metropolitan area. *Cities*, 27(5), 358–368. <https://doi.org/10.1016/j.cities.2010.05.001>
- Song, H., Xie, Y., Gu, X., & Wang, C. (2022). Land reserve potential evaluation based on GIS and big data analysis: A case study of Nansha District, Guangzhou, China. 496–508. *Proceedings of the ICCREM*. <https://doi.org/10.1061/9780784484562.054>
- Štrumbelj, E., & Kononenko, I. (2014). Explaining prediction models and individual predictions with feature contributions. *Knowledge and Information Systems*, 41(3), 647–665. <https://doi.org/10.1007/s10115-013-0679-x>
- Su, S., Wang, Z., Li, B., & Kang, M. (2022). Deciphering the influence of TOD on metro ridership: An integrated approach of extended node-place model and interpretable machine learning with planning implications. *Journal of Transport Geography*, 104, 103455. <https://doi.org/10.1016/j.jtrangeo.2022.103455>

- Su, S., Zhao, C., Zhou, H., Li, B., & Kang, M. (2022). Unraveling the relative contribution of TOD structural factors to metro ridership: A novel localized modeling approach with implications on spatial planning. *Journal of Transport Geography*, *100*, 103308. <https://doi.org/10.1016/j.jtrangeo.2022.103308>
- Sung, H., & Oh, J.-T. (2011). Transit-oriented development in a high-density city: Identifying its association with transit ridership in Seoul, Korea. *Cities*, *28*(1), 70–82. <https://doi.org/10.1016/j.cities.2010.09.004>
- Taki, H. M., Maatouk, M. M. H., Qurnfulah, E. M., & Aljoufie, M. O. (2017). Planning TOD with land use and transport integration: A review. *Journal of Geoscience, Engineering, Environment, and Technology*, *2*(1), 84–94. <https://doi.org/10.24273/jgeet.2017.2.1.17>
- Tao, T., Wu, X., Cao, J., Fan, Y., Das, K., & Ramaswami, A. (2020). Exploring the nonlinear relationship between the built environment and active travel in the Twin Cities. *Journal of Planning Education and Research*, *43*(3), 0739456X20915765. <https://doi.org/10.1177/0739456X20915765>
- van Wee, B., & Handy, S. (2016). Key research themes on urban space, scale, and sustainable urban mobility. *International Journal of Sustainable Transportation*, *10*(1), 18–24. <https://doi.org/10.1080/15568318.2013.820998>
- Wang, F., Zheng, Y., Wu, W., & Wang, D. (2022). The travel, equity and wellbeing impacts of transit-oriented development in Global South. *Transportation Research Part D: Transport and Environment*, *113*, 103512. <https://doi.org/10.1016/j.trd.2022.103512>
- Wu, Z., Qiao, R., Zhao, S., Liu, X., Gao, S., Liu, Z., ... & Jiang, Q. (2022). Nonlinear forces in urban thermal environment using Bayesian optimization-based ensemble learning. *Science of the Total Environment*, *838*, 156348. <https://doi.org/10.1016/j.scitotenv.2022.156348>
- Xiao, L., Lo, S., Liu, J., Zhou, J., & Li, Q. (2021). Nonlinear and synergistic effects of TOD on urban vibrancy: Applying local explanations for gradient boosting decision tree. *Sustainable Cities and Society*, *72*, 103063. <https://doi.org/10.1016/j.scs.2021.103063>
- Yang, H., Zheng, R., Li, X., Huo, J., Yang, L., & Zhu, T. (2022). Nonlinear and threshold effects of the built environment on e-scooter sharing ridership. *Journal of Transport Geography*, *104*, 103453. <https://doi.org/10.1016/j.jtrangeo.2022.103453>
- Yang, J., Cao, J., & Zhou, Y. (2021). Elaborating non-linear associations and synergies of subway access and land uses with urban vitality in Shenzhen. *Transportation Research Part A: Policy and Practice*, *144*, 74–88. <https://doi.org/10.1016/j.tra.2020.11.014>
- Yen, B. T. H., Feng, C.-M., & Lee, T.-C. (2023). Transit-oriented development strategy in Taiwan: An application of land value capture. *Asian Transport Studies*, *9*, 100094. <https://doi.org/10.1016/j.eastsj.2022.100094>
- Zhang, Z., Qian, Z., Zhong, T., Chen, M., Zhang, K., Yang, Y., ... & Yan, J. (2022). Vectorized rooftop area data for 90 cities in China. *Scientific Data*, *9*, 66. <https://doi.org/10.1038/s41597-022-01168-x>
- Zhou, J., & Yang, Y. (2021). Transit-based accessibility and urban development: An exploratory study of Shenzhen based on big and/or open data. *Cities*, *110*, 102990. <https://doi.org/10.1016/j.cities.2020.102990>