The study explores the nonlinear effects of public transport accessibility on the urban development of Chongqing, a mountainous city in China. Based on Chongqing's metro service areas (MSAs) and big open data, point-of-interest (POI) data obtained from the Map Application Programming Interface (API) has been used to represent the level of urban development. Comprehensive city, road, and metro maps, including data from open street maps, were utilized to procure information on road and metro accessibility. Employing Gradient Boosting Decision Trees (GBDT), a supervised machine-learning technique, the study effectively evaluates the accessibility of public transport in Chongqing by examining the nonlinear interplay between public transportation accessibility and urban development.

1 Intro

1.1 Effects of public transport systems on urban development patterns

Public transport investments can shape urban form and land use in several ways. Some studies highlight how high-capacity transit enables more compact, high-density urban development by reducing reliance on automobiles (Hurst, 2011; Knowles, 2012). Others examine public transport's role in achieving sustainable urbanization goals like reducing energy use, air pollution, and traffic congestion (Sharma and Patil, 2022). Additionally, research has identified symbiotic relationships where transportation infrastructure influences surrounding land uses, which in turn generate new travel demand patterns influencing future transport investments (Rode *et al.*, 2017).

The transit-oriented development (TOD) concept has emerged as an intensive, mixed-use urban model centering growth around transit nodes and promoting public transport use (Calthorpe, 1993). While early TOD research focused on impacts to travel behavior, more recent work has analyzed effects on urban development intensity, land values, job accessibility and economic development (Kamruzzaman *et al.*, 2014; Dong, 2016). Many studies have begun to explore this relationship through techniques such as spatial analysis, descriptive statistics, and regression modeling (Xiao *et al.*, 2021; Zhou and Yang, 2021), these efforts have predominantly assumed linear associations between transit accessibility metrics and indicators of urban development, such as employment/population densities and concentrations of points of interest. The potential for nonlinear effects, wherein the marginal impacts of transit accessibility on urban development may diminish or accelerate beyond certain thresholds, has been relatively underexplored (Xu and Yang, 2019).

1.2 Public Transit Accessibility Evaluation

The concept of accessibility encapsulates the ease with which desired destinations and activities can be reached from a given location (Geurs & van Eck, 2001). It serves as a critical determinant of both transit ridership levels and land use dynamics. Scholars have developed a variety of methodologies to evaluate public transit accessibility, typically grounded in the spatial distributions of transit stops/stations, service frequencies, travel times, and the destinations served (Mavoa et al., 2012).

Empirical evidence suggests that higher levels of transit accessibility tend to be correlated with more compact urban development patterns and elevated real estate valuations (Wang *et al.*, 2015; Gong, 2021). The relationship between transit accessibility and urban development may be moderated by factors such as a city's spatial structure, predominant transit modes, and physical geography. Many research has focused on monocentric cities situated in relatively flat terrain, where accessibility exhibits a distance-decay pattern emanating from the central business district (CBD). The dynamics of this relationship in mountainous or polycentric cities with more dispersed development patterns have not been extensively studied (Hickman and Hall, 2008).

Despite this extensive theoretical and empirical research, gaps remain in understanding the precise nature of the relationship between public transport accessibility and urban development levels. The majority of quantitative studies to date have assumed linear relationships between transit accessibility metrics and urban development variables. Potential nonlinearities and threshold effects have not been adequately explored. This study addresses this gap through a case analysis combining big data, machine learning, and interpreting nonlinear effects. Despite the growing availability of new big data sources to measure the built environment and human activities, much of the empirical TOD research still relies on traditional data sources such as population/employment statistics, which may not fully capture urban development patterns.

2 Q

What is the relationship between public transport accessibility and urban development in mountainous cities like Chongqing?

What are the specific impacts of different modes of public transport (e.g., metro, bus, street accessibility) on urban development?

3 method

To investigate the relationship between public transport accessibility and urban development levels in the mountainous city of Chongqing, this study employs a data-driven analytical approach combining big data sources, machine learning algorithms, and interpretable modeling techniques.

The study employs Gradient Boosting Decision Trees (GBDT) to model the potentially nonlinear relationships between the collected variables. GBDT is an ensemble method that incrementally combines weak regression tree models to optimize predictive performance. Crucial advantages include high accuracy, ability to automatically handle non-linearities and interactions, and inbuilt prevention of overfitting.

To interpret the resulting GBDT model's nonlinear effects and each variable's relative importance, a method called SHapley Additive exPlanations (SHAP) is used. SHAP computes Shapley values, a game theory-based approach to fairly distribute the prediction among feature contributions. Visual techniques like dependency plots are then employed to scrutinize variable influence across their data ranges.

The complete analytical workflow spans data collection and cleaning, GBDT model training on Chongqing's metro service area samples using the curated variables, interpreting the GBDT's nonlinear effects via SHAP values and dependence plots, and finally deriving planning implications based on the nonlinear effects revealed for different types of stations/neighborhoods.

4 data

We collected data in year 2024, and the indicator system utilized in this study was tailored to the particular context after being derived from the one used in Shenzhen (Xiao *et al.*, 2021).

The level of urban development is quantified using point of interest (POI) data obtained from a major online mapping platform's API. POIs represent geographic locations and entities crucial for daily urban activities like restaurants, shopping centers, businesses, leisure entertainment, life sevices, medical services, real estate, and schools. Their density and concentration can serve as a proxy for development intensity. The Chongqing Municipal Government's TOD special planning document also mandates that comprehensive development areas be built using a station radius length of 600 meters, while over 730,000 filtered POIs falling within 600m radius metro service areas were collected.

Regional urban development is facilitated by well-maintained streets. Public transport accessibility measures were derived from several spatial data sources, which was calculated using the proportion of streets in an MSA having betweenness values in the top 30%. Road network data from OpenStreetMaps enabled calculating street-level betweenness centrality, which counts how many times a link's midpoint follows the shortest routes to connect any two street segments that are less than 600 meters apart.

Other relevant variables extracted include the number of bus stops in each metro service area, walking times to the central business district, station characteristics (interchange status), and average residential property prices as a control variable.

5 ex

6 analysis

The GBDT model is built with LightGBM. The number of POIs is set as a dependent variable, and station attributes, accessibility to public transportation, and control variables are set as independent variables.

The mean Shap value for each variable is represented by the horizontal coordinates. The average marginal contribution of a variable over all other variables is its Shap value. The graph's left side displays the variables in decreasing order of global relevance, while the graph's right side illustrates how each MSA's variable values affect urban development.

7 dis

Q

The study found nonlinear and threshold effects of public transport accessibility factors like closeness centraility, street accessibility, and number of bus stops on urban development levels measured by number of POIs in metro service areas of Chongqing.

Metro betweenness and street accessibility measures had the strongest positive influence, followed by some other metro accessibility metrics. Bus stop density had a relatively weaker effect.

The predominant factor contributing to the expansion of urban areas in Chongqing is the density of bus stations, which underscores the critical role of access to public transportation in fostering urban development.

The significance of housing cost as a variable suggests its potential as an influencing factor on the growth patterns within MSAs. High real estate values often mirror the attractiveness and thriving economic conditions of a locale, indicating that residential zones are key contributors to the growth of urban regions. The challenging geographical position of Chongqing could lead to escalated construction expenses, which in turn may inflate property prices. Areas that are strategically planned around transit routes and amenities are likely to experience an appreciation in property value.

The study also indicates a dispersion of urban development throughout various city nodes rather than a concentration in the CBD, as the impact of a location within the CBD is minimal. This observation counters the initial prediction of CBDs as focal points of vigorous development and highlights a more decentralized urban growth. The assessment does not reveal any substantial influence of either the age of transit stations or their status as transfer hubs on urban expansion. This insight emerged despite initial expectations of such factors serving as indicators of development longevity or nodal activity intensity. The analysis also reflects that the evaluation of transfer station significance was conducted using a simplistic binary approach, and that the estimation of CBD proximity exclusively considered public transit, omitting the impact of personal vehicular access. Despite the prominence of the Jiefangbei-Jiangbeizui-Danzishi district as the recognized core of Chongqing, urban development is not confined to this region alone.

According to Manout et al. (2018), transit accessibility is widely regarded as a crucial objective in the promotion of transit use and policies. Accessibility to public transportation and urban development have been linked in a number of studies (Rode et al., 2017; Zhou & Yang, 2021). To the best of our knowledge, no previous research has examined the nonlinear relationship between the two, though. The objective of this research is to examine, through quantitative analysis, the non-linear and threshold effects of public transportation accessibility on urban development.

Because there is a lack of data specifically for mountainous cities, the accessibility and transport characteristics used in the study, which were adapted from earlier research on plain cities, might not be applicable there. Furthermore, the study's definition of "MSA" is based on conventional circular areas, which might not be suitable for hilly cities with erratic road systems. Future research could address these limitations by establishing a more varied system for evaluating accessibility characteristics based on the unique features of road networks and pedestrian systems in mountainous cities, as well as by using social media check-in data and other forms of data to obtain more accurate POI data.