Cities 158 (2025) 105661

Contents lists available at ScienceDirect

Cities

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

Evaluation of green space influence on housing prices using machine learning and urban visual intelligence

Yanqing Xu [a], Ruidun Chen [a], Hongyu Du [b], Meixu Chen [c], Cong Fu [a,*], Yuchen Li [d,e,**]

[a] School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, Hubei 430079, China
[b] Institute of Ecology and Sustainable Development, Shanghai Academy of Social Sciences, Shanghai 200020, China
[c] Geographic Data Science Lab, Department of Geography and Planning, University of Liverpool, Liverpool L69 7ZT, United Kingdom
[d] School of Geography, University of Leeds, Leeds LS2 9JT, United Kingdom
[e] MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Cambridge CB2 0QQ, United Kingdom

ARTICLE INFO

Keywords:
Urban greenness
Housing market value
Street view imagery
Random Forest model
Urban planning

ABSTRACT

Green spaces are recognised for enhancing the aesthetic value and health benefits in urban environments, which, in turn, can influence housing prices. This study evaluates the impact of visible green spaces on housing prices in Lucas County, USA, employing an innovative approach that contrasts land use data (NGVI) and street view imagery (AGVI) as quantified indicators. Leveraging a Random Forest model from 2017 to 2019, we determined the contribution of green spaces to housing prices. The Analytic Hierarchy Process (AHP) was then used to score each independent variable based on its ranking performance, thereby assessing the significance of methodological differences in environmental valuation. Our findings reveal that while AGVI typically contributes more to housing price evaluations than NGVI, the primary determinants of housing prices are still the intrinsic property characteristics and socioeconomic factors, furthermore, we observed temporal variability in the effects of visible green space on housing prices. While previous research often suggested a clear link between green space and higher property values, our result indicates this relationship may be more location-dependent. Our research highlights the importance of not overestimating the economic impact of green spaces when planning urban development. Furthermore, our research underscores the necessity of adopting a diverse methodological framework when appraising environmental attributes in housing markets, considering both objective land use data and subjective visual assessments.

1. Introduction

Housing has always been an important issue related to people's livelihood. Housing prices within cities and urban areas act as a market signal of desirability (Galster & Rothenberg, 1991). Higher prices reflect factors like job opportunities, amenities, and quality of life (Gibbons & Machin, 2008). By analysing spatial variations in housing prices, we can evaluate a city's development patterns, identify areas for investment or revitalization, and assess socio-economic disparities. Studies have identified various factors influencing housing prices. At the national level, government revenue, real estate investment, and land value play a key role in China (Yang, 2022). In Lithuania, economic factors like GDP and unemployment alongside policy measures like macroprudential policies and past housing prices are significant determinants (Cohen & Karpavičiūtė, 2017). Cross-country analyses highlight the influence of affordability metrics like rent-to-income and price-to-rent ratios, alongside broader economic indicators like urbanization, per capita GDP, inflation, and demographic trends (Tripathi, 2019). Beyond national and economic factors, studies have explored the impact of property characteristics. Fan et al. (2006) employed a decision tree approach in Singapore, demonstrating the importance of features like floor area, model type, age, and level within a building.

Looking beyond the property itself, research suggests that a house's value is also influenced by its surroundings and neighbourhood amenities, such as accessibility to utilities, social context, and physical environment (Hadavi et al., 2018; Kwon et al., 2019; Lee et al., 2017; Plane & Klodawsky, 2013). Understanding these environmental factors, unlike inherent property characteristics, offers valuable insights for

* Corresponding author.
** Correspondence to: Y. Li, School of Geography, University of Leeds, Leeds LS2 9JT, United Kingdom.
E-mail addresses: cong.fu@whu.edu.cn (C. Fu), y.li3@leeds.ac.uk (Y. Li).

https://doi.org/10.1016/j.cities.2024.105661
Received 4 March 2024; Received in revised form 14 October 2024; Accepted 9 December 2024
Available online 21 December 2024

urban planning as they can be strategically incorporated into development projects. Chen et al. (2020) found a non-linear relationship between housing prices and urban environmental elements in Shanghai, with higher prices associated with access to green spaces and water features. Similarly, Julius et al. (2020) discovered that environmental characteristics like vegetation, sewage systems, and water supply significantly influenced housing prices in Gweru, Zimbabwe, along with proximity to the central business district. The non-linear relationships between greenness and housing prices likely arise due to threshold effects and diminishing returns. In regions where green space is initially sparse, small increases in green coverage may lead to significant increases in property value. However, as the amount of green space continues to rise, its marginal impact on housing prices diminishes, as buyers may place less additional value on green space once a certain threshold is met. This non-linear pattern suggests that the relationship between green space and property values is more complex than linear models imply, with local context playing a crucial role. Although it's recognised that environmental features can affect property values, their precise financial impact often remains nuanced and can be overshadowed by broader urban development objectives. This complexity sometimes leads to these values being overlooked in urban planning and policy decisions, causing the risk of commercial expansion and urban growth to compromise local environmental quality. Moreover, the challenge of accurately measuring these environmental attributes at a localised level persists, as traditional methods for environmental audit can be prohibitively expensive and inherently subjective.

Since the inception of Rosen's classical hedonic pricing model in 1974, quantifying the environmental influence on housing prices at a micro-scale has witnessed a reduction in complexity. Consequently, there has been an upswing in scholarly investigations scrutinising the environmental factors that influence housing prices (Geng et al., 2015; Goodman & Thibodeau, 2003; Liu et al., 2020). Understanding the link between visible green spaces and housing prices is crucial, as access to nature has been shown to affect property values (Bockarjova et al., 2020) and this knowledge can inform urban planning decisions, promoting development that integrates green spaces and fosters sustainable, high-value communities. The reasons why green spaces can influence housing prices are varied, including elements such as improving the aesthetic value of a neighbourhood and promoting physical activities (Lu, 2019; Sang et al., 2016). Nevertheless, it is imperative to acknowledge that the approaches used for quantifying green space indicated notable variations despite their position as influential factors in housing prices that can not be ignored in prior research. Research consistently shows that the influence of green space on housing prices varies based on factors such as the size, quality, and location of the green space. Kim et al. (2018) found that larger tree and urban forest areas positively contribute to property values, while fragmented, isolated, and irregularly shaped landscape patterns have a negative impact. Trojanek et al. (2018) highlighted the significant positive effects of proximity to urban green areas on apartment prices, particularly for newer buildings and post-transformation housing estates. Cho et al. (2008) added a spatial perspective, noting that the amenity values of different green space features vary based on the degree of urbanization, suggesting the need for site-specific land use management.

Traditional quantitative methods such as questionnaire surveys, often used alongside land use data analysis (Liu et al., 2020; Piaggio, 2021; Zambrano-Monserrate et al., 2021), play a valuable role in capturing residents' experiences and perceptions of green spaces. These methods provide direct insight into the subjective value residents place on green spaces, which can significantly influence property valuations. However, a limitation of traditional surveys is their labor-intensive nature. Conducting surveys at a detailed geographic level across a large urban scale can be challenging. This can potentially lead to underrepresentation of diverse resident perspectives.

To address this limitation, our study employs street view imagery as an alternative approach to capture perceived green space. This method allows for a more comprehensive assessment at a broader urban scale while still capturing the crucial element of subjective perception. Using street-view images effectively addresses this deficiency (Yin et al., 2023). Street view images are crucial in urban analysis and geographic science, offering a human-centric perspective to describe the environment (Biljecki & Ito, 2021). Simultaneously, it is imperative to underscore that diverse street view images include extensive semantic information, facilitating a thorough and detailed depiction of environmental elements from various geographic perspectives (Middel et al., 2019). Past technological constraints have posed challenges in extracting information from street-view images. Using such data for academic research started gaining more attention in the evolution of machine learning, which introduced Convolutional Neural Networks (CNNs) as a pivotal tool for information extraction from street view images (Wu et al., 2020). CNNs, structured with convolutional layers, pooling layers, and fully connected layers, constitute deep learning models explicitly designed for image recognition and computer vision tasks, enabling the identification of objects or segmentation of features in images. The application of CNNs has facilitated the extraction of features from street view images, contributing to the growing integration of urban elements in diverse research domains. These investigations span a range of topics, including understanding criminal behaviour, analysing the urban landscape, and exploring the social and environmental factors that influence road accidents and health outcomes (Li et al., 2022; McKee et al., 2017; Stiles et al., 2022; Wu et al., 2020).

In recent years, semantic segmentation techniques based on CNNs have been used to extract visible green spaces from street view images, and studies have been conducted to assess their impact on house prices (Wu et al., 2020; Wu et al., 2022). However, no study has analysed the differences in assessment results in which land use data is used to quantify visible green spaces compared to street view images. Analysing the differences in assessments of visible green spaces between street view imagery and land use data is vital, as street view imagery, captured from the street, may miss green spaces not visible from the streetscape, while land use data may omit greenery that is only observable at the eye-level perspective of residents. This comparison is necessary to bridge the gap between the macro accuracy of land use data and the micro visibility of green spaces that street view imagery provides, affecting housing price evaluations. More critically, it is imperative to gauge the magnitude of influence these differences may exert on homebuyers' assessment of housing prices. Moreover, housing prices demonstrate non-stationarity in the time dimension, and the impact of diverse factors on housing prices, including visible green spaces, can also indicate temporal variations (Soltani et al., 2022). Consequently, assessing whether differences in assessment results indicate temporal heterogeneity becomes a notable concern. This study aims to resolve these challenges.

In this research, we comprehensively analysed the variations in evaluation outcomes that may arise from employing different methodologies to quantify visible green spaces when assessing factors influencing housing prices. Initially, we compiled housing price data from 2017 to 2019 for Lucas County, USA. Based on a comprehensive literature review, we meticulously identified 26 significant factors influencing housing prices. These factors encompass three key domains: housing structure, neighbourhood socio-demographic characteristics, and environmental variables. Subsequently, land use data and street view imagery were independently utilised to quantify an additional independent variable: visible green space. In the model construction phase, recognising the temporal heterogeneity inherent in housing prices, we employed a random forest model to conduct yearly regression analyses of housing prices in Lucas County from 2017 to 2019. During the stage of comparing results, we compared the contributions obtained from quantifying visible green spaces using land use data against those derived from an approach that uses street view images. This comparative analysis aimed to discern the disparities in evaluation results attributable to different quantification methods. Finally, we applied the

Analytic Hierarchy Process to explicate the potential ramifications of these differences on the assessment of housing prices. These findings enhance our understanding of the intricate relationship between visible green spaces and housing prices and offer valuable reference points for future research endeavours in housing price prediction. Fig. 1 presents the overall experimental workflow of this study. Data processing, analysis methods, and other experimental details will be thoroughly discussed in the "Data and Methods" sections.

2. Data and methods

2.1. Study area

This study was conducted in Lucas County, Ohio, USA. The location of the study area is illustrated in Fig. 2. Lucas County borders Lake Erie to the east, and the Maumee River to the southeast, which flows into the lake. Lucas County is a part of the Toledo Metropolitan Area (also named Greater Toledo) which includes Fulton, Lucas, and Wood. The Greater Toledo area has strong ties to Metro Detroit which is located 40 miles north. According to the Census 2020 data, Lucas County has a total area of 596 mile^2, of which 341 mile^2 is land (57.2 %) and 255 mile^2 is water (42.8 %). The population was 431,279 and the population density was 1264.7/sq. mi. There were 200,856 housing units at an average density of 589 per square mile. The racial makeup of the county was 73.6 % white, 20.5 % Black, or African American (https://www.census.gov/quickfacts/fact/table/lucascountyohio/PST045221).

2.2. Housing price data

We obtained housing sales data from the Lucas County Auditor's Office from 2017 to 2019, accessible at https://co.lucas.oh.us/. Given the specific focus of this research, all data pertained to commercial real estate was meticulously omitted, resulting in a refined dataset comprising 17,251 house sales records (85.6 % of the original house price data). Each house has its property value, which is the last transaction price of the house.

2.3. Independent variables

Auditor's database also contains variables that characterise the house, such as the year built, the year of the sale, the size of the house area, the size of the building area, the number of bedrooms, the number of bathrooms, the garage area, and the quality of the house, from which we selected 8 to add to the independent variables.

In addition to the characteristics of the house itself, the socioeconomic conditions in the geographical area where the house is situated also influence the house price. To address this, we accessed block group-level census data specific to Lucas County, USA, derived from the 2020 American Community Survey (ACS). Subsequently, varying socioeconomic attributes were considered as independent variables after a thoughtful selection, including population demographics, ethnic composition, income levels, educational attainment, employment statistics, poverty rates, and healthcare coverage for each block group. The total number of independent variables selected was 11, and most variables are expressed as percentages except for population.

The built environment surrounding a property is a critical determinant that affects its value (Qiu et al., 2022). The built environment of a neighbourhood can vary significantly from one place to another, especially in a socially divided U.S. city (Lee et al., 2020). Understanding the urban built environment of the neighbourhood at a detailed level may enrich the semantics of the place, reveal a pattern of heterogenous urbanity, and help better understand its association with price. In this study, the term 'locational amenities' pertains to the facilities near the house and constituted one of the initial factors examined in quantifying the built environment. The Point of Interest (POI) information was used to show the location characteristics of nearby properties, which include parks, grocery stores, hospitals, universities, and schools. The distance from each property to the nearest POI was calculated using Network Analysis in ArcGIS 10.8. Second, street connectivity and walkability were added as built environmental variables in this research to ensure the consistency and comprehensiveness of variable selection with related research (Wang et al., 2013; Xu & Wang, 2014). Street connectivity was defined by the number of intersections along a specific street network or in an area. Intersections with a starting or ending node of an edge or an intersection of 3-way or more edges were included in the connectivity index calculation. Intersection density corresponds closely to block size - the greater the intersection density, the smaller the blocks. Small blocks make a neighbourhood walkable. Walkability was quantified by (http://www.walkscore.com/) based on the algorithm developed by Front Seat Management (http://www.frontseat.org/). It calculates the Euclidean distances from the point of interest to nearby amenities such as food, retail, education, parks, restaurants, recreation, and entertainment, and then integrates them by a linear combination of these distances with weights that account for facility type priority and a distance decay function. There are seven independent variables related to the built environment, as described above, and 26 independent variables in total, excluding visible green space. Table 1 provides a detailed description of each of the 26 variables and their data characteristics.

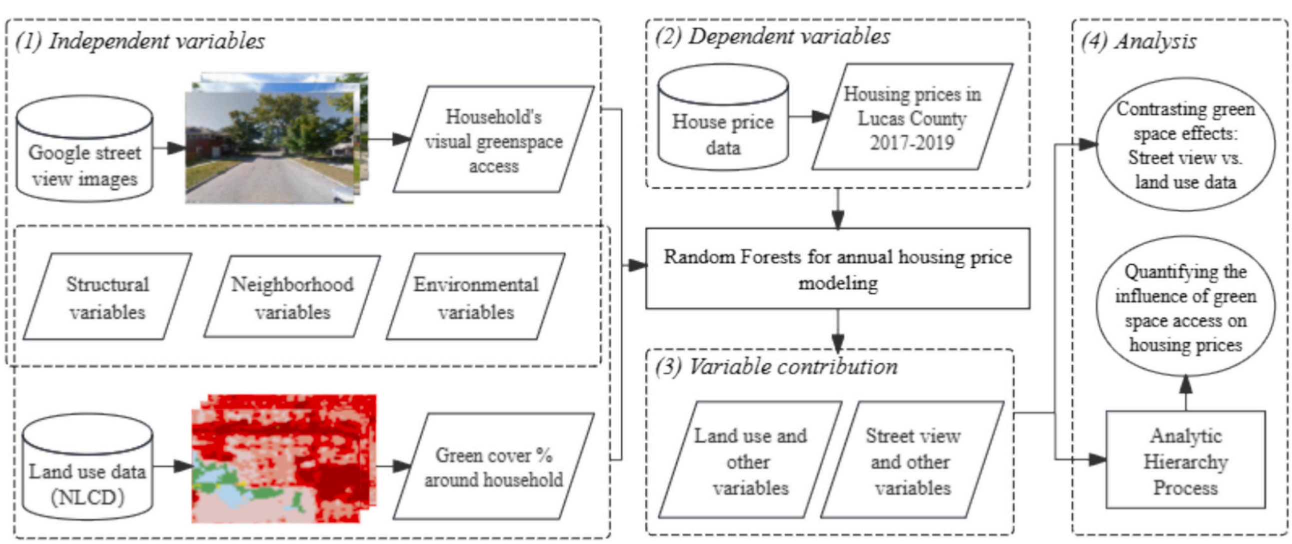

Fig. 1. Overall workflow.

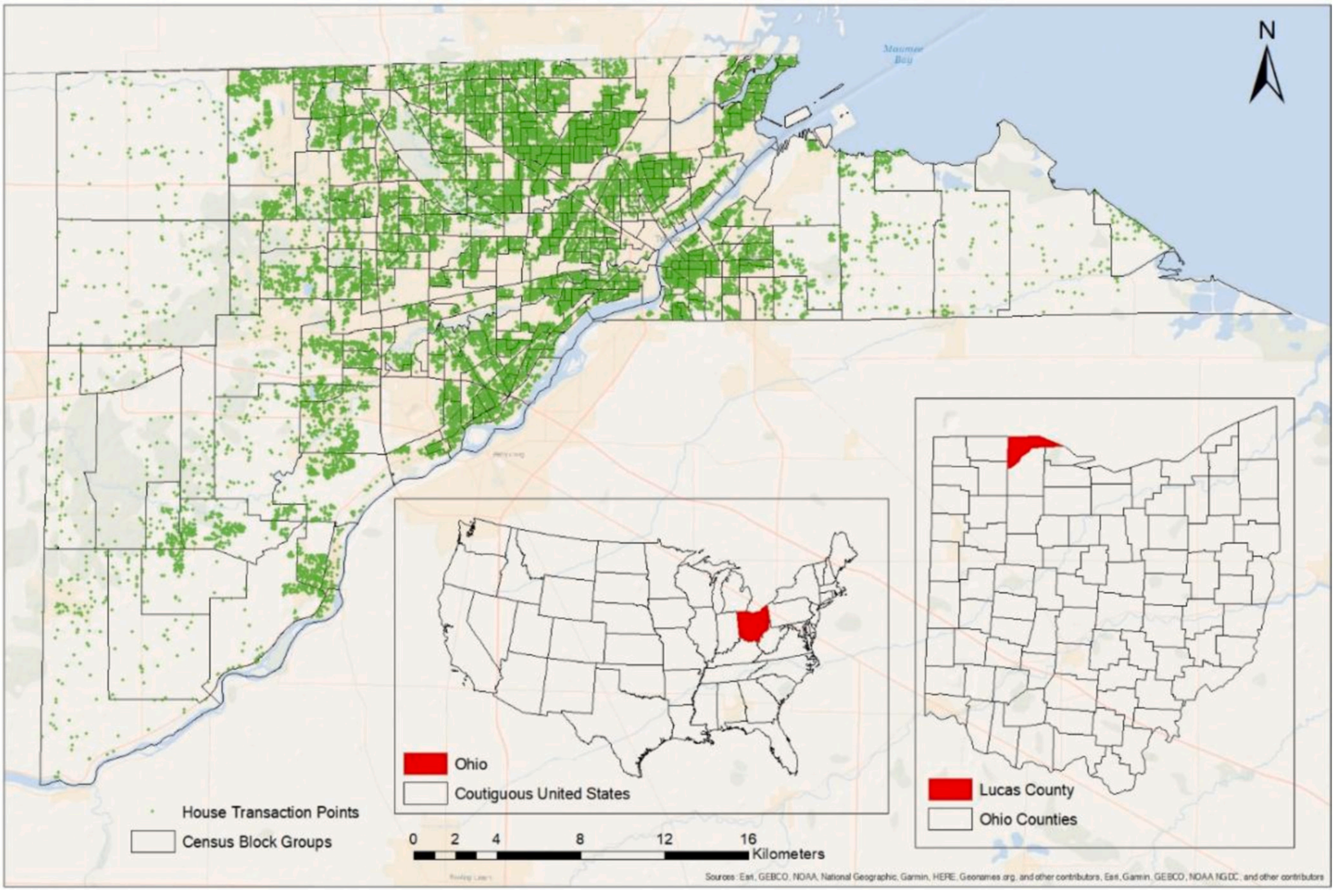

Fig. 2. Overview of the study area and house traction points.

2.4. Visible green space and calculations

To conduct a comparative analysis, we utilised two methods to calculate the independent variable of most significant concern for this study: visible green spaces.

2.5. Land use data and calculation of visible green space in multiple buffer areas

In the first approach, we used land use data from the National Land Cover Database (NLCD) for 2016 and 2019. As the housing price data spans from 2017 to 2019, to improve precision, the quantification of visible green spaces for houses in 2017 relied on NLCD data from 2016, while for the years 2018 and 2019, NLCD data from 2019 was employed. This dataset classifies land into 20 distinct categories, and within this classification, 13 of these categories are considered to potentially encompass visible green spaces (land category codes 41, 42, 43, 51, 52, 71, 72, 73, 74, 81, 82, 90 and 95). A detailed legend illustrating these land categories is accessible at the National Land Cover Database Class Legend and Description. Due to the complexity of ascertaining the precise influence of nearby visible green spaces on house price, this study established buffer areas centred on each house with radii of 76.2 m (250 ft), 106.7 m (350 ft), and 137.2 m (450 ft). The buffer radius was selected regarding the average length of the shortest side of a street block in some major cities throughout the United States, typically within the range of approximately 60.96 to 121.92 m (200 to 400 ft) (Baron, 2023). Then, the 13 aforementioned land categories were considered as a single land category and named the category of visible green space (CVG). Each house's visible green space area is quantified by calculating the proportion of pixels within the respective buffer area corresponding to the CVG category relative to the total number of pixels. This proportion is referred to as the NLCD Extracted Geen View Index (NGVI), in this study.

2.6. Street view images and semantic segmentation using deeplabv3+

We downloaded street view images in the second approach using Google Street View API. To guarantee the collected images are densely enough to cover the streetscape, we retrieved the street view imageries along the Topologically Integrated Geographic Encoding and Referencing (TIGER) Road Network with a fixed distance interval of 100 m (Li et al., 2022). For each location, four street view images are extracted from different angles (0°, 90°, 180°, 270°) to show the surrounding built environment comprehensively. Moreover, we generated a 50-m buffer catchment area for each property, referenced the work of Wu and other scholars (Wu et al., 2022), and only the street view images collected within the buffer were selected as the environment description corresponding to the house. The decision to generate a 50-m buffer instead of the previously used 76.2 m, 106.7 m, or 137.2 m buffers for measuring NGVI was attributed to our emphasis on visible green space rather than accessible and walkable green space. Subsequently, a pre-trained Deeplabv3+ model was utilised to conduct semantic segmentation on street view images acquired from Google. Before this analysis, 100 Google Street View images were employed to fine-tune the neural network (Brostow et al., 2009; Chen et al., 2018). These images were categorised into three distinct classes, namely "building," "vegetation," and "other." The details of the segmentation process are depicted in Fig. 3.

The classification accuracy of the Deeplabv3+ model in classifying vegetation achieved a remarkable accuracy of 0.963. Subsequently, we conducted Green View Index (GVI) calculations for each specific location in four distinct directions (0°, 90°, 180°, 270°), using the outcomes

Table 1
Descriptions of the variables.

Variable	Definition	Mean	Max	Min	Std.
Price	Residential property sales price in USD	103,044	282,000	6800	67,824
Structural variables					
TLA	Total land area	1549	7493	400	580.60
AGE	Age of a residential property, up to the year transacted	65.47	190	−4	31.66
LOTSIZE	Size of the piece of land	15,184	2,091,316	956	42,566
GARAGESQFT	Square feet of garage	397.07	5376	0	237.80
FULLBATH	Number of full bathrooms	1.37	6	1	0.56
HALFBATH	Number of half bathrooms	0.37	3	0	0.51
BEDRMS	Number of bedrooms	2.91	4	0	0.83
COND	Condition of the house	2.98	8	1	0.41
Neighbourhood socio-demographic variables					
MEDHHINC	Median household income	58,970	176,528	0	29,388
POVERTY	The percentage of families whose poverty status has been determined in the last 12 months	12.32	100	0	15.04
NOSCHOOL	The percentage of people who have no school complete	0.71	9.39	0	1.32
HIGHSCHOOL	The percentage of people who have a high school degree or equivalent	20.76	73.51	0.84	8.62
BACHELOR	The percentage of people who have bachelor's degree	11.77	37.13	0	7.78
WHITE	White percentage	75.93	100	0	22.87
ASIAN	Asian percentage	1.54	29.41	0	2.69
HISPANIC	The percentage of the Hispanic population	6.80	52.4	0	7.31
VACANT	The percentage of vacant property	9.92	55.23	0	10.20
HEALTHINSU	The percentage of people who are covered by health insurance	23.10	48.01	1.98	8.17
EMPLOYED	The percentage of people who are employed	47.77	80.02	13.33	9.58
Environmental variables					
DIST_GROCERY	The road network distance for each house to its nearest grocery store	1800	20,968	26.86	2146
DIST_UNIVERSITY	The road network distance in meters for each house to its nearest university	4535	29,063	63.23	3765
DIST_SCHOOL	The road network distance in meters for each house to its nearest school (primary, secondary, and high school)	1027	13,326	18.88	951.50
DIST_HOSPITAL	The road network distance in meters for each house to its nearest hospital	3952	22,335	55.70	2897
DIST_PARK	The road network distance in meters for each house to its nearest park	1232	9906	20.47	1000
WALKSCORE	Measures walkability on a scale from 0 to 100 based on walking routes to destinations.	36.16	87	0	20.52
STREETCONNECT	Number of intersections per square mile	112.30	371.53	1.23	69.17

of the segmentation process. We further derived the Average Green View Index (AGVI) by taking the mean of GVI values obtained from these four directions (Wu et al., 2022). The formulae for GVI and AGVI computation are provided as follows.

$$GVI_i = \frac{Number\ of\ vegtation\ pixel_i}{Number\ of\ pixel_i}$$
$$AGVI = \frac{\sum_{i=1}^{i=n} GVI_i}{n} \tag{1}$$

i is the i^{th} street view among the n street views for a given location. The number of street views per location is n instead of 4 because some directions may be missing street view data.

2.7. Random forest model

The RF model represents an ensemble approach comprising decision trees, which is valuable in regression analysis (Čeh et al., 2018). During

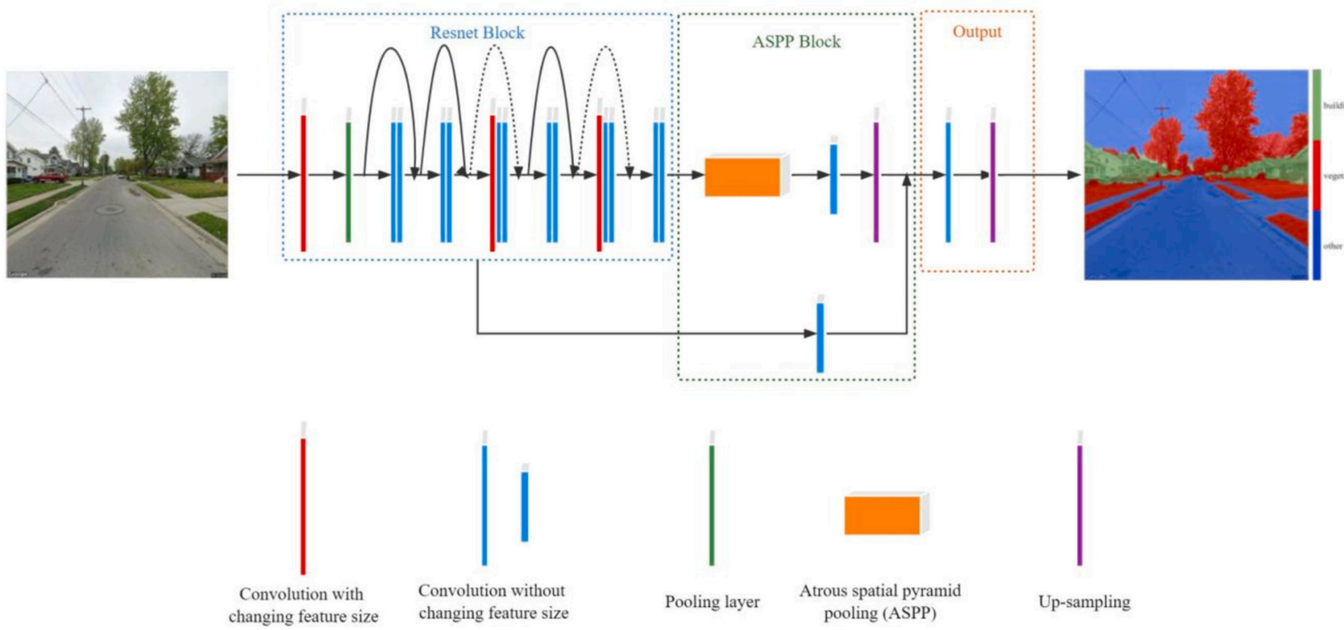

Fig. 3. Workflow of deeplabv3+ network for semantic segmentation.

the regression process, the RF model systematically draws random subsamples from the complete dataset and utilizes these subsamples to construct individual decision tree models. Subsequently, a comprehensive outcome is obtained by amalgamating the results of multiple decision tree models through the "bagging" method. A prominent advantage of the RF model is that it is independent of data distribution, eliminating the necessity for multicollinearity assessments. Furthermore, it effectively addresses the challenge of overfitting. It is worth highlighting that the RF model offers the capability to ascertain the importance of each independent variable concerning the dependent variable, enabling a comprehensive evaluation of the influence of independent variables on the dependent variable (Wu et al., 2022). When performing regression tasks, RF evaluates the importance of the independent variables by calculating the reduction in mean squared error (MSE) at each node split (Meinshausen & Ridgeway, 2006). Specifically, during the construction of each decision tree, the reduction in MSE for each independent variable is calculated, accumulated, and then averaged, resulting in the "importance score" of that variable (Friedman et al., 2009; Scornet et al., 2015). In this study, we calculated the ratio of the importance score of each variable to the total "importance score" of all variables, which is referred to as "contribution" in this study.

To analyse the differences in contributions when utilising NGVI to represent visible green space in contrast to utilising AGVI, we first used those 26 independent variables above, with NGVI as the 27th independent variable. Subsequently, we constructed a RF model with house prices as the dependent variable and calculated the contribution. Following this, we replaced NGVI with AGVI, reconstituted the RF model, and calculated the contribution again. Recognising the potential temporal changes in the relationship between visible green space and housing prices, we conducted separate modelling on the house price data for 2017, 2018, and 2019. The dataset was partitioned into training and testing subsets, with a 90 % and 10 % division. Model performance was assessed through ten-fold cross-validation to ensure rigorous accuracy assessment. The evaluation indicators of the model are coefficient of determination (R^2), mean absolute error (MAE), and root mean square deviation (RMSE) (Čeh et al., 2018; Hjort et al., 2024; Wu et al., 2022). It is essential to note that although RF as a non-parametric model shows robustness to multicollinearity in predictive performance, the mechanism during prediction differs from that used to compute importance scores. Dormann et al. (2013) found that multicollinearity may affect the RF model's calculation of the importance scores of the independent variable (Dormann et al., 2013). Therefore, before modelling, we verified the multicollinearity of all independent variables using the Variance Inflation Factor (VIF). The results showed that the maximum value of all VIFs did not exceed 5, indicating that there is theoretically no severe multicollinearity (Čeh et al., 2018; Zuur et al., 2010).

3. Results

3.1. The influence of two ways of quantifying green space on contributions

Based on the results of the RF model constructed using 2017 house price data, when the buffer radius is specified at 76.2 m, the influence of NGVI on housing prices is notably limited, standing at a mere 0.75 %. Following a comprehensive investigation of all 27 independent variables, NGVI is positioned at the 27th rank concerning its contribution. Nevertheless, as the buffer radius is extended to 106.7 m, there is a marginal increment in the contribution of NGVI, reaching 0.79 %. Notably, within this context, the rank of NGVI in terms of its contribution remains unchanged. Furthermore, as the buffer radius is extended to 137.2 m, the contribution of NGVI on housing prices experiences an upturn, escalating to 0.92 %, and concurrently elevating the rank of NGVI to the 25th position (Fig. 4). When utilising NGVI as an indicator for quantifying visible green spaces, the R^2 values of the RF model consistently surpass 0.7, irrespective of the length of the buffer radius. Concurrently, the MAE is confined to 23,586 to 23,604, while the RMSE spans 32,300 to 32,349 (Table 2). These data indicate a high level of model accuracy, and the analysis above related to the contribution of NGVI is reliable. It is noteworthy that this study includes a total of 27 independent variables, with NGVI ranked between 25th and 27th in terms of importance. Based on the 2017 modelling results for Lucas County, the overall impact of green space on housing prices in this specific region is relatively small.

Fig. 4 illustrates the contribution of independent variables when the RF model is constructed using 2017 house price data (A) is the value and ranking of the contribution of each independent variable when the NGVI is used to quantify the visible green space and the buffer radius is 76.2 m (B) is the value and ranking of the contribution of each independent

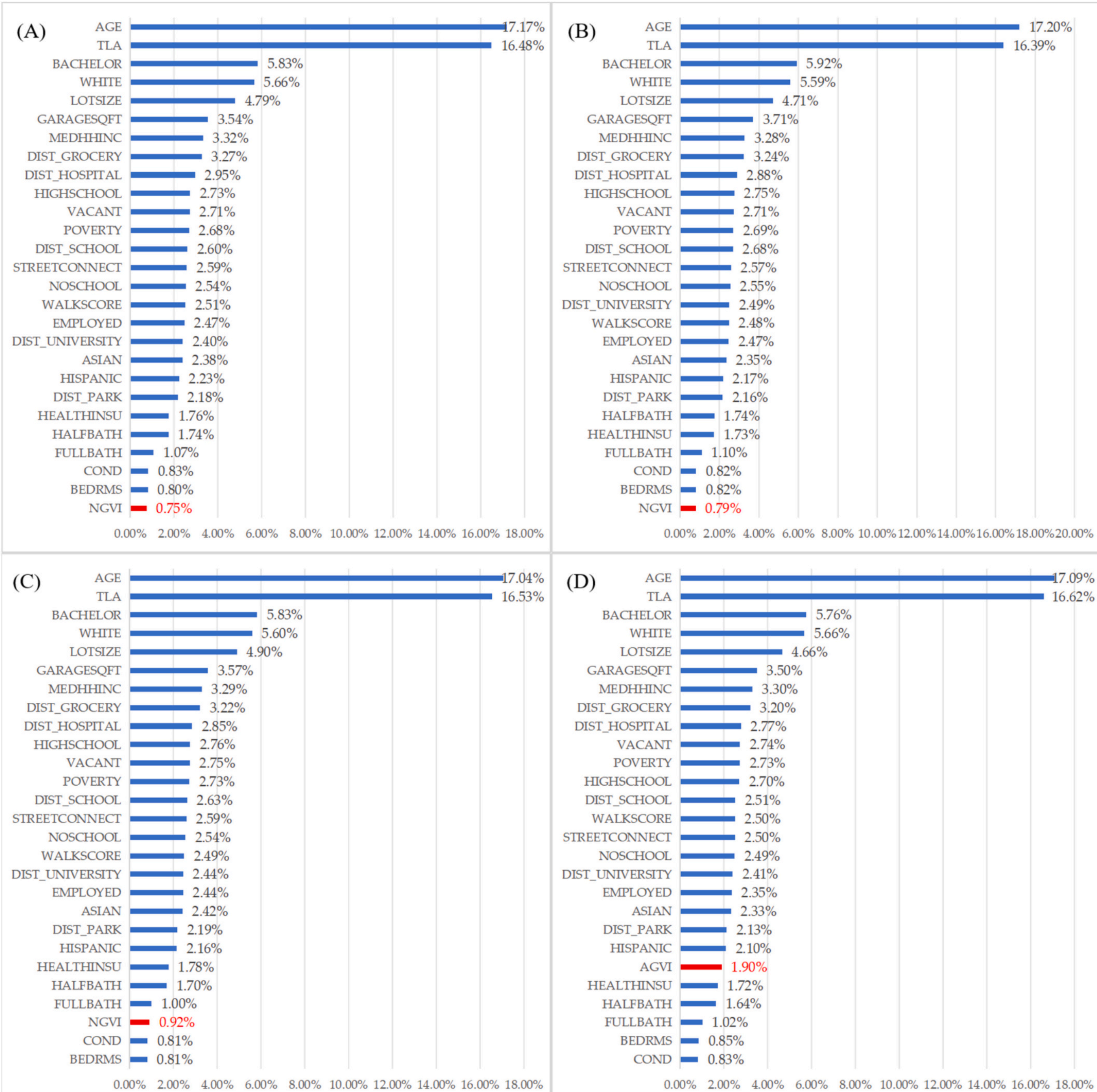

Fig. 4. Value and ranking of the contribution of each independent variable, 2017.

Table 2
The R^2, MAE and RMSE values of RF models.

Types	2017			2018			2019		
	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE
NGVI (76.2 m)	0.746	23,604	32,300	0.751	23,940	32,477	0.726	26,282	38,123
NGVI (106.7 m)	0.747	23,596	32,332	0.748	24,005	32,677	0.725	26,359	38,162
NGVI (137.2 m)	0.746	23,586	32,349	0.748	23,990	32,653	0.725	26,359	38,144
AGVI	0.747	23,594	32,312	0.748	23,999	32,650	0.726	26,288	38,171

variable when the NGVI is used to quantify the visible green space and the buffer radius is 106.7 m (C) is the value and ranking of the contribution of each independent variable when the NGVI is used to quantify the visible green space and the buffer radius is 137.2 m (D) is the value and ranking of the contribution of each independent variable when the AGVI is used to quantify the visible green space.

When using the AGVI to quantify visible green spaces, the results obtained from the RF model constructed with housing price data from 2017 indicate notable differences compared to those obtained when using the NGVI to quantify visible green spaces. Specifically, the contribution of AGVI stands at 1.90 %, positioning it at the 22nd rank among all 27 variables. The contribution value and ranking exceed NGVI's (Fig. 4). This observation implies that using land-use data for quantifying visible green spaces underestimates their contribution to housing prices compared to the quantification using street view images. Notably, when AGVI is utilised for quantifying visible green spaces, the R^2 value of the RF model is 0.747, accompanied by an MAE of 23,594 and an RMSE of 32,312. Collectively, these indicators' values indicate a high overall model accuracy (Table 2). Although the importance ranking of green space increased when quantified using AGVI, this change did not affect the conclusion that the overall impact of green space on housing prices in Lucas County remains relatively small.

Nonetheless, it is essential to acknowledge that the results obtained from the RF model established using housing price data from 2017 may not fully consider the potential temporal variations in the relationship between housing prices and their influencing factors. As a result, an investigation was conducted into the results when the house price data imported the RF model is from 2018. The findings indicate that when the

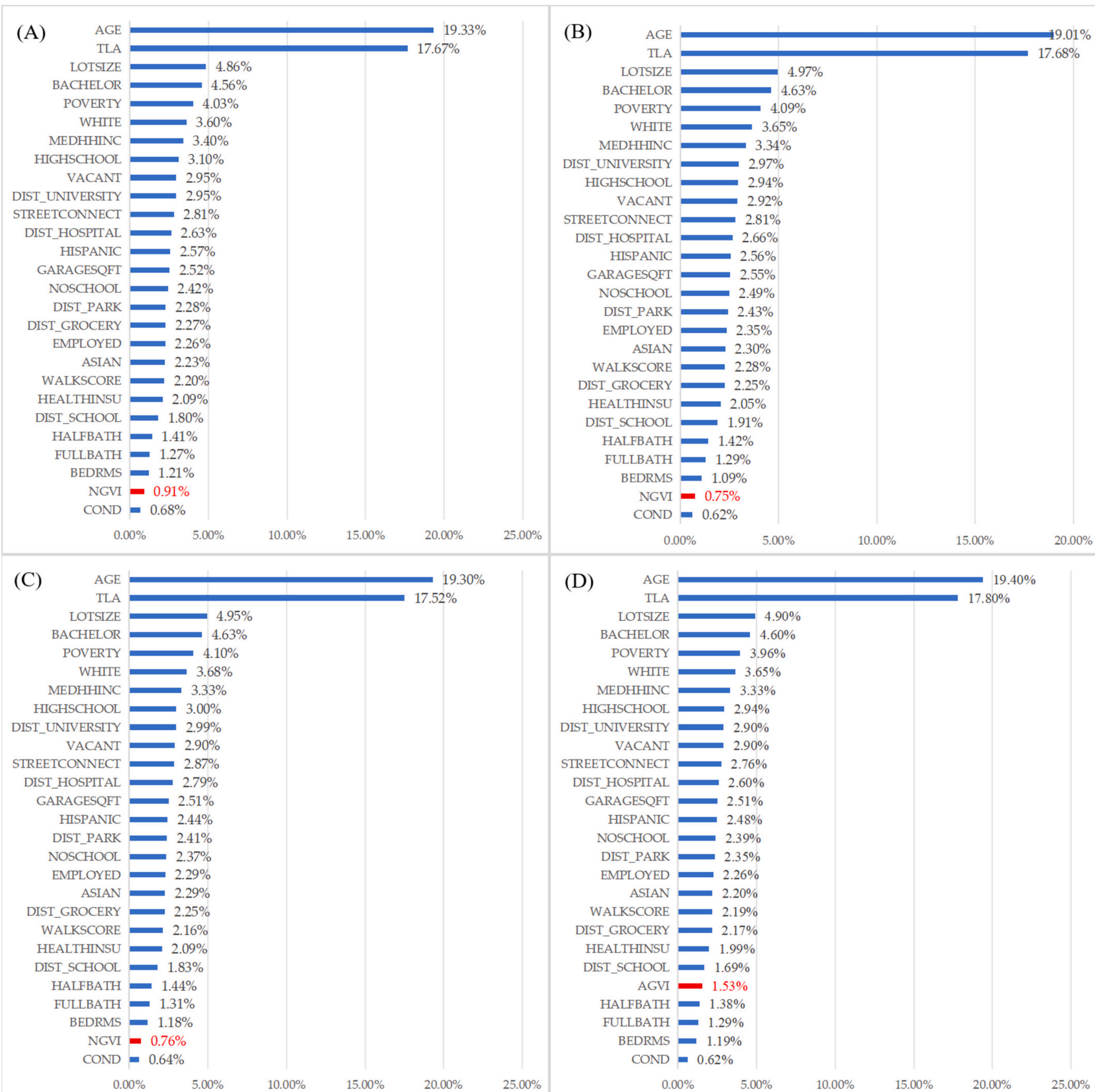

Fig. 5. Value and ranking of the contribution of each independent variable, 2018.

radius of the buffer area is set as 76.2 m, the contribution of NGVI stands at 0.91 %, positioning it as the 26th variable among all 27 independent variables. However, as the buffer radius is expanded to 106.7 m, the contribution of NGVI decreases to 0.75 %, but the rank of NGVI in terms of its contribution remains unchanged. The rank of NGVI still maintains relative stability when the buffer radius is extended to 137.2 m, where the contribution of NGVI experiences a minor increase to 0.76 %. Furthermore, when quantifying visible green spaces using street view images, the contribution value and ranking of AGVI also indicate an increase relative to those of NGVI. Specifically, the contribution value of AGVI reaches 1.53 %, sitting at the 23rd rank (Fig. 5). Consequently, it can be affirmed that using NGVI for quantifying visible green spaces tends to underestimate their influence on housing prices compared to AGVI. Importantly, this underestimation issue excludes RF models constructed solely with 2017 housing price data. The R^2 of all models is higher than 0.7, and the overall accuracy does not change much from the RF model constructed using the 2017 house price data, which indicates that the above analysis is credible (Table 2). From the perspective of the importance of green space in relation to housing prices, although its ranking fluctuated compared to the 2017 results, the highest ranking was 23rd. This indicates that in Lucas County in 2017 and 2018, green space did not demonstrate particularly high importance in influencing housing prices.

Fig. 5 shows the ranking of the contribution of each independent variable when the RF model is constructed using 2018 house price data (A) is the value and ranking of the contribution of each independent variable when the NGVI is used to quantify the visible green space, and the buffer radius is 76.2 m (B) is the value and ranking of the

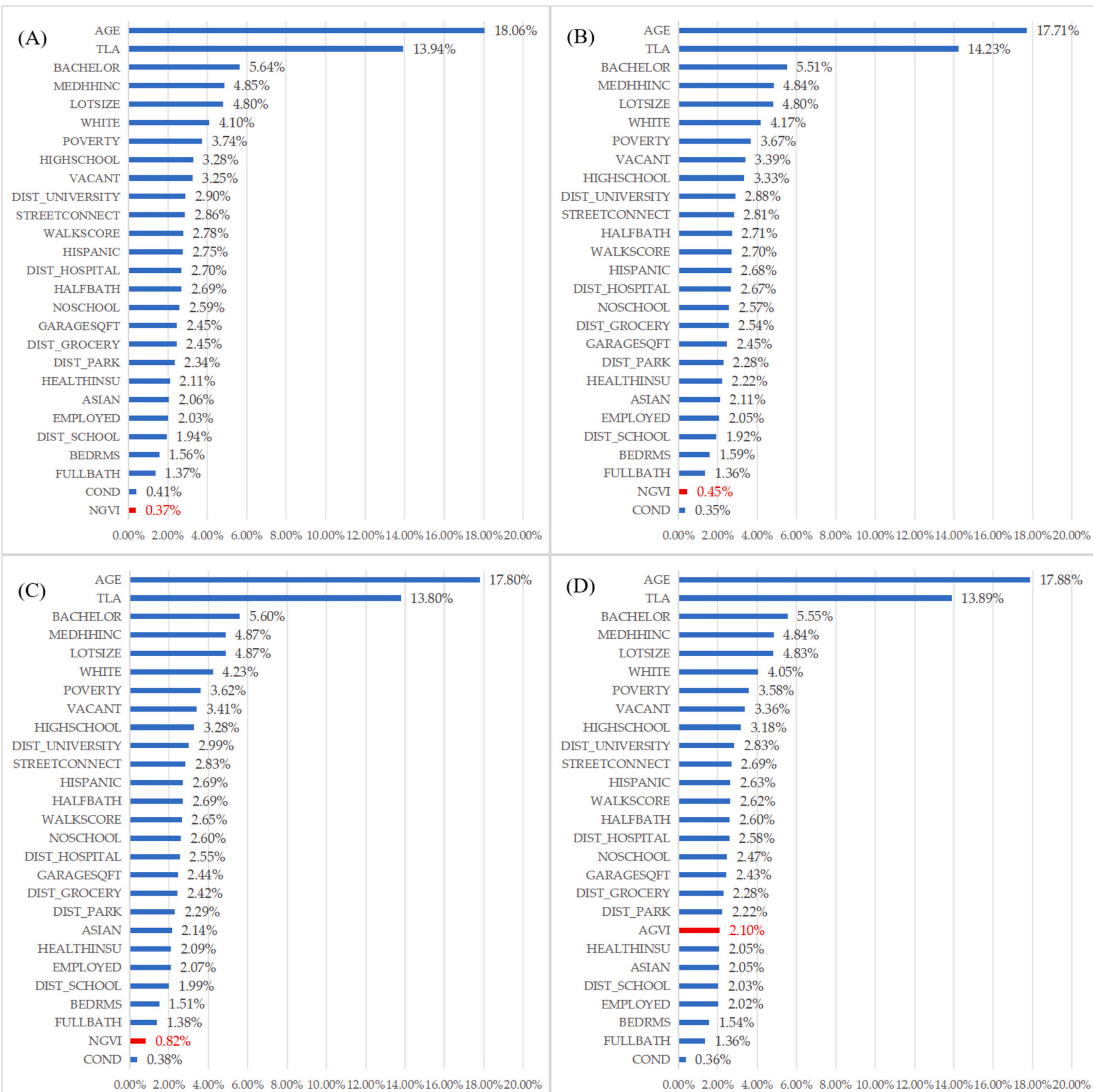

Fig. 6. Value and ranking of the contribution of each independent variable, 2018.

contribution of each independent variable when the NGVI is used to quantify the visible green space and the buffer radius is 106.7 m (C) is the value and ranking of the contribution of each independent variable when the NGVI is used to quantify the visible green space and the buffer radius is 137.2 m (D) is the value and ranking of the contribution of each independent variable when the AGVI is used to quantify the visible green space.

When constructing the RF model using housing price data from 2019, the previously discussed issue of underestimation becomes more conspicuous. To expound further, the highest contribution of NGVI is 0.82 %, positioning it at the 26th rank among the 27 independent variables. In contrast, the AGVI makes a noteworthy contribution of 2.10 %, securing the 20th position (Fig. 6). The difference between the contribution values of AGVI and the highest contribution values of NGVI is most pronounced during this specific temporal context. The same pattern is observed in the contrast between AGVI's contribution ranking and NGVI's highest contribution rankings. This accentuates the potential influence of quantification methodologies on the values and ranking of visible green space contributions and signifies that this influence is not constant, demonstrating temporal heterogeneity. This observation is supported by a comprehensive analysis of all the results above. Hence, the practice of annual modelling of housing price data in this study is reasonable. Furthermore, the overall importance of green space in housing prices did not undergo obvious changes between 2017 and 2019. Therefore, the impact of green space on housing prices in Lucas County was not substantial during these three years. Finally, It is crucial to underscore that the RF model still demonstrates a notably high overall accuracy, thereby ensuring the reliability of the findings derived in this study (Table 2).

Fig. 6 demostrates the contribution of each independent variable when the RF model is constructed using 2019 house price data (A) is the value and ranking of the contribution of each independent variable when the NGVI is used to quantify the visible green space, and the buffer radius is 76.2 m (B) is the value and ranking of the contribution of each independent variable when the NGVI is used to quantify the visible green space, and the buffer radius is 106.7 m (C) is the value and ranking of the contribution of each independent variable when the NGVI is used to quantify the visible green space and the buffer radius is 137.2 m (D) is the value and ranking of the contribution of each independent variable when the AGVI is used to quantify the visible green space.

3.2. *Characterisation of the contribution of visible green space to house prices*

This academic research conducted a year-by-year modelling of housing prices in Lucas County from 2017 to 2019. In this context, it is evident that the contribution values and rankings of various variables would naturally change over this period. These changes pose a challenge when attempting to assess the relative importance of each independent variable. Therefore, this study counted the number of times the respective variables ranked in the top 10 % to the top 50 % of all independent variables in the RF models that quantified visible green space using AGVI. The importance of variables that repeatedly rank in the top 10 % differs from those that repeatedly rank in the top 20 % but do not make it to the top 10 %. Hence, we adopted the Analytic Hierarchy Process (AHP), a decision-making framework that aids stakeholders in reaching consensus and determines the optimal course of action in situations where decision-making is complex and not easily quantifiable (Saaty, 1980). Through the AHP method, it is possible to ascertain the importance of each scale relative to the overall objective.

This study aims to "evaluate the relative importance of each independent variable", with the criteria being the "top 10% (top 3), 20% (top 5), 30% (top 8), 40% (top 11), and 50% (top 14)", these five criteria. The scaling rules were developed based on these five criteria. The criterion of being in the top 10 % is the most critical, followed by being in the top 20 %, and so forth. Accordingly, a score of 5 was assigned to the top 10 %, 4 to the top 20 %, 3 to the top 30 %, 2 to the top 40 %, and 1 to the top 50 %. Utilising these scaled data, SPSS was employed to construct the judgment matrix and to generate the weight scores. The weight scores obtained in this study are as follows, an independent variable received 0.41742 points each time it achieved a top 10 % ranking, 0.26337 points for a top 20 % ranking, 0.16023 points for a top 30 % ranking, 0.09748 points for a top 40 % ranking, and 0.06150 points for a top 50 % ranking. The Consistency Ratio (CR) is used to test whether the obtained weight scores contain logical errors. The test is passed when CR < 0.1. If the test is not passed, the scales need to be revised. In this study, all weight scores have passed the test (Saaty, 1980). If an independent variable's contribution ranking is in the top 10 %, it will also be recorded in both the "10 %", "20 %", "30 %", "40 %" and "50 %" fields in this study. Finally, all independent variables were reordered based on the results of the AHP analysis. The reordered results show the relative importance of each independent variable (Table 3).

The findings suggest that the primary determinants of house prices continue to be the intrinsic characteristics of the properties themselves and the socioeconomic attributes of the neighbourhoods in which they are situated (Table 3). The influence of most components of the built environment, including visible green spaces as measured by NGVI or AGVI, on house prices is relatively modest. The evidence in this study shows that the highest contribution of NGVI or AGVI to house prices is limited to 2.1 % in Lucas County (Fig. 6). Furthermore, AHP analysis results indicate that NGVI and AGVI garner weighting values 0. This signifies that, despite potential underestimation in contribution

Table 3
The results of the AHP analysis of each variable.

Independent variables	10 %	20 %	30 %	40 %	50 %	AHP	Types
TLA	3	3	3	3	3	3.00	house
AGE	3	3	3	3	3	3.00	house
BACHELOR	2	3	3	3	3	2.58	socioeconomic
LOTSIZE	1	3	3	3	3	2.17	house
MEDHHINC	0	1	3	3	3	1.22	socioeconomic
WHITE	0	1	3	3	3	1.22	socioeconomic
POVERTY	0	1	2	3	3	1.06	socioeconomic
VACANT	0	0	1	3	3	0.64	socioeconomic
HIGHSCHOOL	0	0	1	2	3	0.54	socioeconomic
GARAGESQFT	0	0	1	1	2	0.38	house
DIST_GROCERY	0	0	1	1	1	0.32	built environment
DIST_UNIVERSITY	0	0	0	2	2	0.32	built environment
STREETCONNECT	0	0	0	2	2	0.32	built environment
DIST_HOSPITAL	0	0	0	1	2	0.22	built environment
HISPANIC	0	0	0	0	2	0.12	socioeconomic
WALKSCORE	0	0	0	0	2	0.12	built environment
HALFBATH	0	0	0	0	1	0.06	house
DIST_SCHOOL	0	0	0	0	1	0.06	built environment
COND	0	0	0	0	0	0.00	house
BEDRMS	0	0	0	0	0	0.00	house
FULLBATH	0	0	0	0	0	0.00	house
HEALTHINSU	0	0	0	0	0	0.00	socioeconomic
AGVI	0	0	0	0	0	0.00	built environment
NGVI[a]	0	0	0	0	0	0.00	built environment
DIST_PARK	0	0	0	0	0	0.00	built environment
ASIAN	0	0	0	0	0	0.00	socioeconomic
EMPLOYED	0	0	0	0	0	0.00	socioeconomic
NOSCHOOL	0	0	0	0	0	0.00	socioeconomic

[a] Table 3 presents a secondary analysis of the results obtained when quantifying green spaces using AGVI. However, it has already been concluded that NGVI is underestimated compared to AGVI, as AGVI's AHP score is 0, meaning all NGVI scores must also be 0. Both are displayed here for ease of analysis.

associated with quantifying visible green spaces using NGVI, it does not emerge as a significant factor in scrutinising the primary determinants of housing prices in Lucas County. Put differently, even in the presence of measurement errors related to visible green spaces, the fundamental evaluation of housing prices remains hardly affected in Lucas County. This finding diverges from existing studies and can be attributed to the spatial heterogeneity in the influence of visible green spaces on housing prices (Goodchild, 2003; Su et al., 2021; Wu et al., 2022). Simultaneously, when utilising housing price data from the same area, the degree of underestimation in the contribution to visible green spaces due to the application of NGVI indicates temporal inconsistency annually.

4. Discussion

This research contributes to the body of knowledge on environmental economics, particularly the concept of hedonic pricing, by highlighting the methodological nuances in valuing green spaces within urban housing markets. Hedonic price models posit that the value of a good (in this case, a house) can be decomposed into its individual attributes, including environmental characteristics like green space (Rosen, 1974). These attributes contribute to the overall utility derived from the good, and individuals are willing to pay a premium for properties with desirable attributes. Our research aligns with this framework by investigating how access to green space influences housing prices. We hypothesize that the presence and quality of green spaces surrounding a property enhance its utility for potential buyers, leading to higher market values.

The primary contribution of this study lies in its focus on non-metropolitan regions, offering insights into the valuation of green spaces in areas that differ significantly from the wealthier, high-density metropolitan regions typically studied. While prior studies, particularly those focused on metropolitan areas, have shown a positive correlation between larger green areas and property values (Nicholls & Crompton, 2005), we suggest that while green spaces are valued additions to neighbourhoods, their economic influence on housing prices is less significant than the core attributes of the properties and the socioeconomic context. This contradicts the traditional emphasis on the amenity value of green spaces in real estate assessments (Gibbons et al., 2014; Kong et al., 2007). Lucas county, though not a major metropolis, serves as a unique case study as it represents a growing segment of the American population residing in non-metropolitan areas. We argue the specific characteristics of a place can influence how green spaces impact property values. The underlying mechanisms through which green spaces affect housing prices may vary depending on factors like community needs, local demographics, and the overall availability of green space within a particular city. Our findings in Lucas county demonstrate that while green spaces are valued amenities, their influence on housing prices might be less pronounced compared to core property attributes and the socioeconomic context. This observed difference from previous research underscores the importance of considering location-specific factors when evaluating the economic impact of green spaces. Moreover, temporal heterogeneity is found in the impacts of visible green space on housing prices, proving inaccuracies within a localised spatial and relatively brief temporal context, necessitating a large-scale and long-term study. This study expands the discussion on spatial heterogeneity by examining how green spaces influence housing prices in areas with lower levels of urbanization and more diverse economic structures. Last but not least, we have to acknowledge our study utilizes the hedonic pricing framework to explore the correlations, not causal relationships. This approach aligns with the inherent limitations of hedonic models, which focus on associations rather than causation. Our choice of RF reflects this focus. While RFs don't definitively prove causality, they excel at uncovering correlations, which is precisely our objective. We acknowledge that future research interested in causal effects could benefit from a more intricate quasi-experimental design.

While this study finds that the impact of green spaces on housing prices in Lucas County is smaller than in previous research, this difference can be attributed to the distinct characteristics of the area. Many prior studies have focused on metropolitan regions, which often feature higher-income populations and more concentrated urban development. These metropolitan areas, though valuable for research, may not always represent the typical living environments of most counties. In contrast, Lucas County is more representative of an ordinary, non-metropolitan area where factors such as housing, safety, and transportation may take precedence for residents. According to Maslow's hierarchy of needs, individuals tend to prioritize basic necessities before focusing on higher-order needs like aesthetics and well-being, which green spaces primarily enhance (Lu, 2019; Maslow, 1971; Ulrich, 1984). As a result, green spaces may have a smaller influence on housing prices in areas like Lucas County. The stronger effect observed in previous studies may reflect the unique characteristics of wealthier, urbanized regions, which are not necessarily typical of the broader population. Theoretically, our findings challenge the conventional wisdom that green space always has a very essential impact on property values, suggesting instead that regional and socioeconomic factors heavily mediate this relationship.

From a data harmonisation standpoint, conducting environmental audits using diverse data sources and measuring objects and subjects presents various advantages and disadvantages. Therefore, it's essential to consider the multifaceted aspects of this approach thoroughly. Dennis and James (2016) emphasised the need to account for user participation and social-ecological contexts in evaluating urban green spaces. Daniels et al. (2018) supported this by proposing a multidimensional perspective that integrates ecological, microclimatic, and social aspects in assessing green space structures. Seaman et al. (2010) further underscored the importance of subjective experiences, such as feelings of integration and inclusion, in influencing the use of urban green spaces. Baycan-Levent et al. (2009) provided a practical application of this approach, using a multi-criteria evaluation to compare the “green performance” of European cities based on indicators of green space availability, changes, planning, financing, and performance. Inspired by prior insight, this study introduces a comparative dimension, measuring the impact of green space visibility from both macro (NGVI) and micro (AGVI) perspectives. Our results advocate for a balanced approach that considers objective and subjective assessments of green spaces, as each method captures different facets of urban greenness. For instance, while NGVI offers a broader categorisation of land use, AGVI provides a street-level view that reflects the immediate visual environment of residents. This dual approach could help reconcile discrepancies between large-scale planning and individual-level urban experience. Although the integration of land use data and street-view data has been explored in previous research, our approach is unique in its application to non-metropolitan areas, a context that remains under-explored. Additionally, by using both NGVI and AGVI, we provide a multi-scale analysis that captures both broad categorizations of green space as well as street-level views of green space visibility. This dual approach allows for a more nuanced understanding of how green spaces impact housing prices at different levels of urban experience.

In terms of methodology, this study presents an approach that integrates a non-linear model with the AHP. This combination not only enhances the accuracy of the model and the credibility of the analytical results but also effectively accommodates temporal heterogeneity. By reflecting model results across different periods, it provides a comprehensive assessment of the importance of independent variables. Many studies have utilised linear models to explore the relationship between environmental factors, economic indicators, and real estate prices. Cellmer et al. (2012) found that greenery, surface water, noise impacts, and landscape features significantly influenced property prices. Grum and Govekar (2016) identified a significant correlation between real estate prices and macroeconomic factors such as unemployment, current account, GDP, and industrial production. Din et al. (2001) compared different real estate valuation models, including linear regression models, and found that they produced similar price indices. Chiarazzo et al. (2014) and Chiarazzo et al. (2014) discovered that hedonic multiple linear regression models, accounting for spatial dependence, provided a better fit for examining the influence of environmental factors and accessibility on real estate prices, highlighting the importance of managing non-linear relationships and interactions among variables in market analysis. In summary, existing research has revealed the non-linear characteristics of the relationship between environmental factors, economic indicators, and real estate prices, indicating that linear or generalized linear models may struggle to establish such relationships effectively, thereby impacting predictive accuracy. Conducting interpretability analysis and evaluating models with insufficient accuracy can often lead to questionable credibility (Aiken, 1991; Lipton, 2018). RF, with its complex operational mechanisms, has been widely recognised for effectively capturing non-linear relationships between independent and dependent variables, thus achieving high simulation accuracy. Therefore, we opted to employ RF, a non-linear model, to ensure the high credibility of our analysis results. Furthermore, existing research has examined the challenges of heterogeneity across different time periods in interpreting results from time-sliced regression experiments from multiple perspectives (Cameron, 2005; Cattaneo et al., 2016). By utilising AHP to transform temporal heterogeneity into a scoring problem based on criteria layers, we integrate time-varying data into a cohesive result, rendering the conclusions more intuitive. This methodological advancement not only enhances interpretability but also reinforces the robustness of our findings.

One limitation of the current study is the reliance on NGVI and AGVI as the primary indicators of green space, which may not fully capture the multifaceted nature of urban greenery and its nuanced effects on housing prices. For instance, the quality, maintenance, and vegetation types are not differentiated from these indices, potentially leading to an undervaluation of green spaces that provide more fantastic ecosystem services or recreational opportunities. Additionally, the temporal scope of our study does not allow for the observation of long-term trends in housing prices influenced by changing urban policies or climate adaptation strategies related to green spaces. Finally, our NGVI data corresponds to annual NLCD data, while street view data does not. Theoretically, street view data should also be annual for more accurate comparisons. However, Google Street View does not provide an interface for obtaining annual street view data; therefore, only the most recently updated data is accessible, which impacts the assessment results. Further more, while this study employed a quantitative approach to assess the influence of green space on housing prices, a limitation is the exclusion of residents' subjective perceptions of green space value. Future research could benefit from incorporating a mixed-methods approach, combining quantitative analysis with qualitative data collection through in-depth interviews or focus groups. This would provide a richer understanding of how residents perceive green spaces and how these perceptions translate into housing preferences. Moreover, our study employed a case study area in a non-metropolitan region. This approach may contribute to the apparent contradiction between our findings and previous research. We suspect this difference stems from the spatial heterogeneity of the underlying mechanisms by which green spaces influence housing prices. However, the limitation of our study is also the focus on a single case study area. A multi-city study with diverse urban profiles could help validate our assumptions regarding the influence of location on green space valuation. Finally, our analysis only considered green spaces within a buffer zone surrounding each household. Future research could explore the impact of larger green spaces, such as parks and recreational areas, that may enhance community vibrancy and desirability, potentially influencing housing prices. These attributes could be valuable considerations for future studies.

Future research should aim to incorporate a broader set of green space characteristics, including biodiversity, park amenities, and accessibility, to provide a more comprehensive valuation. Longitudinal studies could elucidate the evolving influence of green spaces on housing prices over time, particularly in response to urban development and environmental policy changes. Moreover, integrating public perception surveys and qualitative assessments could enrich the understanding of how residents value green spaces beyond their visual presence. Further exploration into the intersectionality of green space benefits with socioeconomic and demographic factors would also offer a more detailed landscape of urban green space valuation. More importantly, expanding the geographical scope of such studies could validate the model's applicability across different urban settings and cultural contexts, potentially leading to universally applicable urban planning guidelines. Lastly, In recent years, explainable machine learning methods such as SHAP (SHapley Additive exPlanations) have been demonstrated to reveal non-linear relationships more effectively, particularly in uncovering local effects and variable interactions (Lundberg, 2017). However, this study emphasizes the assessment of broader patterns and overall impacts of different green space measurements, and the calculation of variable contributions within RF suffices to meet this research objective; thus, further analysis using SHAP was not conducted. Nevertheless, SHAP could prove valuable in future studies that require a more granular understanding of variable contributions, especially in contexts similar to this one.

5. Conclusion

The empirical evaluation of the influence of green spaces on housing prices using both NGVI and AGVI provides a nuanced understanding of their complex economic impact. While green spaces contribute positively to housing values, the extent of this effect varies depending on factors like property characteristics, socioeconomic conditions, and even the specific type or quality of green space. This highlights the need for future research to explore these nuances in greater detail.

Furthermore, this study underscores the importance of incorporating both objective land-use data (like NGVI) and subjective visual assessments (like AGVI) in urban planning and policy-making. This multifaceted approach allows for a more comprehensive evaluation of environmental attributes within housing markets.

By adopting this balanced approach, urban development initiatives can be guided by a deeper understanding of how green spaces contribute to neighbourhood livability and desirability. This can lead to informed decisions that promote sustainable and equitable urban environments, without solely focusing on the immediate economic benefits of green spaces.

CRediT authorship contribution statement

Yanqing Xu: Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Ruidun Chen:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Hongyu Du:** Writing – review & editing, Visualization, Software, Methodology. **Meixu Chen:** Writing – review & editing. **Cong Fu:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation. **Yuchen Li:** Writing – review & editing, Writing – original draft, Validation, Supervision.

Funding

This study was supported by National Natural Science Foundation of China (42471444).

Data availability

Data will be made available on request.

References

Aiken, L. S. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.

Baron, c. (2023). How many acres is a city block? Discover with calculator 8/12/2023. https://www.measuringknowhow.com/how-many-acres-is-a-city-block-discover-with-calculator/#google_vignette.

Baycan-Levent, T., Vreeker, R., & Nijkamp, P. (2009). A multi-criteria evaluation of green spaces in European cities. *European Urban and Regional Studies, 16*(2), 193–213.

Biljecki, F., & Ito, K. (2021). Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning, 215*.

Bockarjova, M., Botzen, W. J. W., Van Schie, M. H., & Koetse, M. J. (2020). Property price effects of green interventions in cities: A meta-analysis and implications for gentrification. *Environmental Science & Policy, 112*, 293–304.

Brostow, G. J., Fauqueur, J., & Cipolla, R. (2009). Semantic object classes in video: A high-definition ground truth database. *Pattern Recognition Letters, 30*(2), 88–97.

Cameron, A. C. (2005). *Microeconometrics: Methods and applications*. Cambridge University.

Cattaneo, M. D., Titiunik, R., Vazquez-Bare, G., & Keele, L. (2016). Interpreting regression discontinuity designs with multiple cutoffs. *The Journal of Politics, 78*(4), 1229–1248.

Čeh, M., Kilibarda, M., Lisec, A., & Bajat, B. (2018). Estimating the performance of random forest versus multiple regression for predicting prices of the apartments. *ISPRS International Journal of Geo-Information, 7*(5).

Cellmer, R., Senetra, A., & Szczepanska, A. (2012). The effect of environmental factors on real estate value. In *, vol. 20012. FIG working week* (pp. 1–13).

Chen, L., Yao, X., Liu, Y., Zhu, Y., Chen, W., Zhao, X., & Chi, T. (2020). Measuring impacts of urban environmental elements on housing prices based on multisource data—A case study of Shanghai, China. *ISPRS International Journal of Geo-Information, 9*(2), 106.

Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer vision – ECCV 2018* (p. 11211). Cham: Springer.

Chiarazzo, V., dell'Olio, L., Ibeas, Á., & Ottomanelli, M. (2014). Modeling the effects of environmental impacts and accessibility on real estate prices in industrial cities. *Procedia-Social and Behavioral Sciences, 111*, 460–469.

Cho, S. H., Poudyal, N. C., & Roberts, R. K. (2008). Spatial analysis of the amenity value of green open space. *Ecological Economics, 66*(2–3), 403–416.

Cohen, V., & Karpavičiūtė, L. (2017). The analysis of the determinants of housing prices. *Independent journal of management & production, 8*(1), 049–063.

Daniels, B., Zaunbrecher, B. S., Paas, B., Ottermanns, R., Ziefle, M., & Roß-Nickoll, M. (2018). Assessment of urban green space structures and their quality from a multidimensional perspective. *Science of the Total Environment, 615*, 1364–1378.

Dennis, M., & James, P. (2016). Considerations in the valuation of urban green space: Accounting for user participation. *Ecosystem Services, 21*, 120–129.

Din, A., Hoesli, M., & Bender, A. (2001). Environmental variables and real estate prices. *Urban Studies, 38*(11), 1989–2000.

Dormann, C. F., et al. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography, 36.1*, 27–46.

Fan, G. Z., Ong, S. E., & Koh, H. C. (2006). Determinants of house price: A decision tree approach. *Urban Studies, 43*(12), 2301–2315.

Friedman, J., Hastie, T., & Tibshirani, R. (2009). *The elements of statistical learning: Data mining, inference, and prediction*.

Galster, G., & Rothenberg, J. (1991). Filtering in urban housing: A graphical analysis of a quality-segmented market. *Journal of Planning Education and Research, 11*(1), 37–50.

Geng, B., Bao, H., & Liang, Y. (2015). A study of the effect of a high-speed rail station on spatial variations in housing price based on the hedonic model. *Habitat International, 49*, 333–339.

Gibbons, S., & Machin, S. (2008). Valuing school quality, better transport, and lower crime: Evidence from house prices. *Oxford Review of Economic Policy, 24*(1), 99–119.

Gibbons, S., Mourato, S., & Resende, G. M. (2014). The amenity value of English nature: A hedonic price approach. *Environmental and Resource Economics, 57*, 175–196.

Goodchild, M. F. (2003). *The fundamental laws of GIScience*. University of California, Santa Barbara: Invited talk at University Consortium for Geographic Information Science.

Goodman, A. C., & Thibodeau, T. G. (2003). Housing market segmentation and hedonic prediction accuracy. *Journal of Housing Economics, 12*(3), 181–201.

Grum, B., & Govekar, D. K. (2016). Influence of macroeconomic factors on prices of real estate in various cultural environments: Case of Slovenia, Greece, France, Poland and Norway. *Procedia Economics and Finance, 39*, 597–604.

Hadavi, S., Kaplan, R., & Hunter, M. R. (2018). How does perception of nearby nature affect multiple aspects of neighbourhood satisfaction and use patterns? *Landscape Research, 43*(3), 360–379.

Hjort, A., Scheel, I., Sommervoll, D. E., & Pensar, J. (2024). Locally interpretable tree boosting: An application to house price prediction. *Decision Support Systems, 178*, Article 114106.

Julius, M., Sanderson, A., Roux, P. L., Simion, M., & Zenzile, M. (2020). Effects of environmental quality on urban housing prices: A hedonic multiple linear regression model approach. *Journal of Environmental Assessment Policy and Management, 22* (01n02), 2250003.

Kim, J. H., Li, W., Newman, G., Kil, S. H., & Park, S. Y. (2018). The influence of urban landscape spatial patterns on single-family housing prices. *Environment and Planning B: Urban Analytics and City Science, 45*(1), 26–43.

Kong, F., Yin, H., & Nakagoshi, N. (2007). Using GIS and landscape metrics in the hedonic price modeling of the amenity value of urban green space: A case study in Jinan City, China. *Landscape and urban planning, 79*(3–4), 240–252.

Kwon, M., Pickett, A. C., Lee, Y., & Lee, S. (2019). Neighborhood physical environments, recreational wellbeing, and psychological health. *Applied Research in Quality of Life, 14*, 253–271.

Lee, J., Irwin, N., Irwin, E., & Miller, H. J. (2020). The role of distance-dependent versus localized amenities in polarizing urban spatial structure: A spatio-temporal analysis of residential location value in Columbus, Ohio, 2000–2015. *Geographical Analysis, 53*(2), 283–306.

Lee, S. M., Conway, T. L., Frank, L. D., Saelens, B. E., Cain, K. L., & Sallis, J. F. (2017). The relation of perceived and objective environment attributes to neighborhood satisfaction. *Environment and Behavior, 49*(2), 136–160.

Li, Y., Miller, H. J., Root, E. D., Hyder, A., & Liu, D. (2022). Understanding the role of urban social and physical environment in opioid overdose events using found geospatial data. *Health & Place, 75*, Article 102792.

Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue, 16*(3), 31–57.

Liu, T., Hu, W., Song, Y., & Zhang, A. (2020). Exploring spillover effects of ecological lands: A spatial multilevel hedonic price model of the housing market in Wuhan, China. *Ecological Economics, 170*, Article 106568.

Lu, Y. (2019). Using Google Street View to investigate the association between street greenery and physical activity. *Landscape and Urban Planning, 191*, Article 103435.

Lundberg, S. (2017). *A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874*.

Maslow, A. H. (1971). *The farther reaches of human nature*. New York: Viking Press.

McKee, P., Erickson, D. J., Toomey, T., Nelson, T., Less, E. L., Joshi, S., & Jones-Webb, R. (2017). The impact of single-container malt liquor sales restrictions on urban crime. *Journal of Urban Health, 94*(2), 289–300.

Meinshausen, N., & Ridgeway, G. (2006). Quantile regression forests. *Journal of Machine Learning Research, 7*(6).

Middel, A., Lukasczyk, J., Zakrzewski, S., Arnold, M., & Maciejewski, R. (2019). Urban form and composition of street canyons: A human-centric big data and deep learning approach. *Landscape and Urban Planning, 183*, 122–132.

Nicholls, S., & Crompton, J. L. (2005). The impact of greenways on property values: Evidence from Austin, Texas. *Journal of Leisure Research, 37*(3), 321–341.

Piaggio, M. (2021). The value of public urban green spaces: Measuring the effects of proximity to and size of urban green spaces on housing market values in San José, Costa Rica. *Land Use Policy, 109*, Article 105656.

Plane, J., & Klodawsky, F. (2013). Neighbourhood amenities and health: Examining the significance of a local park. *Social Science & Medicine, 99*, 1–8.

Qiu, W., Zhang, Z., Liu, X., Li, W., Li, X., Xu, X., & Huang, X. (2022). Subjective or objective measures of street environment, which are more effective in explaining housing prices? *Landscape and Urban Planning, 221*.

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy, 82*(1), 34–55.

Saaty, T. L. (1980). *The analytic hierarchy process: Planning, priority setting, resource allocation*.

Sang, Å. O., Knez, I., Gunnarsson, B., & Hedblom, M. (2016). The effects of naturalness, gender, and age on how urban green space is perceived and used. *Urban Forestry & Urban Greening, 18*, 268–276.

Scornet, E., Biau, G., & Vert, J. P. (2015). Consistency of random forests. *Ann. Statist., 43* (4), 1716–1741.

Seaman, P. J., Jones, R., & Ellaway, A. (2010). It's not just about the park, it's about integration too: Why people choose to use or not use urban greenspaces. *International Journal of Behavioral Nutrition and Physical Activity, 7*(1), 1–9.

Soltani, A., Heydari, M., Aghaei, F., & Pettit, C. J. (2022). Housing price prediction incorporating spatio-temporal dependency into machine learning algorithms. *Cities, 131*, Article 103941.

Stiles, J., Li, Y., & Miller, H. J. (2022). How does street space influence crash frequency? An analysis using segmented street view imagery. *Environment and Planning B: Urban Analytics and City Science, 49*(9), 2467–2483.

Su, S., He, S., Sun, C., Zhang, H., Hu, L., & Kang, M. (2021). Do landscape amenities impact private housing rental prices? A hierarchical hedonic modeling approach based on semantic and sentimental analysis of online housing advertisements across five Chinese megacities. *Urban Forestry & Urban Greening, 58*, Article 126968.

Tripathi, S. (2019). *Macroeconomic determinants of housing prices: A cross country level analysis*.

Trojanek, R., Gluszak, M., & Tanas, J. (2018). The effect of urban green spaces on house prices in Warsaw. *International Journal of Strategic Property Management, 22*(5), 358–371.

Ulrich, R. S. (1984). View through a window may influence recovery from surgery. *Science, 224*(4647), 420–421. https://doi.org/10.1126/science.6143402

Wang, F., Wen, M., & Xu, Y. (2013). Population-adjusted street connectivity, urbanicity and risk of obesity in the U.S. *Applied Geography, 41*, 1–14.

Wu, C., Du, Y., Li, S., Liu, P., & Ye, X. (2022). Does visual contact with green space impact housing prices? An integrated approach of machine learning and hedonic modeling based on the perception of green space. *Land Use Policy, 115*.

Wu, C., Peng, N., Ma, X., Li, S., & Rao, J. (2020). Assessing multiscale visual appearance characteristics of neighbourhoods using geographically weighted principal component analysis in Shenzhen, China. *Computers, Environment and Urban Systems, 84*.

Xu, Y., & Wang, L. (2014). GIS-based analysis of obesity and the built environment in the U.S. *Cartography and Geographic Information Science, 42*(1), 9–21.

Yang, W. (2022). The determinants of housing prices: Evidence from Chinese cities. In *, vol. 151. SHS web of conferences* (p. 01003). EDP Sciences.

Yin, C., Peng, N., Li, Y., Shi, Y., Yang, S., & Jia, P. (2023). A review on street view observations in support of the sustainable development goals. *International Journal of Applied Earth Observation and Geoinformation, 117*, Article 103205.

Zambrano-Monserrate, M. A., Ruano, M. A., Yoong-Parraga, C., & Silva, C. A. (2021). Urban green spaces and housing prices in developed countries: A two-stage quantile spatial regression analysis. *Forest Policy and Economics, 125*, Article 102420.

Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution, 1*(1), 3–14.