



OPEN ACCESS

EDITED BY

Yibin Ao,
Chengdu University of
Technology, China

REVIEWED BY

Jixiang Liu,
Xiamen University, China
Zhuolin Tao,
Beijing Normal University, China

*CORRESPONDENCE

Lin Zhou
✉ zhoulunup@pku.edu.cn

SPECIALTY SECTION

This article was submitted to
Environmental Health and Exposome,
a section of the journal
Frontiers in Public Health

RECEIVED 09 November 2022

ACCEPTED 12 December 2022

PUBLISHED 06 January 2023

CITATION

Fan C, Gai Z, Li S, Cao Y, Gu Y, Jin C,
Zhang Y, Ge Y and Zhou L (2023) Does
the built environment of settlements
affect our sentiments? A multi-level
and non-linear analysis of Xiamen,
China, using social media data.
Front. Public Health 10:1094036.
doi: 10.3389/fpubh.2022.1094036

COPYRIGHT

© 2023 Fan, Gai, Li, Cao, Gu, Jin,
Zhang, Ge and Zhou. This is an
open-access article distributed under
the terms of the [Creative Commons
Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,
distribution or reproduction in other
forums is permitted, provided the
original author(s) and the copyright
owner(s) are credited and that the
original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution
or reproduction is permitted which
does not comply with these terms.

Does the built environment of settlements affect our sentiments? A multi-level and non-linear analysis of Xiamen, China, using social media data

Chenjing Fan¹, Zhenyu Gai¹, Shiqi Li¹, Yirui Cao¹, Yueying Gu¹,
Chenxi Jin¹, Yiyang Zhang¹, Yanling Ge¹ and Lin Zhou^{2,3*}

¹School of Landscape Architecture, Nanjing Forestry University, Nanjing, China, ²School of Public Administration and Policy, Renmin University of China, Beijing, China, ³Institute of Industrial Economics of Chinese Academy of Social Sciences, Beijing, China

Introduction: Humans spend most of their time in settlements, and the built environment of settlements may affect the residents' sentiments. Research in this field is interdisciplinary, integrating urban planning and public health. However, it has been limited by the difficulty of quantifying subjective sentiments and the small sample size.

Methods: This study uses 147,613 Weibo text check-ins in Xiamen from 2017 to quantify residents' sentiments in 1,096 neighborhoods in the city. A multilevel regression model and gradient boosting decision tree (GBDT) model are used to investigate the multilevel and nonlinear effects of the built environment of neighborhoods and subdistricts on residents' sentiments.

Results: The results show the following: (1) The multilevel regression model indicates that at the neighborhood level, a high land value, low plot ratio, low population density, and neighborhoods close to water are more likely to improve the residents' sentiments. At the subdistrict level, more green space and commercial land, less industry, higher building density and road density, and a smaller migrant population are more likely to promote positive sentiments. Approximately 19% of the total variance in the sentiments occurred among subdistricts. (2) The proportion of green space and commercial land, and the density of buildings and roads are linearly correlated with residents' sentiments. The land value is a basic need and exhibits a nonlinear correlation with sentiments. The plot ratio, population density, and the proportions of industrial land and the migrant population are advanced needs and are nonlinearly correlated with sentiments.

Discussion: The quantitative analysis of sentiments enables setting a threshold of the influence of the built environment on residents' sentiments in neighborhoods and surrounding areas. Our results provide data support for urban planning and implementing targeted measures to improve the living environment of residents.

KEYWORDS

neighborhoods, subdistricts, sentiment analysis, natural language processing, social media text data

1. Introduction

As an important comprehensive indicator to measure people's quality of life and mental health, subjective sentiments have attracted attention from various fields. According to the World Health Organization, 322 million people were affected by depression worldwide in 2017 (1). Due to rapid urbanization, the number of people with depression in China has reached 54 million; 73.6% are in a state of psychological sub-health, and 16.1% have psychological problems of varying degrees (2). Although residents' subjective sentiments are largely influenced by individual and family status, such as income, marriage, age, economic status, genetic indicators, and individual subjective indicators, environmental indicators comprise 40–50% of all factors affecting subjective sentiments (3). Studies have shown that people spend an average of 87% of their time indoors and about half of their time in their neighborhoods due to increasing urbanization (4). There is widespread awareness that improvements in the built environment of neighborhoods may improve the residents' sentiments. Research on the relationship between the living environment and psychological factors is a core topic in urban planning and environmental psychology, and the improvement of residents' sentiments is one goal of urban planning. Research on sentiments and living environments encompasses areas as diverse as inequality (5), space deprivation (6), and policy, which are critical to the health of rapidly urbanizing cities in developing countries.

Empirical studies based on questionnaires have shown that land prices (7, 8), location (9, 10), spatial form (11), and the built environment (such as the plot ratio, greening rate, property ownership, transportation organization, and density) (12–14) in neighborhoods may affect residents' sentiments. However, the largest challenge in this field is the large-scale quantification of sentiments. The concept of subjective sentiments has been typically used to evaluate an individual's sentiments (15). However, subjective sentiments are difficult to monitor or quantify in real time. For example, most studies relied on questionnaires, which have limitations, such as limited quantitative measurements, low coverage, recovery, and efficiency, and difficulty in replicating the results. In the past 2 years, few studies have used big data to quantify the sentiments of urban residents (16, 17). However, these data have rarely been used to conduct neighborhood-level research on residents' sentiments. More importantly, as discussed in the next section, objective, large-sample, non-discrete, and reproducible quantitative sentiment measurements may provide a detailed reference for studying multilevel and non-linear relationships between the built environment and sentiments.

This study uses social media text data and natural language processing (NLP) to quantify the residents' sentiments in Xiamen, China. A multilevel regression model is established at the neighborhood and subdistrict levels to investigate the

relationship between the built environment and residents' sentiments at different levels. We use the gradient boosting decision tree (GBDT) model to evaluate the non-linear correlation between variables with a significant impact. We attempt to answer the following questions: (1) What is the sentiment difference between residents living in different subdistricts of the city? (2) Which are the built environment indicators affecting individual subjective sentiments at the neighborhood and subdistrict levels? (3) Is the relationship between the built environment indicators and the sentiments non-linear or linear? This study uses objective social media text data to quantify long-term sentiments instead of short-term happiness (18) to provide a reference for interdisciplinary research on urban planning and public mental health. In addition, determining whether the relationship between the sentiments and the built environment at multiple levels is linear or non-linear is critical for optimizing the built environment of neighborhoods to improve residents' sentiments.

The rest of this article is organized as follows. Section 2 is a literature review of the quantification of sentiments, the built environment at multiple levels, and non-linear studies of sentiments to identify current research problems. Section 3 introduces the data, sentiment quantification methods, variables of the built environment, and multilevel and non-linear regression modeling methods. Sections 4 and 5 present the results and discussions. The final section summarizes the paper and discusses policy implications.

2. Literature review

2.1. Research on the quantitative analysis of residents' sentiments

Questionnaires are commonly used in sentiment studies and social research (19). However, this method may not be objective and may not reflect the psychological state of the subjects. The results of psychological state studies are influenced by the subjective feelings of the research participants and by the questionnaire design. Uncertain and confounding indicators may exist, such as the same questions applicable to different environmental conditions and inappropriate measurement methods. Moreover, discrete variables are typically used in questionnaires (20). Thus, the results need to be reclassified and scored to conform to a normal distribution and meet the requirements of statistical inference. A reproducible quantitative method can avoid some of the shortcomings of traditional surveys.

A limited number of studies have used social media data for sentiment analysis to determine the objective sentiments of residents in cities. Social media capture thousands of interactions between individuals and large groups over a long

period. Semantic analysis has been used to analyze large samples of highly objective and spatially and temporally resolved data to study the sentiments and wellbeing of individuals and societies. These data can be used to assess mental health and public sentiments (21, 22). One advantage of using social media data rather than questionnaires and interviews is the large sample size for analyzing sentiments (23). Text analysis and geographical analysis have been used to process social media data to obtain non-discrete and reproducible quantitative sentiment data with high spatial and temporal resolution (24). In general, the use of social media data to quantify sentiments is a widely used and accepted method in the academic community.

2.2. Complex relationship between the built environment and residents' sentiments

2.2.1. Multilevel analysis of the impact of the built environment on residents' sentiments

Studies on the influence of the built environment on residents' sentiments have been conducted primarily at two levels: the neighborhood and the surrounding environment (25). Baker and Steemers (26) stated that "In Britain, we spend, on average, as much as 90% of our time inside buildings, 70% of it in our own homes" (26). Therefore, the built environment of a neighborhood probably has the largest influence on residents' sentiments. Most studies are in agreement. For example, it has been shown that unsafe, inadequate facilities and poorly designed landscapes can significantly reduce residents' sentiments, potentially leading to psychological stress and mental problems. In contrast, environments with well-designed facilities, beautiful landscapes, low noise, and more daylight are more likely to evoke positive sentiments (27, 28). However, some disagreement exists on the effect of some indicators, such as the impact of building density indicators on sentiments. Most studies found that a higher building density is more unpleasant and results in negative sentiments (29, 30). However, a study conducted in Oslo reported that high building density might promote social relationships to improve residents' sentiments, provided that the environment is safe and not noisy (31).

Urban planning considers spaces outside neighborhoods (e.g., subdistricts, districts, towns) to enrich residents' daily activities (32). Research has focused on five aspects: land use, spatial form, development intensity, property ownership, and transportation organization (Table 1). (1) Early studies focusing on land use have consistently shown that neighborhoods far away from industrial areas have better public health. Safety and welfare can promote a clean environment and improve the quality of life (33, 34). A large amount of urban green space in neighborhoods can provide good air

quality and landscape conditions to enhance sentiments (35–37). In contrast, disagreement has developed over the impact of commercial land use on residents' sentiments. Some studies have found that areas of commercial land around neighborhoods can promote travel, reduce the dependence on cars, and lower residents' travel time and costs (38). However, the proximity of commercial land to residential sites results in more litter, high traffic noise, and low visual quality, potentially evoking negative sentiments (37). (2) Studies on the spatial form found that a large proportion of mixed land use reduced the average walking distance of residents from their homes to sites of interest and increased social interaction, improving the residents' sentiments (39). However, some empirical studies did not produce consistent results. Foord (40) observed that mixed land use improved the convenience and diversity of amenities for residents to meet their lifestyle needs. Cao (41) found that mixed land use in the Twin Cities, MN, provided more amenities but also resulted in more noise, traffic congestion, and possibly stranger danger, resulting in positive and negative effects on the residents' sentiments. However, the overall impact was statistically insignificant. (3) Higher development intensity results in higher population density and diverse impacts on the residents' sentiments. Some studies reported that a higher population density caused overcrowding, unemployment, poverty, and mental stress (31, 42, 43). Other studies suggest that a higher population density may improve residents' sentiments by enabling them to walk through their neighborhoods (23, 44, 45). (4) Some studies on property ownership reported that the migrant population caused a sense of insecurity and instability, increasing the mental stress of residents. When the proportion of the migrant population reached a specific size, the formation of group identity caused a stabilization of the sentiments (46, 47). (5) Many studies have examined the impact of road design on residents' sentiments. Some found that a higher road density provided increased connectivity between neighborhoods and significantly reduced congestion, improving residents' sentiments. It has also been argued that a high road density in neighborhoods can reduce the quality of life in a subdistrict due to landscape fragmentation. Too many road crossings can reduce access efficiency and make residents' travel experiences less enjoyable (48, 49). It has also been found that residents' sentiments are considerably influenced by traffic efficiency (e.g., rush hour, and traffic lights) (50) and that transportation organization and sentiments may not be correlated.

Early empirical studies primarily used simple linear regression models to explore the impact of the built environment on residents' sentiments. Subsequently, more complex regression models, such as multiple linear regression and structural equations, were used (62, 63). Although theoretical and empirical evidence suggests that the impact of the built environment on sentiments is multilayered, most studies focused on a single level and individuals.

TABLE 1 Studies of the relationship between the built environment and residents' sentiments.

| Indicators | Positive correlation | Negative correlation | Non-significance |
|--|-------------------------|----------------------|------------------|
| Built environment of the neighborhood | | | |
| Land value | (8), (7), (20)* | | |
| Plot ratio | | (25), (10) | |
| Landscape | (27), (28), (20)*, (10) | | |
| Land use | | | |
| Green space | (33), (34), (10) | | |
| Industrial land | | (51) | |
| Commercial land | (38) | (37) | |
| Spatial form | | | |
| Mixed land use | (40) | (10) | (41), (52) |
| Accessibility of facilities | (46), (47), (53)* | | |
| Development intensity | | | |
| Building density | (31) | (43), (42) | |
| Population density | (54), (55) | (30), (53)* | |
| Property ownership | | | |
| Proportion of migrant population | | (56), (57) | |
| Transportation organization | | | |
| Road density | (49), (58) | (59) | (50), (60) |
| Road intersections | | (48), (61) | |

*Indicates studies with non-linear correlations.

2.2.2. Studies on the non-linear correlation between the built environment and sentiments

Existing studies show that the built environment has non-linear relationships with overall sentiments (64). Referring to Maslow's theory, residents' needs regarding the environment can be divided into three categories: basic needs, intermediate needs, and advanced needs (65). Negative sentiments occur when basic needs are not met and vice versa. Positive sentiments occur when advanced needs are met, but negative sentiments do not occur when they are not met. Negative sentiments occur when intermediate needs are not met and vice versa (20, 66). Studies on residential environments found a non-linear relationship between negative sentiments and basic needs such as street lighting, residential safety, absence of noise, and nearby facilities (53, 67). A non-linear relationship was also observed between sentiments and advanced needs, such as diverse architectural styles, outstanding education, and good streetscape design (68, 69). However, questionnaire methods used in most sentiment analysis studies provide mostly discrete data, and there may be

errors in analyzing non-linear relationships. Using an ordinal scale to classify the sentiments of residents is subjective and does not provide a trend, making it difficult to determine whether the variance in the data is the result of random errors or curve fitting when assessing non-linear relationships.

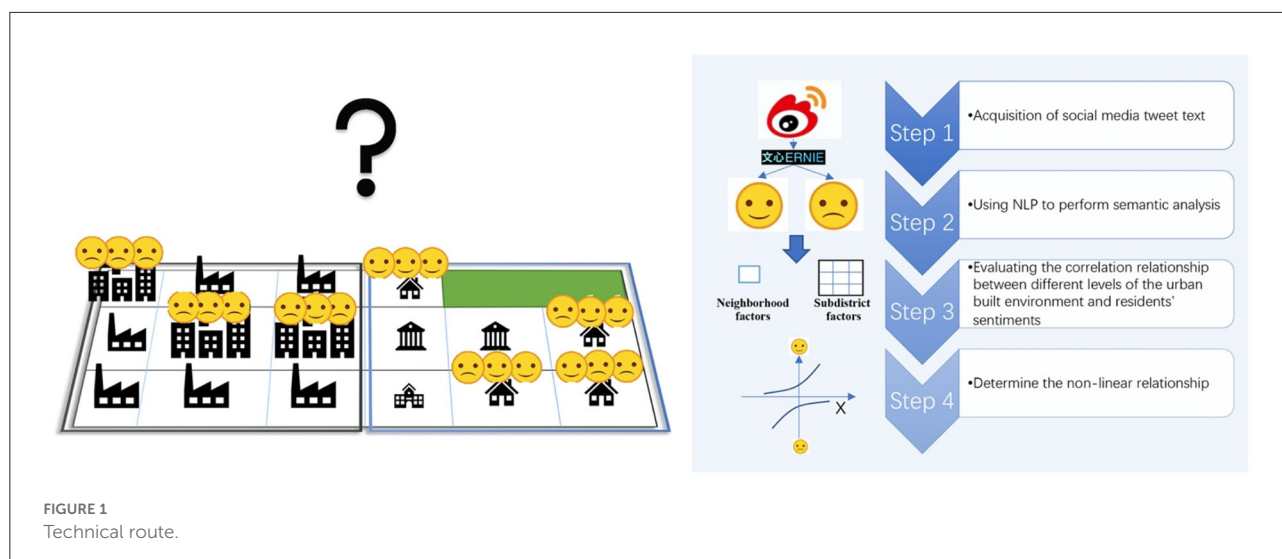
Although theoretical and empirical studies in public health and planning indicate a multilevel and non-linear correlation between the spatial environment and residents' short-term satisfaction, most multi-level analyses have focused on the individual and the environment. In contrast, the impact of multi-level differences in the spatial environment on residents' long-term sentiments has been based on theoretical approaches, and few empirical studies have been conducted. However, public health studies focused more on long-term sentiments than short-term satisfaction (18). Researchers started to use social media data to analyze the long-term sentiments of people ~2 years after the development of NLP techniques (70–72). These studies found complex relationships between people's sentiments and the built environment in boroughs with different region, no quantitative analysis was conducted. Furthermore, most studies used discrete data from questionnaires, which are highly subjective and contingent, making it difficult to replicate the results and assess the non-linear relationship between sentiments and the environment. Assessing the variable and non-linear relationship between the different elements of the built environment and the multi-level needs of residents requires more accurate and comprehensive data.

3. Data and methods

We use tweet text data from social media platforms and the sentiment knowledge enhanced pretraining (SKEP) algorithm to score the sentiments. Multilevel regression analysis and non-linear correlation analysis are used to assess the relationship between the built environment and residents' sentiments. The approach uses 4 steps: (1) Acquisition of social media tweet texts and data cleaning. (2) Using NLP to perform semantic analysis of the social media tweets. (3) Establishing a multilevel regression model to evaluate the correlation relationship between different levels of the urban built environment and residents' sentiments. (4) Using the GBDT model to determine the non-linear relationship between different indicators of the urban built environment and residents' sentiments (Figure 1).

3.1. Study area

The study area is Xiamen, China. Xiamen is located in East China in the southeast of Fujian Province; it has 50 subdistricts. The city covers an area of 1,699.39 square kilometers, and the permanent population of Xiamen was 4.01 million in 2017. Xiamen has repeatedly ranked first in China's



economic life survey as the happiest city in China. It has an excellent urban environment, a comfortable climate, and is a safe city with high urban development and numerous social and cultural activities (73). In addition, Xiamen's urban development ranks high in China, attracting a large migrant population for work. The complex population structure has also contributed to the formation of many urban villages, which are called *Chengzhongcun* in China (74). These are high-density villages surrounded by urban communities that have poor living conditions and are located in areas with high land prices. In addition, the old city of Xiamen is limited by the terrain, and the building density and housing prices are much higher than those of the new city. Therefore, the built environment of Xiamen is highly unbalanced and complex, making it ideal for this type of research. Many studies on the residents' sentiments and lives focused on tourism and housing have been conducted in Xiamen (70, 75, 76). We used the 2018 land-use map of Xiamen City to select the high-grade and medium-grade residential land (Code R1 and R2) and urban villages (Code R3) for this research (Figure 2).

3.2. Data sources

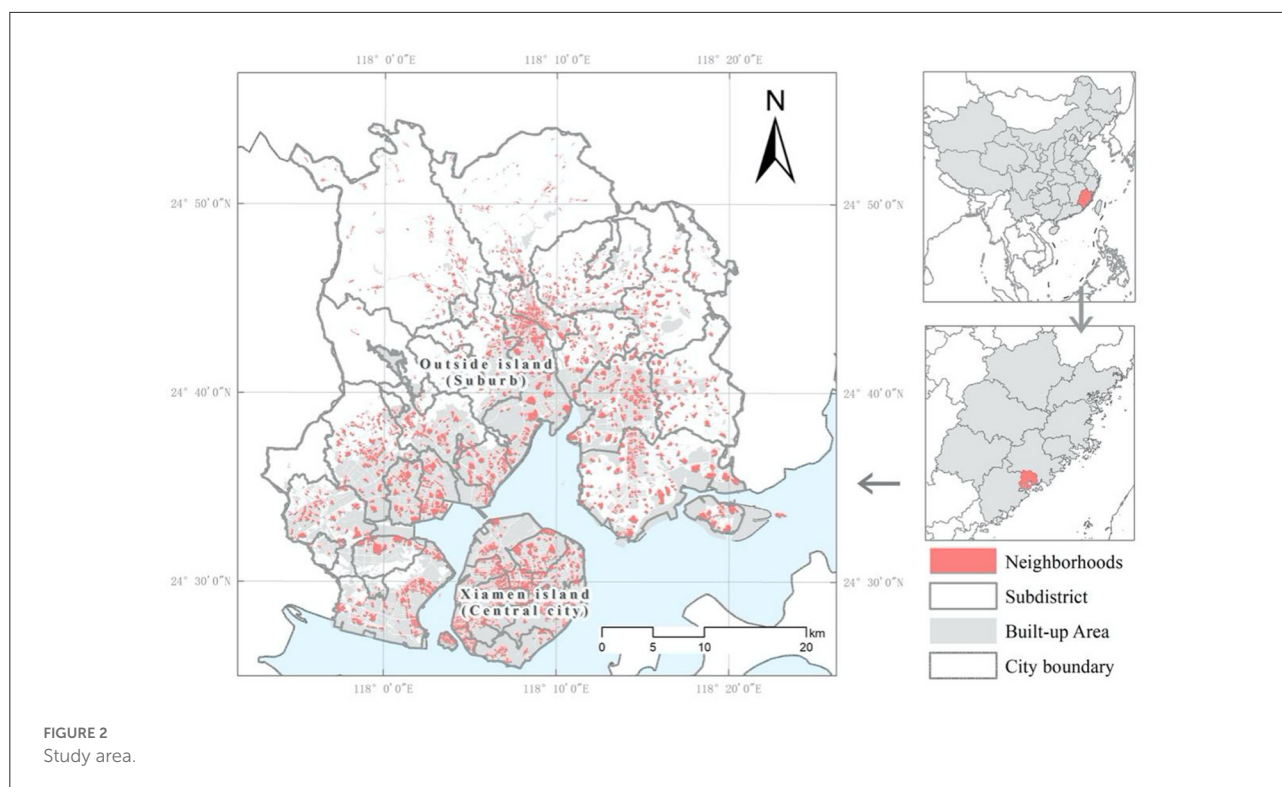
3.2.1. Data of residents' sentiments obtained from social media tweets

Social media tweet data include the social media user's geographic information (latitude and longitude coordinates), time, and text. Since check-in events are based on people's conscious behavior, people only post at a location if they stay for a relatively long time and think they have information worth recording. Thus, these tweets reflect the user's psychological state. The amount of data is larger than that obtained from questionnaires. Unlike mobile phone data and night lights, social

media check-in data contain information on human sentiments; thus, they are more suitable for studying the state of residents (77, 78).

We used Sina Weibo check-in data. Sina Weibo is a widely used social networking platform in China. Users are encouraged to check in frequently, recording their daily activity patterns and behaviors. The data used in this article was obtained by crawling the annual Weibo tweet data using the Sina Weibo Application Programming Interface (API) in 2017, including public data, such as Weibo tweets, generation time, user ID, and location. The data were filtered to remove repetitive, garbled, and other meaningless text or symbols in the text, such as URLs, HTML, tags, curly single or double quotes, email addresses, and non-ASCII characters (8, 79). The Weibo data with a neighborhood was extracted, and a total of 146,147 tweets were obtained. There were 1,096 neighborhoods with more than 10 tweets, exceeding the amount of data obtained from a questionnaire.

The quantification of the residents' sentiments was performed using deep learning classifiers. They can identify emojis and text in social media tweets, analyze subjective texts with sentiment overtones, and score the sentiments (70). We used the open access model SKEP based on ERNIE 3.0 to analyze the sentiments (80, 81). First, we used the pre-trained model dataset to build a learning platform based on sentiment knowledge. Then, we manually labeled 50,000 Weibo texts with sentiments. Machine learning based on pre-trained samples teaches computers how to quantify sentiments in tweets (82). Unlike scoring methods, such as sentiment dictionaries and cloud sentiment analysis (8, 83), the SKEP method is similar to human subjective thinking, and the method is reproducible. The sentiment scores of the tweet text have a range of [0, 1]. The closer the score is to 1, the more positive the sentiments are, and vice versa. The sentiment analysis preprocessing model was shared in



Figshare (<https://doi.org/10.6084/m9.figshare.21524391>). A comparison of the publicly available Chinese corpus (https://console.bce.baidu.com/ai/?_=1667982697826&fromai=1#/ai/nlp/sentiment/dict/list) annotated by Baidu Intelligent Cloud Sentiment analysis showed that the classification accuracy of the SKEP model was 95.8%. The precision and recall for the positive sentiments were 95.7 and 96.4%, and those for the negative sentiments were 96 and 95.1%, respectively, demonstrating the reliability of the sentiment classification results are reliable.

3.2.2. Built environment data

We obtained statistical data for the neighborhood and subdistricts (Table 1). In China, a neighborhood is the smallest residential unit for urban planning and management. It is constructed by a developer or the village collective, and the boundaries of the neighborhood determine in urban planning depend on the land granted. A subdistrict consists of several neighborhoods and is an administrative unit in China. It is subject to the urban planning goals of the government.

The neighborhood data was obtained from the website of *Fangtianxia*, a Chinese real estate market website (<https://www.fang.com/>). Typically, Internet real estate data are collected by designated staff and uploaded to the Internet by relevant institutions. The data include neighborhood construction information, such as neighborhood location, boundary, land value, number of households, building density, plot ratio, and

greening rate. We extracted the neighborhood boundaries from the current map of the urban master plan and calculated the information on living facilities, population, public facilities, and other information within 1 km of the neighborhood.

Urban dwellers spend most of their time in neighborhoods that provide the most important functions of daily life. Multiple indicators affect sentiments (Table 1). For example, the housing area, building density, plot ratio, and greening rate may affect the living experience of residents. The number of households, the proportion of the migrant population, and population density affect residents' social activities. Property fees and land values affect the living expenses of residents. Cultural, public, and transportation facilities affect the living experience of residents. We used two types of indicator statistics; one is a numerical variable for the number of neighborhoods, population structure, and land value. The other is a categorical variable, e.g., for cultural facilities and proximity to water.

The subdistrict characteristics were extracted from the 2018 land-use map of Xiamen City, including the subdistrict area, name, and region. We derived statistics for the subdistrict, including land use, spatial form, development intensity, property ownership, and transportation. The average of the sentiments for each neighborhood was used as the dependent variable.

Since a subdistrict is used by urban residents for jurisdiction and communication activities, the spatial elements have the largest impact on the residents' sentiments. For example,

different land-use types of the subdistrict can provide different functions. The public service facilities may affect the residents' accessibility, and the building density may reflect the development level of the subdistrict. The proportion of the migrant population in the subdistrict may affect the interpersonal experiences of residents, and road density may affect traffic quality.

The indicators and their calculation are listed in Table 2; X_n is a neighborhood-level indicator, and S_n is a subdistrict-level indicator.

3.3. Statistical analysis of the correlation between residents' sentiments and built environment indicators

3.3.1. Multilevel regression model

The sentiments of urban dwellers may have a multilevel relationship with the built environment. Some scholars have conducted hierarchical studies on the built environment at different administrative levels and under different development conditions to determine the impact of the built environment factors on residents' sentiments at multiple levels (20, 84). Our sample data had a hierarchical structure, with low-level neighborhood data nested within high-level subdistrict data. The objective of this study is to determine the effect of the built environment on the population's happiness. Thus, we focus more on public health (18) than individual attributes or satisfaction (85). Therefore, a multilevel regression model was used to analyze the differences between subdistricts and neighborhoods. Multi-level regression models can also explain the relationship of variables at different levels with the dependent variable. As a result, we used neighborhood-level indicators and subdistrict-level indicators to construct a two-level regression model.

The first step is to analyze the data hierarchy using an empty model with no explanatory variables. In this case, there are two levels; i is the neighborhood, and j is the subdistrict. Y_{ij} represents the observed variable of neighborhood i in subdistrict j . The model is defined as:

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij}. \tag{1}$$

The change in the intercept between neighborhoods can be expressed as:

$$\beta_{i0} = \gamma_{00} + u_{i0}. \tag{2}$$

The empty model is defined as:

$$Y_{ij} = \gamma_{00} + u_{i0} + \varepsilon_{ij}. \tag{3}$$

where γ_{00} is the mean intercept, β_{i0} is the neighborhood-level intercept, u_{i0} is the random effect of the neighborhood-level intercept, and ε_{ij} represents the estimated neighborhood-level difference in the built environment.

The inter-group correlation coefficient (ICC) can be calculated using the empty model. It is defined as the ratio of the variance between groups to the total variance:

$$ICC = \frac{\sigma_{0j}^2}{(\sigma_{0j}^2 + \sigma^2)}. \tag{4}$$

The empty model [Equation (3)] can estimate the variation across all subdistricts. σ_{0j}^2 is the neighborhood-level variance, and σ^2 is the subdistrict-level variance. The significance of $\hat{\sigma}_{0j}$ and the size of the ICC determines whether the sentiment difference is significantly affected by the subdistrict and whether a multi-level model is required.

If there is a difference, a subdistrict-level variable is added to Equation (3) to create Equation (5). Five models are constructed: land use, spatial form, development intensity, property ownership, and transportation organization. The significant built environment variables at the neighborhood level are retained, and the subdistrict-level variables are added.

$$\beta_{i0} = \gamma_{00} + \gamma_{i0}S_j + u_{i0}. \tag{5}$$

$$Y_{ij} = \beta_{i0} + \gamma_{0j}X_i + \varepsilon_{ij}. \tag{6}$$

Equations (5) and (6) are combined to obtain the final model:

$$Y_{ij} = \gamma_{00} + \gamma_{0j}X_i + \gamma_{i0}S_j + u_{i0} + \varepsilon_{ij}. \tag{7}$$

where X_j denotes the neighborhood variables, and S_j denotes the subdistrict-level variables that remain in the final model only if they are significant. $\gamma_{00} + \gamma_{0j}X_i + \gamma_{i0}S_j$ is the fixed effect, γ_{0j} is the main effect of explanatory variable X_i at the neighborhood level, γ_{i0} is the main effect of the explanatory variable S_j at level 2, and $u_{i0} + \varepsilon_{ij}$ is the random effects.

An empty model was constructed to determine the difference in people's sentiments between neighborhoods and subdistricts. We established a single-level model with five indicators at the neighborhood level. We then added variables at the subdistrict level (land use, spatial form, development intensity, property ownership, and transportation organization) to establish different models. Several indicators were removed to prevent multicollinearity (86, 87).

3.3.2. Non-linear GBDT regression model

The sentiments of city dwellers may have a non-linear relationship with the built environment. We used the GBDT method to describe the non-linear relationship between sentiments and neighborhood spatial features. The GBDT uses decision trees and gradient boosting regression trees.

TABLE 2 Research variables and their calculation methods.

| Level | Type | Indicator | Code | Unit | Calculation method |
|--------------------|--------------------------------|-------------------------------------|----------------|------------------------------|--|
| Dependent variable | | Sentiment score | Y | - | Quantitative analysis of social media tweet data, details in 3.2. |
| Neighborhood level | Land value | Land value | X ₁ | 10,000 RMB yuan | Average housing prices in public neighborhoods obtained from real estate websites (https://www.fang.com/). The price of houses that cannot be sold, such as urban villages, are estimated using spatial interpolation. |
| | Neighborhood built environment | Plot ratio | X ₂ | - | Total neighborhood building area/neighborhood land area. The data was obtained from real estate websites (https://www.fang.com/). |
| | | Population density | X ₃ | Number of people per hectare | Neighborhood population/neighborhood land area. The data was obtained from real estate websites (https://www.fang.com/). |
| | | Proximity to water | X ₄ | - | 0 = Close to the water. 1 = Not close to the water. |
| Subdistrict level | Land use | Proportion of green space | S ₁ | % | Total area of green space/total area of the subdistrict. The data was obtained from land-use maps. |
| | | Proportion of industrial land | S ₂ | % | Total area of industrial land/total area of the subdistrict. The data was obtained from land-use maps. |
| | | Proportion of commercial land | S ₃ | % | Total area of commercial land / total area of the subdistrict. The data is derived from land-use maps. |
| | Spatial form | Mixed land use | S ₄ | - | We used the information entropy Shannon-Wiener Index to calculate it: $H = -\sum (P_i) (\ln P_i)$, where P_i is the proportion of land use derived from land-use maps. |
| | | Number of public service facilities | S ₅ | Number | The number of public services in the subdistrict derived from the facilities in the current map. |
| | Development intensity | Building density | S ₆ | % | Subdistrict floor area/total subdistrict area. Building density is derived from housing boundaries calculated by the Housing and Urban-Rural Development Bureau. |
| | Property ownership | Proportion of migrant population | S ₇ | % | Number of migrant population/total subdistrict population. The data was obtained from the 2017 Population Census. |
| | Transportation organization | Road density | S ₈ | m/km ² | Total subdistrict road length/total subdistrict area. The data was obtained from the Traffic Bureau's road status map. |

- Indicates that the indicator is dimensionless.

The gradient boosting is based on the residuals of the previous tree (53). We used the mean of the response variable to predict and calculate the residuals and determine the difference between the observed and predicted values. Then, a tree was added to predict the residuals. The new predicted value of the response was the sum of the predicted values from the previous step. The predicted residuals were multiplied by the learning rate (a number from 0 to 1), and the new residuals were obtained by subtracting the new predicted values from the observed values. We repeated the second step until the addition of a tree did not improve the prediction result or the maximum number of trees was reached. The GBDT method uses additive regression models by sequentially fitting a simple parametric function to

the current residuals using least squares at each iteration (88). We selected index x_i with a significant correlation in the multi-level regression model.

First, the optimal constant model (Equation 8) was initialized to minimize the loss function $L(Y_{ij}, \gamma)$.

$$f_0(x) = \arg \min_{\gamma} \sum_N L(Y_{ij}, \gamma). \tag{8}$$

In the second step, each iteration of m had four sub-steps (a - d). First, we calculated the negative gradient using Equation (9). Next, we fit a regression tree to the target. The third sub-step was to calculate the gradient descent step size based on different tree

expansions using Equation (10). In the last sub-step, Equation 11, the model was updated based on the results of Equation (10).

For $m = 1$ to M :

(a) For $i = 1, 2, \dots, N$, calculate the pseudo-residuals

$$r_{im} = - \left[\frac{\partial L(Y_{ij}, f(x_i))}{\partial f(x_i)} \right]_{f=f_{m-1}}. \quad (9)$$

(b) Fit the target edge of the regression tree and determine the terminal area $R_{jm}, j = 1, 2, \dots, m$.

(c) For $j = 1, 2, \dots, m$, compute

$$\gamma_{jm} = \arg \min_y \sum_{x \in R_{jm}} L(Y_{ij}, f_{m-1}(x_i) + \gamma). \quad (10)$$

(d) Update

$$f_m(x_i) = f_{m-1}(x_i) + \sum_{j=1}^m \gamma_{jm} I(x_i \in R_{jm}). \quad (11)$$

The third step is to generate the final model as follows:

$$\hat{f}(x_i) = f_M(x_i). \quad (12)$$

where Y_{ij} is the sentiment mean, γ is the step size, r_{im} is the pseudo-residual, R_{jm} is the terminal region, M is the number of iterations, N is the number of eigenvalues, and J is the size of each constituent tree.

The regression curve was plotted according to the model fitting results, and the influence of different indicators on the sentiments was assessed. Positive sentiments occur when an intermediate need is met and vice versa. According to the three-factor theory, positive sentiments also occur when advanced needs are met, but negative sentiments do not occur when they are not met (89). The thresholds in the non-linear relationships were analyzed (Figure 3).

4. Results

4.1. Results of variable preprocessing

Table 3 list the descriptive statistics of the mean sentiments and variables at different levels. Figure 4 illustrates the high accuracy of the tweet data classification and the distribution of the non-discrete sentiment data. It was verified that the mean sentiments of the neighborhood followed a normal distribution, and the sample size was sufficiently large to reflect the sentiments of the neighborhood residents.

The average sentiment in Xiamen is 0.66, representing an average of 66 positive sentiments per 100 Weibo texts. The average sentiment on Xiamen Island is 0.67, and the average value outside of the island is 0.63, indicating a higher sentiment level on the island than outside the island. Figure 5 depicts the difference in the sentiments for different subdistricts. Therefore, it is necessary to perform multilevel regression.

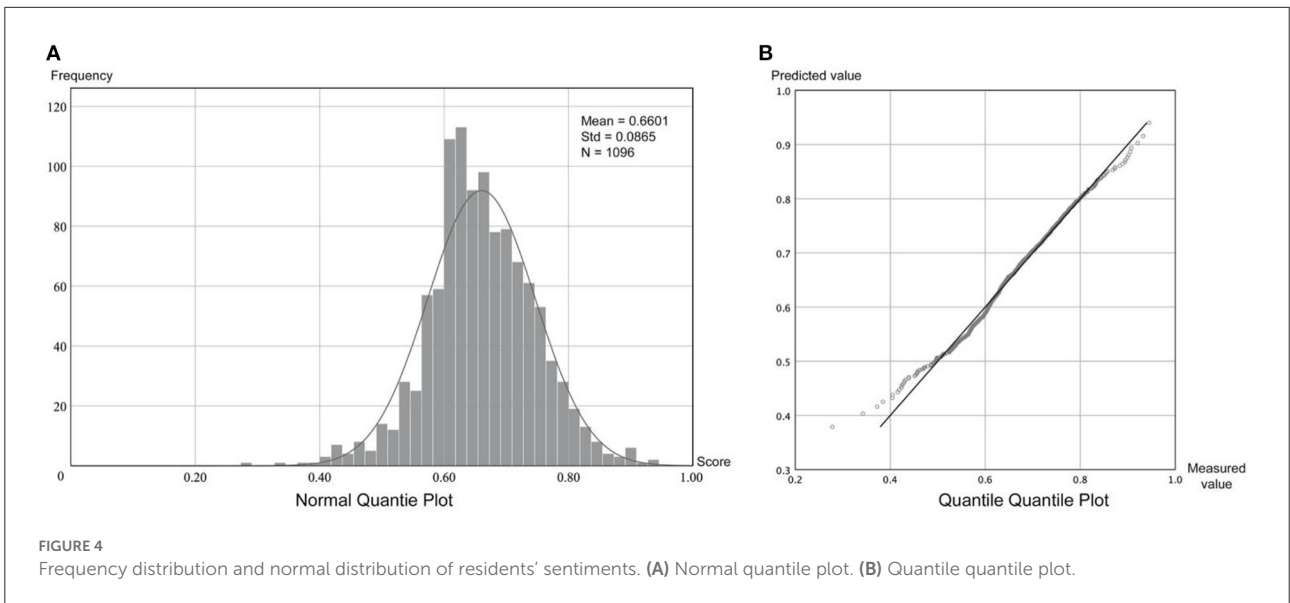
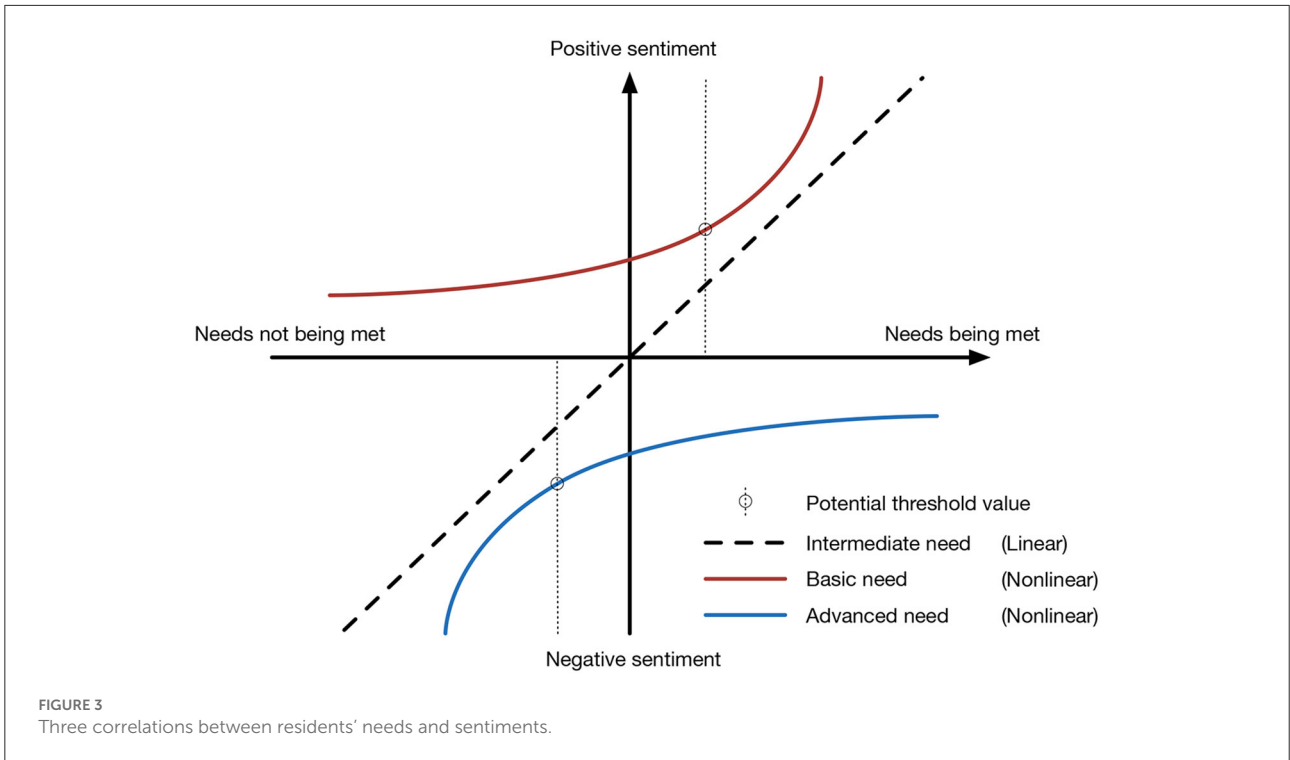
4.2. Multilevel regression results

The multi-level regression results in Table 4 show that in the empty model, the intergroup variance of the two-level regression model is 0.1986, the within-group variance is 0.8270, the ICC is 0.1937, and the $p < 0.001$, indicating that ~19% of the total variance in the sentiments occurred among subdistricts. The variance inflation factor (VIF) of model 1 is < 1.5 , indicating no multicollinearity between the indicators. At the neighborhood level, the land value and proximity to water are significantly positively correlated with the sentiments. The plot ratio and population density are significantly negatively correlated with the sentiments. At the subdistrict level, the proportion of green space, the proportion of commercial land, building density, and road density are significantly positively correlated with the sentiments. The proportion of industrial land and the proportion of the migrant population are significantly negatively correlated with the sentiments. Mixed land use and the number of public service facilities were not significantly correlated with the sentiments.

4.3. Non-linear regression results obtained from GBDT model

According to the results of the multilevel regression model, we selected indicators with significant correlations and used them in the GBDT model to assess the non-linear relationship between different indicators on residents' sentiments.

We set the parameters in the GBDT model by focusing on the key parameters, such as the number of trees and the shrinkage, following Fan et al. (89). We used the following GBDT model parameters: the number of trees was 10,000, the shrinkage was 0.0001, and the minimum number of observations in the terminal nodes of the trees was 20. The results are plotted in Figure 6. Since there were positive and negative correlations between different indicators, we plotted the growth of the built environment in the positive direction of the horizontal axis. The larger the indicator value, the more influence it has. We observed different degrees of non-linear correlations between the neighborhood and subdistrict variables that significantly affected the sentiments (Figure 6, Table 5). The land value had a greater impact on sentiments when the indicator value was low, and it was difficult to meet residents' needs. This is referred to as a basic need in the non-linear relationship. Population density, the proportion of industrial land, building density, and the proportion of the migrant population had a larger influence on sentiments when the indicator values were high, i.e., an advanced need in a non-linear relationship. A linear relationship was observed between the sentiments and the proportion of green space and the proportion of commercial land. Table 5 summarizes the non-linear characteristics and thresholds.



5. Discussion

5.1. Advantages of using social media data for sentiment quantification

The data used in previous related studies had either low spatial resolution (or a fuzzy classification was used)

or lacked quantitative information. Using questionnaires to survey residents' sentiments is costly and provides small sample sizes, making it difficult to capture residents' sentiments objectively. Inconsistent results may be obtained, and it is not possible to analyze the non-linear correlation between the built environment and sentiments. Using social media data for sentiment analysis allows for quantitative analysis

TABLE 3 Descriptive statistics of research variables.

| Code | Indicator | Number | Min | Max | Mean |
|----------------|-------------------------------------|--------|--------|-----------|-----------|
| Y | Average sentiment score | 1,090 | 0.28 | 0.94 | 0.66 |
| X ₁ | Land value | 1,090 | 0.70 | 14.20 | 4.57 |
| X ₂ | Plot ratio | 1,090 | 0.00 | 9.20 | 1.88 |
| X ₃ | Population density | 1,090 | 0.00 | 0.05 | 0.02 |
| X ₄ | Proximity to water | 1,090 | 0 | 1 | 0.28 |
| S ₁ | Proportion of green space | 50 | 0.01 | 52.04 | 18.01 |
| S ₂ | Proportion of industrial land | 50 | 0.00 | 27.48 | 6.51 |
| S ₃ | Proportion of commercial land | 50 | 0.00 | 14.53 | 5.92 |
| S ₄ | Mixed land use | 50 | 0.16 | 3.54 | 2.43 |
| S ₅ | Number of public service facilities | 50 | 27 | 211 | 102 |
| S ₆ | Building density | 50 | 0.36 | 30.16 | 13.19 |
| S ₇ | Proportion of migrant population | 50 | 7.40 | 84.47 | 45.87 |
| S ₈ | Road density | 50 | 936.64 | 27,989.16 | 13,000.30 |

of urban sentiments. Questionnaires may be subjective, with leading questions and a specific survey context. In contrast, social media data capture the residents' sentiments (90). The SKEP model enables the efficient quantification of the sentiments of urban residents. It provides a score of the sentiments and has the advantages of low cost, a large sample size, and quantitative and reproducible results (80).

5.2. Multilevel relationships between sentiments and the built environment

We observed differences in the relationships between sentiments and the built environment between the neighborhood and subdistrict.

At the neighborhood level, housing with a high land value and close proximity to the waterfront evoked positive sentiments, consistent with previous research (7, 8, 28). A low plot ratio and low population density resulted in negative sentiments, which is in line with residents' needs for high-quality housing. Our research also revealed the relationship between population density and sentiments. Previous research found that a high-population density worsened environmental conditions and increased noise, causing negative sentiments (30, 53). However, several studies found that an increase in population density in low-density areas improved the residents' sentiments by increasing opportunities for interpersonal interactions (54, 55). However, in compact cities like Xiamen, residents prefer low-density settlements. Reducing the population density of residential areas can

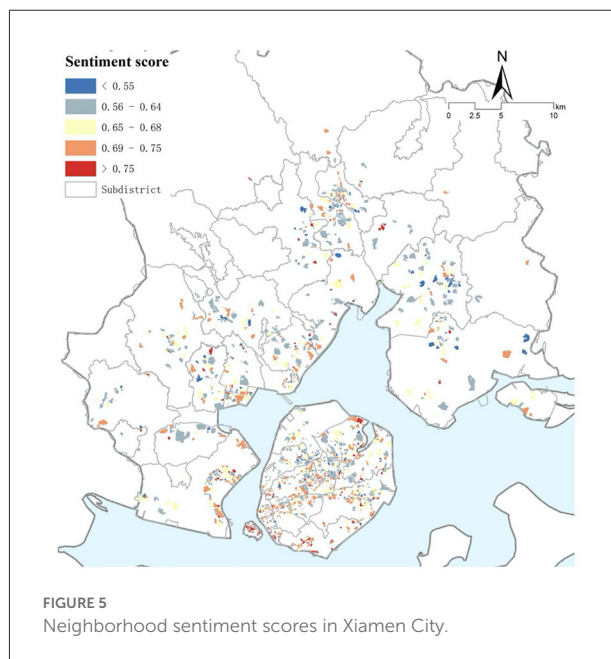


FIGURE 5 Neighborhood sentiment scores in Xiamen City.

improve the quality of life; thus, it is one objective of urban planners (91).

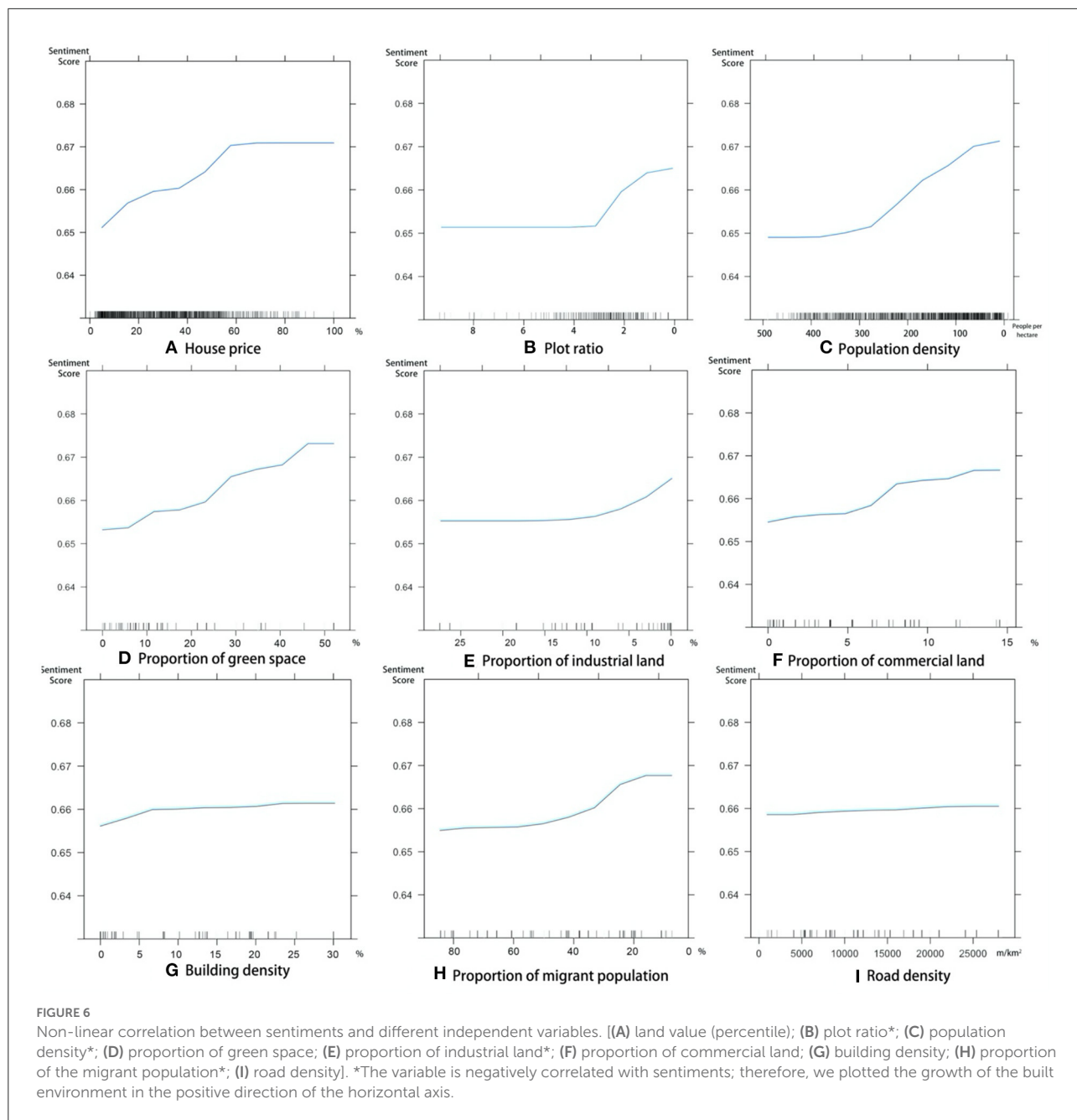
At the subdistrict level, more green spaces and commercial land improved the sentiments of the residents, whereas a large proportion of industry resulted in negative sentiments. An increase in the proportion of commercial land use significantly improved residents' sentiments, suggesting that residents enjoy the convenience offered by nearby commercial facilities (38) and are not bothered by the negative impacts, such as traffic congestion and noise, that some studies have associated with

TABLE 4 Two-level regression results for neighborhoods and subdistricts.

| Level | Dimension | Indicator | Empty model | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 | Model 6 | Model 7 |
|--------------|--------------------------------|-------------------------------------|-------------|-----------|-----------|-----------|----------|-----------|-----------|-----------|
| Level 1: | Land value | Land value | | 0.290*** | 0.173*** | 0.143*** | 0.299*** | 0.243*** | 0.221*** | 0.256*** |
| Neighborhood | Neighborhood built environment | Plot ratio | | -0.097*** | -0.073* | -0.078** | -0.093** | -0.096*** | -0.075* | -0.097*** |
| | | Population density | | -0.085* | -0.092** | -0.143*** | -0.082* | -0.135*** | -0.071* | -0.132** |
| | | Proximity to water | | 0.152* | 0.150* | — | 0.150* | 0.144* | 0.140* | 0.147* |
| Level 2: | Land use | Proportion of green space | | | 0.116*** | | | | | |
| Subdistrict | | Proportion of industrial land | | | -0.127*** | | | | | |
| | | Proportion of commercial land | | | | 0.262*** | | | | |
| | Spatial form | Mixed land use | | | | | — | | | |
| | | Number of public service facilities | | | | | — | | | |
| | Development intensity | Building density | | | | | | 0.106* | | |
| | Property ownership | Proportion of migrant population | | | | | | | -0.170*** | |
| | Transportation organization | Road density | | | | | | | | 0.092* |
| | | ICC | | 0.194 | | | | | | |
| | | AIC | 2,959.54 | 3,035.07 | 3,008.43 | 2,997.26 | 3,043.21 | 3,034.57 | 3,010.45 | 3,035.28 |

— Indicates no significant effect.

*** Represents significant at the 0.001 level, ** represents significant at the 0.01 level, and * represents significant at the 0.05 level.



commercial land use (37). A higher proportion of the migrant population in the neighborhood significantly worsened the residents' sentiments. Previous studies have also concluded that too many foreign renters affected the living experience of local residents and made it difficult for outsiders to integrate into local life (56, 57), resulting in negative sentiments. A high road density in a subdistrict indicates a higher level of development (49), increasing traffic and meeting the travel needs of more residents.

5.3. Non-linear correlation between sentiments and the built environment

The GBDT model can predict complex non-linear associations and is particularly efficient when the non-linear associations differ for different independent variables (89). Figure 6 shows the relationship between different indicators and the sentiments.

TABLE 5 Summary of relationships.

| Level | Code | Indicator | Type | Threshold |
|--------------|----------------|----------------------------------|-------------------|-----------|
| Neighborhood | X ₁ | Land value | Basic need | 60% |
| | X ₂ | Plot ratio | Advanced need | 3 |
| | X ₃ | Population density | Advanced need | 300 |
| Subdistrict | S ₁ | Proportion of green space | Intermediate need | * |
| | S ₂ | Proportion of industrial land | Advanced need | 10% |
| | S ₃ | Proportion of commercial land | Intermediate need | * |
| | S ₆ | Building density | Intermediate need | * |
| | S ₇ | Proportion of migrant population | Advanced need | 40% |
| | S ₈ | Road density | Intermediate need | * |

*Indicates that the indicator has no threshold.

The land value is a basic need of residents. These indicators had non-linear relationships with sentiments. The home owners' sentiments can be significantly affected by a change in the housing price when the price is <60% of the maximum value (about 60,000 yuan per square meter in Xiamen). At this price, most people in low-income and middle-income groups can own a house. Home owners are not affected by a change in house prices if the price is higher than 60% of the maximum value.

In contrast, the plot ratio, population density, proportion of industrial land, and the proportion of the migrant population were advanced needs and had a non-linear correlation with sentiments. The plot ratio had a threshold of 3. A plot ratio of 0–3 usually indicates low-rise and middle-rise housing. The lower the plot ratio in this range, the more spacious the living place; thus, a decrease in the plot ratio improved the residents' sentiments. When the plot ratio was more than 3 in middle-rise and high-rise residential houses, an increase in the plot ratio did not affect the housing type; therefore, there was a negligible effect on the sentiments. The lower the population density in the neighborhood, the more positive the sentiments of the inhabitants were at densities of <300 people per hectare. At greater densities, the sentiment remained relatively stable. The threshold for the proportion of industrial land within the subdistricts was 10% (e.g., industrial workers' living quarters, industrial attached neighborhoods). Above this threshold, the sentiments of the residents were not significantly

affected. In contrast, below 10%, as the percentage decreased, the sentiments of the residents improved substantially, which is in line with the planning and construction criteria of keeping neighborhoods away from industrial areas to improve their quality (33, 34). When the migrant population in the neighborhood was <20%, the sentiments of the residents were largely unaffected and remained high. Between 20 and 40%, a significant increase in the migrant population caused a significant decrease in the sentiments in the neighborhood. The residents exhibited more xenophobia, and the migrant population showed a decrease in their sense of belonging (56, 57). When the proportion of the migrant population exceeded 40%, the sentiments stabilized.

In general, advanced need represents the residents' desire for a high-quality residential life and should be the focus of high-quality construction in urban renewal projects. The findings of this article indicate inequalities in the sentiments in Xiamen due to differences in the built environment of neighborhoods and subdistricts. These differences may lead to inequalities in the health of residents (92). Therefore, planners and designers should focus on meeting the various needs of urban residents to promote health equity (93). High-density development in Xiamen currently meets the needs of residents at the middle and lower levels, although residents desire a high-quality built environment with low plot ratios, low population density, low proportion of industrial land, and low proportion of the migrant population in the neighborhood and district. These factors are considered in the current people-oriented urban renewal and transformation projects in Xiamen. The non-linear correlations and thresholds obtained in this study (Table 5) can inform decision-makers in other cities to meet the needs of residents at different stages of development. This information enables the use of planning and design tools in a targeted manner to improve the living experience of residents.

6. Conclusion

We used social media data and the SKEP model to quantify residents' sentiments in Xiamen, China. Multilevel regression models and GBDT models were used to investigate the effects of various indicators on sentiments in the neighborhood and subdistrict and determine the non-linear correlation between different indicators and sentiments. The multilevel regression results showed that neighborhoods with a higher land value, lower plot ratio, lower population density, and water frontage were more likely to evoke positive sentiments. At the subdistrict level, more green spaces and businesses, less industry, higher building and road densities, and a smaller migrant population were more likely to result in positive sentiments. Approximately 19% of the variability in the residents' sentiments was explained by the subdistrict indicators. We used the GBDT model to derive the non-linear correlations between the sentiments and different

indicators. The proportion of green space and commercial land and building and road density were linearly correlated with residents' sentiments. The land value is a basic need and exhibited a non-linear correlation with sentiments. The plot ratio, population density, the proportion of industrial land, and the proportion of the migrant population were advanced needs and had non-linear correlations with sentiments. The basic needs should be met first during planning and construction, whereas advanced needs are a direction for developing a high-quality living environment.

Quantitative analysis of sentiments can provide powerful data support for urban planners and decision-makers. This analysis is superior to using traditional questionnaires, which provide discrete data that do not reflect global attributes. Combining social media data and intelligent sentiment analysis algorithms can greatly improve the efficiency and accuracy of sentiment quantification. Priorities for planning and construction can be determined based on the needs of residents at different levels. Decision-makers should consider the correlation of different indicators with sentiments to optimize resource allocation. Future research should focus on assessing sentiments with larger sample sizes and including more built environment variables (e.g., climate, city size, etc.) to reveal the complex relationship between the urban built environment and residents' mental health from a people-centered perspective. In addition, our study on the relationship between environment and human emotions is exploratory in nature. There are currently no models that explain both multilevel and non-linear relationships. Further analyses of multilevel and non-linear relationships are needed.

Data availability statement

The original contributions presented in the study are included in the article/supplementary

material, further inquiries can be directed to the corresponding author.

Author contributions

CF led the project and provided the idea for this research. CF and ZG designed the research and wrote the paper. CF, ZG, SL, and CJ collected, analyzed, and validated the data. ZG created the figures. LZ revised and supervised the manuscript. YC, YGu, CJ, YZ, and YGe helped with the programming. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by the National Natural Science Foundation of China (51908309 and 51878367).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. WHO. *Depression and Other Common Mental Disorders: Global Health Estimates*. Geneva: World Health Organization (2017).
2. HMBCMA. *White Paper on Mental Health of Urban Residents in China*. Xi'an: Health Management Branch of Chinese Medical Association (2018).
3. Melton C. *The Drivers of Health: 4 Factors That Influence Individual and Population Health*. Tennessee: The Sycamore Institute (2018).
4. Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, et al. The national human activity pattern survey (Nhaps): a resource for assessing exposure to environmental pollutants. *J Expo Anal Environ Epidemiol*. (2001) 11:231–52. doi: 10.1038/sj.jea.7500165
5. Caragliu A, Del Bo CF. Smart cities and urban inequality. *Reg Stud*. (2022) 56:1097–112. doi: 10.1080/00343404.2021.1984421
6. Abascal A, Rodríguez-Carreño I, Vanhuysse S, Georganos S, Sliuzas R, Wolff E, et al. Identifying degrees of deprivation from space using deep learning and morphological spatial analysis of deprived urban areas. *Comput Environ Urban Syst*. (2022) 95:101820. doi: 10.1016/j.compenvurbysys.2022.101820
7. Zhang C, Zhang F. Effects of housing wealth on subjective well-being in urban China. *J Hous Built Environ*. (2019) 34:965–85. doi: 10.1007/s10901-019-09651-5
8. Tan MJ, Guan C. Are people happier in locations of high property value? Spatial temporal analytics of activity frequency, public sentiment and housing price using twitter data. *Appl Geog*. (2021) 132:102474. doi: 10.1016/j.apgeog.2021.102474
9. Dang Y, Zhan D, Qiu L, Wu S, Cui Y. Effects of the built environment on residents' subjective well-being and behaviours: a case of Hangzhou, China. *J Hous Built Environ*. (2022). doi: 10.1007/s10901-022-09955-z
10. Dang Y, Dong G, Chen Y, Jones K, Zhang W. Residential environment and subjective well-being in Beijing: a fine-grained spatial scale analysis using a bivariate response binomial multilevel model. *Environ Plan B Urban Analytic City Sci*. (2017) 46:648–67. doi: 10.1177/2399808317723012
11. Mouratidis K. Built environment and social well-being: how does urban form affect social life and personal relationships? *Cities*. (2018) 74:7–20. doi: 10.1016/j.cities.2017.10.020

12. Holy-Hasted W, Burchell B. Does public space have to be green to improve well-being? An analysis of public space across greater London and its association to subjective well-being. *Cities*. (2022) 125:103569. doi: 10.1016/j.cities.2022.103569
13. Coldwell DE, Evans KL. Visits to urban green-space and the countryside associate with different components of mental well-being and are better predictors than perceived or actual local urbanisation intensity. *Landsc Urban Plan*. (2018) 175:114–22. doi: 10.1016/j.landurbplan.2018.02.007
14. Oswald D, Moore T, Baker E. Exploring the well-being of renters during the COVID-19 pandemic. *Int J Hous Policy*. (2022) 1–21. doi: 10.1080/19491247.2022.2037177
15. Fredrickson BL, Branigan C. Positive emotions broaden the scope of attention and thought-action repertoires. *Cogn Emot*. (2005) 19:313–32. doi: 10.1080/02699930441000238
16. Bannister J, O'Sullivan A. Big data in the city. *Urban Studies*. (2021) 58:3061–70. doi: 10.1177/00420980211014124
17. Tran M, Draeger C, Wang X, Nikbakht A. Monitoring the well-being of vulnerable transit riders using machine learning based sentiment analysis and social media: lessons from COVID-19. *Environ Plan B Urban Anal City Sci*. (2022). doi: 10.1177/23998083221104489
18. Zheng SQ, Wang JH, Sun C, Zhang XN, Kahn ME. Air pollution lowers chinese urbanites' expressed happiness on social media. *Nat Human Behav*. (2019) 3:237–43. doi: 10.1038/s41562-018-0521-2
19. Sjöström O, Holst D. Validity of a questionnaire survey: response patterns in different subgroups and the effect of social desirability. *Acta Odontol Scand*. (2002) 60:136–40. doi: 10.1080/000163502753740133
20. Cao J, Hao Z, Yang J, Yin J, Huang X. Prioritizing neighborhood attributes to enhance neighborhood satisfaction: an impact asymmetry analysis. *Cities*. (2020) 105:102854. doi: 10.1016/j.cities.2020.102854
21. Dong E, Du H, Gardner L. An interactive web-based dashboard to Track COVID-19 in real time. *Lancet Infect Dis*. (2020) 20:E215. doi: 10.1016/S1473-3099(20)30509-0
22. Diener E, Oishi S, Lucas RE. Personality, culture, and subjective well-being: emotional and cognitive evaluations of life. *Annu Rev Psychol*. (2003) 54:403–25. doi: 10.1146/annurev.psych.54.101601.145056
23. Melis G, Gelormino E, Marra G, Ferracin E, Costa G. The effects of the urban built environment on mental health: a cohort study in a large Northern Italian City. *Int J Environ Res Public Health*. (2015) 12:14898–915. doi: 10.3390/ijerph121114898
24. Dorostkar E, Najarsadeghi M. How to evaluate urban emotions using twitter social media? *Cities*. (2022) 127:103713. doi: 10.1016/j.cities.2022.103713
25. Guha M, Channon B. Mental health in the built environment. *J Mental Health*. (2020) 29:611–3. doi: 10.1080/09638237.2020.1836559
26. Baker N, Steemers K. *Healthy Homes: Designing with Light and Air for Sustainability and Wellbeing*. London: RIBA Publishing (2019).
27. Qiu Y, Liu Y, Liu Y, Li Z. Exploring the linkage between the neighborhood environment and mental health in Guangzhou, China. *Int J Environ Res Public Health*. (2019) 16:3206. doi: 10.3390/ijerph16173206
28. Youssoufi S, Houot H, Vuidel G, Pujol S, Mauny F, Foltête J-C. Combining visual and noise characteristics of a neighborhood environment to model residential satisfaction: an application using gis-based metrics. *Landsc Urban Plan*. (2020) 204:103932. doi: 10.1016/j.landurbplan.2020.103932
29. He D, Miao J, Lu Y, Song Y, Chen L, Liu Y. Urban greenery mitigates the negative effect of urban density on older adults' life satisfaction: evidence from Shanghai, China. *Cities*. (2022) 124:103607. doi: 10.1016/j.cities.2022.103607
30. Morrison PS. Local expressions of subjective well-being: the New Zealand experience. *Reg Stud*. (2011) 45:1039–58. doi: 10.1080/00343401003792476
31. Mouratidis K. Compact city, urban sprawl, and subjective well-being. *Cities*. (2019) 92:261–72. doi: 10.1016/j.cities.2019.04.013
32. Leverentz A, Pittman A, Skinnon J. Place and perception: constructions of community and safety across neighborhoods and residents. *City Commun*. (2018) 17:972–95. doi: 10.1111/cico.12350
33. Syrbe R-U, Neumann I, Grunewald K, Brzoska P, Louda J, Kochan B, et al. The value of urban nature in terms of providing ecosystem services related to health and well-being: an empirical comparative pilot study of cities in Germany and the Czech Republic. *Land*. (2021) 10:341. doi: 10.3390/land10040341
34. Honold J, Beyer R, Lakes T, van der Meer E. Multiple environmental burdens and neighborhood-related health of city residents. *J Environ Psychol*. (2012) 32:305–17. doi: 10.1016/j.jenvp.2012.05.002
35. Dzhambov AM, Markevych I, Tilov B, Arabadzhiyev Z, Stoyanov D, Gatseva B, et al. Pathways linking residential noise and air pollution to mental ill-health in young adults. *Environ Res*. (2018) 166:458–65. doi: 10.1016/j.envres.2018.06.031
36. Hoisington AJ, Stearns-Yoder KA, Schuldt SJ, Beemer CJ, Maestre JP, Kinney KA, et al. Ten questions concerning the built environment and mental health. *Build Environ*. (2019) 155:58–69. doi: 10.1016/j.buildenv.2019.03.036
37. Ellis CD, Lee S-W, Kweon B-S. Retail land use, neighborhood satisfaction and the urban forest: an investigation into the moderating and mediating effects of trees and shrubs. *Landsc Urban Plan*. (2006) 74:70–8. doi: 10.1016/j.landurbplan.2004.10.004
38. Oosterbaan C, Arku G, Asiedu AB. Conversion of residential units to commercial spaces in Accra, Ghana: a policy dilemma. *Int Plan Stud*. (2012) 17:45–66. doi: 10.1080/13563475.2011.638185
39. Wu W, Chen WY, Yun Y, Wang F, Gong Z. Urban Greenness, mixed land-use, and life satisfaction: evidence from residential locations and workplace settings in Beijing. *Landsc Urban Plan*. (2022) 224:104428. doi: 10.1016/j.landurbplan.2022.104428
40. Foord J. Mixed-use trade-offs: how to live and work in a compact city neighbourhood. *Built Environ*. (2010) 36:47–62. doi: 10.2148/benv.36.1.47
41. Cao X. How does neighborhood design affect life satisfaction? Evidence from Twin Cities. *Travel Behav Soc*. (2016) 5:68–76. doi: 10.1016/j.tbs.2015.07.001
42. Firdaus G. Built environment and health outcomes: identification of contextual risk factors for mental well-being of older adults. *Ageing Int*. (2017) 42:62–77. doi: 10.1007/s12126-016-9276-0
43. Ochodo C, Ndeti DM, Moturi WN, Otieno JO. External built residential characteristics that affect mental health of adults. *J Urban Health*. (2014) 91:908–27. doi: 10.1007/s11524-013-9852-5
44. Khan NY, Ghafoor N, Iftikhar R, Malik M. Urban annoyances and mental health in the city of Lahore, Pakistan. *J Urban Aff*. (2012) 34:297–315. doi: 10.1111/j.1467-9906.2011.00585.x
45. Simone C, Carolin L, Max S, Reinhold K. Associations between community characteristics and psychiatric admissions in an urban area. *Soc Psychiatry Psychiatr Epidemiol*. (2013) 48:1797–808. doi: 10.1007/s00127-013-0667-1
46. Wang F, Wang D. Changes in residential satisfaction after home relocation: a longitudinal study in Beijing, China. *Urban Stud*. (2019) 57:583–601. doi: 10.1177/0042098019866378
47. Chen Y, Lü B, Chen R. Evaluating the life satisfaction of peasants in concentrated residential areas of Nanjing, China: a fuzzy approach. *Habitat Int*. (2016) 53:556–68. doi: 10.1016/j.habitatint.2016.01.002
48. Li X, Sun J-Q. Effects of vehicle-pedestrian interaction and speed limit on traffic performance of intersections. *Physica A*. (2016) 460:335–47. doi: 10.1016/j.physa.2016.05.034
49. Choi D, Ewing R. Effect of street network design on traffic congestion and traffic safety. *J Transport Geog*. (2021) 96:103200. doi: 10.1016/j.jtrangeo.2021.103200
50. Wang S, Yu D, Kwan M-P, Zheng L, Miao H, Li Y. The Impacts of road network density on motor vehicle travel: an empirical study of Chinese cities based on network theory. *Transport Res Part A Policy Pract*. (2020) 132:144–56. doi: 10.1016/j.tra.2019.11.012
51. Wang BY, Li T, Yao ZH. Institutional uncertainty, fragmented urbanization and spatial lock-in of the pen-urban area of China: a case of industrial land redevelopment in panyu. *Land Use Policy*. (2018) 72:241–9. doi: 10.1016/j.landusepol.2017.12.054
52. Bertram C, Goebel J, Krekel C, Rehdanz K. Urban land use fragmentation and human wellbeing. *Land Econ*. (2021) 98:399–420. doi: 10.3368/le.98.2.122019-0175R1
53. Ding C, Cao X, Naess P. Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. *Transport Res Part A Policy Pract*. (2018) 110:107–17. doi: 10.1016/j.tra.2018.02.009
54. Buys L, Miller E. Residential satisfaction in inner urban higher-density brisbane, australia: role of dwelling design, neighbourhood and neighbours. *J Environ Plan Manag*. (2012) 55:319–38. doi: 10.1080/09640568.2011.597592
55. Mouratidis K. Is compact city livable? The impact of compact versus sprawled neighbourhoods on neighbourhood satisfaction. *Urban Stud*. (2017) 55:2408–30. doi: 10.1177/0042098017729109
56. Xiao Y, Miao S, Sarkar C, Geng H, Lu Y. Exploring the impacts of housing condition on migrants' mental health in nanxiang, shanghai: a structural equation modelling approach. *Int J Environ Res Public Health*. (2018) 15:225. doi: 10.3390/ijerph15020225

57. Hu F. Homeownership and subjective wellbeing in urban china: does owning a house make you happier? *Soc Indic Res.* (2013) 110:951–71. doi: 10.1007/s11205-011-9967-6
58. Li HM, Zhang YT, Ao YB, Wang Y, Wang T, Chen YF. Built environment impacts on rural residents' daily travel satisfaction. *Front Ecol Evol.* (2022) 10:931118. doi: 10.3389/fevo.2022.931118
59. Cai X, Wu Z, Cheng J. Using kernel density estimation to assess the spatial pattern of road density and its impact on landscape fragmentation. *Int J Geog Inform Sci.* (2013) 27:222–30. doi: 10.1080/13658816.2012.663918
60. Cui YL, Yu Y, Cai ZY, Wang DH. Optimizing road network density considering automobile traffic efficiency: theoretical approach. *J Urban Plann Develop.* (2022) 148:0402106. doi: 10.1061/(ASCE)UP.1943-5444.0000780
61. Morency P, Archambault J, Cloutier M-S, Tremblay M, Plante C. Major urban road characteristics and injured pedestrians: a representative survey of intersections in Montréal, Quebec. *Can J Public Health.* (2015) 106:e388–94. doi: 10.17269/CJPH.106.4821
62. Evans MDR, Kelley J, Kelley SMC, Kelley CGE. Rising income inequality during the great recession had no impact on subjective wellbeing in Europe, 2003–2012. *J Happiness Stud.* (2019) 20:203–28. doi: 10.1007/s10902-017-9917-3
63. Zagorski K, Evans MDR, Kelley J, Piotrowska K. Does national income inequality affect individuals' quality of life in Europe? Inequality, happiness, finances, and health. *Soc Indi Res.* (2014) 117:1089–110. doi: 10.1007/s11205-013-0390-z
64. Wu X, Cao J, Huting J. Using three-factor theory to identify improvement priorities for express and local bus services: an application of regression with dummy variables in the twin cities. *Trans Res Part A Policy Pract.* (2018) 113:184–96. doi: 10.1016/j.tra.2018.04.003
65. Maslow AH. *Motivation and Personality.* New York, NY: Prabhat Prakashan (1987).
66. Matzler K, Sauerwein E, Heischmidt K. Importance-performance analysis revisited: the role of the factor structure of customer satisfaction. *Serv Indust J.* (2003) 23:112–29. doi: 10.1080/02642060412331300912
67. Wang X, Shao C, Yin C, Guan L. Disentangling the comparative roles of multilevel built environment on body mass index: evidence from China. *Cities.* (2021) 110:103048. doi: 10.1016/j.cities.2020.103048
68. Lieske SN, van den Nouweland R, Han JH, Pettit C, A. Novel hedonic price modelling approach for estimating the impact of transportation infrastructure on property prices. *Urban Stud.* (2019) 58:182–202. doi: 10.1177/0042098019879382
69. Dong H, Qin B. Exploring the link between neighborhood environment and mental wellbeing: a case study in Beijing, China. *Landsc Urban Plan.* (2017) 164:71–80. doi: 10.1016/j.landurbplan.2017.04.005
70. Gai Z, Fan C, Shen S, Ge Y, Shi Z, Li S, et al. Using social media data to explore urban land value and sentiment inequality: a case study of Xiamen, China. *Wirel Commun Mob Comput.* (2022) 2022:1456382. doi: 10.1155/2022/1456382
71. Gao Y, Chen YY, Mu L, Gong SZ, Zhang PC, Liu Y. Measuring urban sentiments from social media data: a dual-polarity metric approach. *J Geogr Syst.* (2022) 24:199–221. doi: 10.1007/s10109-021-00369-z
72. Ji HM, Wang J, Meng B, Cao Z, Yang T, Zhi GQ, et al. Research on adaption to air pollution in chinese cities: evidence from social media-based health sensing. *Environ Res.* (2022) 210:112762. doi: 10.1016/j.envres.2022.112762
73. China Central Television (CCTV). *China Economic Life Survey.* Beijing: National Bureau of Statistics of China (NBCS) (2021).
74. Tian L, Yao Z, Fan C, Zhou L. A systems approach to enabling affordable housing for migrants through upgrading Chengzhongcun: a case of Xiamen. *Cities.* (2020) 105:102186. doi: 10.1016/j.cities.2018.11.017
75. Liu J, Wang B, Xiao L. Non-linear associations between built environment and active travel for working and shopping: an extreme gradient boosting approach. *J Trans Geog.* (2021) 92:103034. doi: 10.1016/j.jtrangeo.2021.103034
76. Yan J, Bao HXH. A prospect theory-based analysis of housing satisfaction with relocations: field evidence from China. *Cities.* (2018) 83:193–202. doi: 10.1016/j.cities.2018.06.022
77. Yan L, Duarte F, Wang D, Zheng S, Ratti C. Exploring the effect of air pollution on social activity in china using geotagged social media check-in data. *Cities.* (2019) 91:116–25. doi: 10.1016/j.cities.2018.11.011
78. Ye C, Zhang F, Mu L, Gao Y, Liu Y. Urban function recognition by integrating social media and street-level imagery. *Environ Plan B Urban Analyt City Sci.* (2020) 48:1430–44. doi: 10.1177/2399808320935467
79. Zhu F, Ning Y, Chen X, Zhao Y, Gang Y. On removing potential redundant constraints for svor learning. *Appl Soft Comput.* (2021) 102:106941. doi: 10.1016/j.asoc.2020.106941
80. Sun Y, Wang S, Feng S, Ding S, Pang C, Shang J, et al. Ernie 3.0: large-scale knowledge enhanced pre-training for language understanding and generation. *arXiv [Preprint].* (2021). arXiv: 2107.02137. doi: 10.48550/arXiv.2107.02137
81. Tian H, Gao C, Xiao X, Liu H, He B, Wu H, et al. Skep: sentiment knowledge enhanced pre-training for sentiment analysis. *ACL.* (2020) 4067–76. doi: 10.18653/v1/2020.acl-main.374
82. Zhu F, Gao J, Yang J, Ye N. Neighborhood linear discriminant analysis. *Pattern Recognit.* (2022) 123:108422. doi: 10.1016/j.patcog.2021.108422
83. Xu L, Li L, Jiang Z, Sun Z, Wen X, Shi J, et al. A novel emotion lexicon for chinese emotional expression analysis on weibo: using grounded theory and semi-automatic methods. *IEEE Access.* (2021) 9:92757–68. doi: 10.1109/ACCESS.2020.3009292
84. Mouratidis K. Neighborhood characteristics, neighborhood satisfaction, and well-being: the links with neighborhood deprivation. *Land Use Policy.* (2020) 99:104886. doi: 10.1016/j.landusepol.2020.104886
85. Kahneman D, Krueger AB. Developments in the measurement of subjective well-being. *J Econ Perspect.* (2006) 20:3–24. doi: 10.1257/089533006776526030
86. Fan C, Tian L, Zhou L, Hou D, Song Y, Qiao X, et al. Examining the impacts of urban form on air pollutant emissions: evidence from China. *J Environ Manage.* (2018) 212:405–14. doi: 10.1016/j.jenvman.2018.02.001
87. Yuan M, Song Y, Huang Y, Hong S, Huang L. Exploring the association between urban form and air quality in China. *J Plan Educ Res.* (2017) 38:413–26. doi: 10.1177/0739456X17711516
88. Jerome HF. Greedy function approximation: a gradient boosting machine. *Annals Stat.* (2001) 29:1189–232. doi: 10.1214/aos/1013203450
89. Fan L, Cao J, Hu M, Yin C. Exploring the importance of neighborhood characteristics to and their nonlinear effects on life satisfaction of displaced senior farmers. *Cities.* (2022) 124:103605. doi: 10.1016/j.cities.2022.103605
90. Roberts H, Sadler J, Chapman L. The value of twitter data for determining the emotional responses of people to urban green spaces: a case study and critical evaluation. *Urban Stud.* (2018) 56:818–35. doi: 10.1177/0042098017748544
91. Okulicz-Kozaryn A, Mazelis JM. Urbanism and happiness: a test of wirth's theory of urban life. *Urban Stud.* (2016) 55:349–64. doi: 10.1177/0042098016645470
92. Fan C, Ouyang W, Tian L, Song Y, Miao W. Elderly health inequality in china and its determinants: a geographical perspective. *Int J Environ Res Public Health.* (2019) 16:2953. doi: 10.3390/ijerph16162953
93. Mehdipanah R, Schulz AJ, Israel BA, Gamboa C, Rowe Z, Khan M, et al. Urban heart detroit: a tool to better understand and address health equity gaps in the City. *J Urban Health.* (2018) 95:662–71. doi: 10.1007/s11524-017-0201-y