IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 June 2023, accepted 5 August 2023, date of publication 9 August 2023, date of current version 22 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3303571

APPLIED RESEARCH

Smart Cities and Access to Nature: A Framework for Evaluating Green Recreation Space Accessibility

JENNY MARTINEZ [1] **AND SACHIT MAHAJAN** [2]
[1]Decision Sciences Group, Universidad Nacional de Colombia, Medellín 050034, Colombia
[2]Computational Social Science (COSS), ETH Zürich, 8092 Zürich, Switzerland

Corresponding author: Sachit Mahajan (sachit.mahajan@gess.ethz.ch)

The work of Jenny Martinez was supported by the Colombian Administrative Department of Science, Technology, and Innovation (Colciencias) National Doctoral under Grant 727. The work of Sachit Mahajan was supported by the Project "CoCi: Co-Evolving City Life," through the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program under Grant 833168.

ABSTRACT As our world becomes increasingly urbanized, smart cities are leading the way in using technology to create more efficient, connected, and sustainable environments. However, amidst all the talk of connectivity and smartness, it's crucial not to lose sight of one of the most basic human needs: access to nature in cities. This research describes a novel open-source framework for investigating the availability and accessibility of green recreation spaces using open-source data and statistical analytic approaches. The framework includes a comprehensive set of tools for data extraction, processing, analysis, and visualization, thereby enabling reproducible geospatial research. We test our framework on five international cities: Medellin, Milan, Chicago, Singapore, and Mumbai. Through geospatial analysis and statistical modeling of data sourced from OpenStreetMaps, we explore and comprehend the characteristics and distribution of spatial accessibility related to green recreation spaces in five cities. We find significant clustering of green recreation spaces in all these cities, indicating that a majority of such spaces are located in close proximity to each other within small areas. Our findings also shed light on the potential implications of unequal distribution of green recreation spaces for the health and well-being of city residents and highlight the need for policies and initiatives that promote equitable access to green recreation spaces in smart cities.

INDEX TERMS Accessibility, cities, geospatial analysis, recreation spaces, open-source.

I. INTRODUCTION

Urbanization is one of the defining trends of the twenty-first century, with cities now housing more than half of the world's population [1]. While cities provide numerous advantages, such as economic opportunities and cultural diversity, they also pose significant challenges, such as traffic congestion, environmental pollution, and social inequality [2], [3], [4], [5], [6]. By harnessing technology to build more sustainable, efficient, and livable urban settings, smart cities have emerged as a possible answer to these difficulties [7]. However, the traditional emphasis on data-centric techniques and connectivity has frequently overlooked the relevance of nature and accessibility in influencing the urban quality of life. Accessibility is a multifaceted concept that encompasses several dimensions, including physical, and socio-economic aspects. In the context of urban planning, accessibility refers to the ease with which people can reach and use essential services and amenities, such as green recreation spaces [8]. These spaces, which include parks, gardens, and other natural environments for recreation, play an important role in promoting physical and mental well-being, fostering social cohesion, and enhancing the overall quality of life for urban residents [9], [10]. As cities continue to grow and become more densely populated, it is important to consider the role that green recreation spaces play in promoting community well-being [11], [12]. Green recreation spaces can provide a variety of benefits to city

The associate editor coordinating the review of this manuscript and approving it for publication was Gang Mei.

residents, including improved mental and physical health, social cohesion, and a sense of community pride [13], [14]. There is a growing body of evidence that demonstrates the positive impact of green spaces on health. For example, a recent study found that living near green space is associated with a lower risk of mortality [15], while another found that exposure to green space can lead to reductions in stress and anxiety [16]. In addition, green recreation spaces can also promote physical activity, which is an important factor in maintaining good health [17]. As we delve deeper into the development of smart cities, it becomes increasingly apparent that the significance of access to nature, such as green recreation spaces has been undervalued. The conventional understanding of what constitutes a smart city has largely centered on the deployment of cutting-edge technologies such as Artificial Intelligence (AI), the Internet of Things, etc. to optimize the efficacy and efficiency of government services [18], [19]. While this approach has yielded positive results in many parts of the world, there remains room for a paradigm shift from a data-driven to a data-informed approach. By leveraging technological advancements and available data, this approach could prioritize promoting community well-being and access to nature, instead of solely prioritizing government efficiency and productivity. As a result, there is a growing need for a more comprehensive strategy that integrates technology, community, and nature to develop sustainable, healthy, and inclusive urban environments that promote livability, sustainability, and resilience. This transition is critical for addressing urbanization-related issues such as limited access to green spaces and the detrimental impact on community well-being.

In this study, we propose an open-source,[1] reproducible, and extendable framework that can be used to investigate the distribution and accessibility of green recreation spaces (as well as other amenities) in cities using open-source data. As compared to the previous works in this area that look at the accessibility of infrastructure like green spaces through the lens of socio-economic factors, we follow a more quantitative approach that looks into the spatial distribution of green recreation spaces by performing statistical analysis to understand their spatial arrangement in the cities followed by a network analysis to understand the accessibility of those spaces. We evaluate the framework on five global cities, namely Medellin, Milan, Chicago, Singapore, and Mumbai, in order to identify potential disparities in green recreation space distribution and access. Our findings suggest that despite the differences in geography, level of development, and planning strategies across these cities, green recreation spaces tend to cluster in certain areas, which impacts their accessibility to citizens. By analyzing the accessibility of green spaces, our approach offers a systematic and quantitative means for city planners to identify areas that require improvement in order to promote greater access for all members of society. In this way, our research contributes to the development of a more equitable and sustainable urban environment.

The rest of the paper is organized as follows: In Section II, we discuss the related work. In Section III, we describe the methods used, followed by the presentation of results in Section IV. The discussion is encapsulated in Section V, and finally, Section VI concludes the study.

II. BACKGROUND

In this section, we briefly go through the related work in the area of accessibility analysis of city amenities like green recreation spaces. We also discuss the commonly used terms as well as methods that are widely used in the literature to discuss the accessibility of social infrastructure in cities as well as the use of digital technology and open-source data for decision-making in smart cities.

The growing availability and use of Volunteered Geographic Information (VGI) has helped to improve the analysis and prediction capabilities of sectors such as urban planning, energy infrastructure, etc. In the past, researchers have used data sources like OpenStreetMap (OSM) data, Mapillary data as well as Google Places data to analyze urban areas [20], [21], [22]. Many studies have investigated the accessibility of urban green spaces [23], [24] as well as other amenities in cities [25], employing various methods and metrics. Some researchers have used distance-based measures, such as the proximity of residential areas to the nearest green space [26], while others have used more sophisticated approaches, such as the two-step floating catchment area method, which accounts for both supply and demand factors [27]. Additionally, researchers have explored the role of socioeconomic factors in shaping accessibility patterns, revealing that lower-income and minority communities often face greater barriers to accessing green spaces [28], [29]. There has also been a significant increase in the use of open-source digital technologies as well as openly available data to map cities as well as create data-driven applications [30], [31], [32], [33]. Several recent works have utilized OSM data [25], [34], remote-sensing data [35], [36] as well as sensor-based data to understand how the resources and infrastructure are distributed in cities and how it has changed over the time [37].

To further understand the existing literature, we conducted a keyword co-occurrence network analysis to gain insights into the evolution of the research field, identify trends, and uncover research opportunities and gaps. This type of analysis allows us to visualize how different keywords, representing various research themes, are interconnected based on their co-occurrence in the same papers. It provides a holistic view of the research landscape, highlighting the main areas of focus and their interrelations.

For this analysis, we used a query[2] to search for relevant papers in the Web of Science database. This query was

[1] https://github.com/sachit27/Accessibility-Analysis

[2] "city" AND "accessibility" AND "green space*."

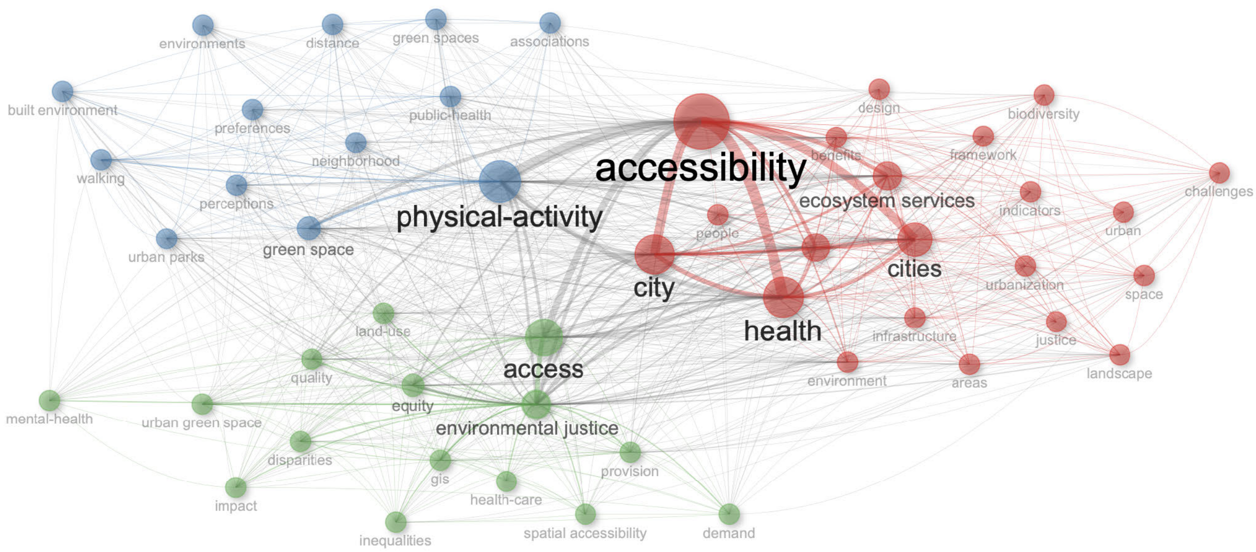

FIGURE 1. Representation of various research themes and clusters, derived from the keyword co-occurrence network analysis of existing literature.

designed to capture papers that focus on the intersection of cities, accessibility, and green spaces, which is the core interest of our study. The search resulted in a total of 640 papers published between 2003 and 2023. From these papers, we extracted keywords and constructed a network with 50 nodes, each representing a unique keyword. Nodes were connected if the corresponding keywords co-occurred in the same papers, with a minimum of 2 edges required for a connection. To identify clusters of closely related keywords, we used the Louvain method for community detection. Based on the keyword co-occurrence network (as shown in Figure 1), we can observe several clusters of research themes and trends in the field of cities, accessibility, and green spaces.

- Cluster 1 (Red): This cluster underscores the exploration of the link between green space accessibility and health outcomes, with a focus on mental and physical health benefits and broader ecosystem services.
- Cluster 2 (Blue): This cluster is focused on the relationship between green space and physical activity, including walking, perceptions of green space, and the built environment.
- Cluster 3 (Green): This cluster emphasizes research on equitable green space distribution and its implications for environmental justice.

While the existing research clusters provide valuable insights into the field of cities, accessibility, and green spaces, our work introduces innovative approaches that address two significant gaps:

- Innovation through Open-Source Tools: Current studies in the field have not extensively utilized open-source tools for analyzing accessibility and green spaces. Our research proposes an open-source framework specifically designed to investigate the availability and accessibility of green spaces. The novelty of our approach lies in the adaptability of open-source tools, which can be widely adopted, integrated with various data sets, and updated continuously. This makes our methodology a versatile and valuable resource for ongoing research in this field.
- Advancement through Quantitative and Statistical Analysis: While there is a lot of discussion around environmental justice and other indicators, there seems to be limited quantitative and statistical analysis that can support data-informed decisions. Our work contributes to addressing this gap by using statistical analytic approaches to explore and comprehend the characteristics and distribution of spatial accessibility related to green recreation spaces.

III. METHODOLOGY

In this section, we will discuss in detail the methodology that is used in this work. Figure 2 provides an overview of the framework that is used for understanding the spatial distribution of green recreation spaces in a city as well as analyzing the accessibility of those recreation spaces. The framework comprises three integrated workflows. The first workflow is used to extract the street network of a location using OSM API. The OSM data is accessed using the ‘‘osmnx’’ package in Python [38]. This package provides easy-to-use functions for retrieving, processing, and manipulating OSM data. This workflow takes the name of the city as well as the network type to extract the street network. In this work, we used the ‘‘walk’’ network type to extract the street network suitable for pedestrians. The extracted street network is then converted into a GeoDataFrame that contains the nodes and edges of the street network. The second step in this workflow entails retrieving the location’s green recreational spaces. We then extract the polygons of green recreation spaces in the city. We specify the tags

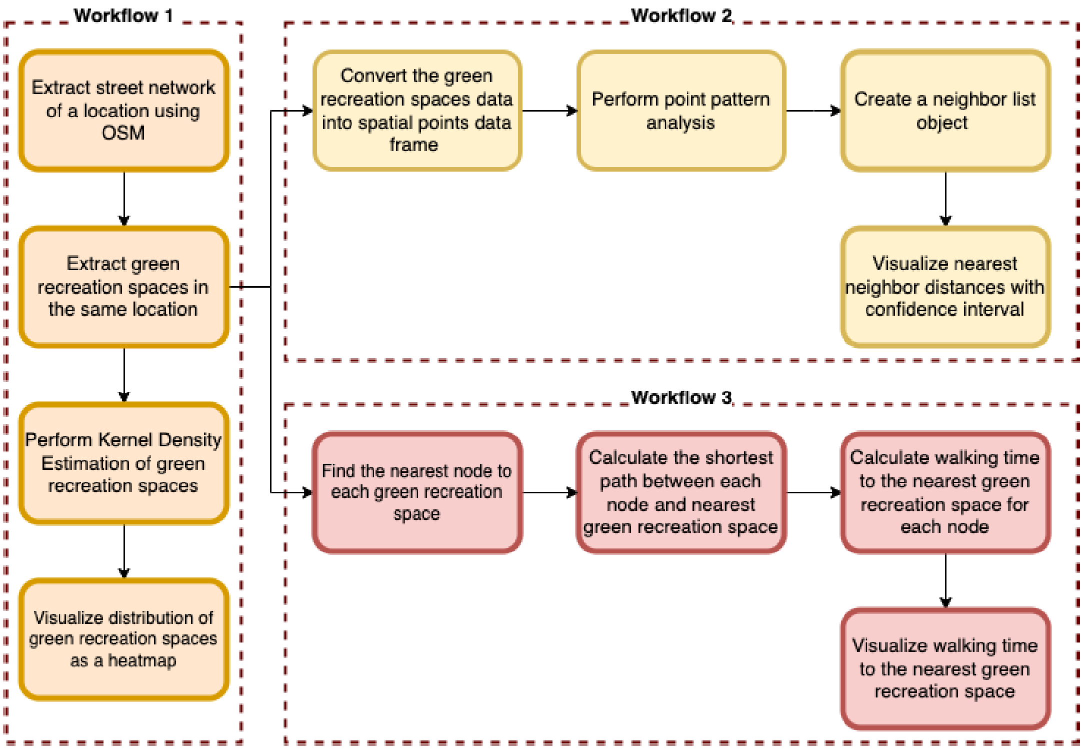

FIGURE 2. A schematic representation of the proposed framework consisting of three workflows.

to narrow down the green spaces with the tags "leisure" and "landuse" that are connected with parks and recreation areas.

The second workflow uses the extracted green recreation spaces data to first visualize the distribution of the green recreation spaces as a density heatmap. The density estimation is performed using a kernel density estimation (KDE) algorithm [39]. KDE calculates the density at each point as a weighted sum of nearby points. This is followed by a point pattern analysis approach to understand the spatial distribution of green recreation spaces. Point pattern analysis is a statistical method used to study the spatial arrangement of points in two-dimensional space. This approach is widely used in geospatial analysis, ecology, and epidemiology to analyze the spatial patterns of different variables [40], [41]. In our case, we are interested in understanding how green recreation spaces are distributed in a region of interest. To do this, we create a point pattern object from the coordinates of the green recreation spaces within the specified window. The window represents the extent of the region of interest. We use the K function [42] to quantify the spatial distribution of points in a point pattern. It is defined as the expected number of points within a certain distance of any point in the pattern, normalized by the density of the points. K function can be represented as:

$$K(d) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{A}\mathbb{I}(\|x_i - x_j\| \leq d) \tag{1}$$

where $K(d)$ is the value of the K-function at distance d, n is the total number of points in the study region, x_i and x_j are the locations of the i-th and j-th points and A is the area of the study region. The function $\mathbb{I}$ returns the value 1 if the distance between points i and j is less than or equal to d, and 0 otherwise. The envelope test is a method of hypothesis testing that is commonly used to determine if the pattern of points is significantly different from a random spatial distribution. It involves generating a large number of simulated point patterns and calculating the K function for each simulation. The results are then used to create an envelope of confidence intervals around the observed K function values. Following the point pattern analysis, we use the nearest neighbor approach [43] to quantify the degree of clustering or dispersion of points in a point pattern. This approach calculates the distances between each point (in this case, the green recreation spaces) and its nearest neighbors. By examining the distances between the points, we can determine if the distribution is random, clustered or dispersed.

To perform the nearest neighbor analysis, we first create a neighbor list object and set the distance range between 0 and 1000 meters to identify the neighboring points within this range. The formula for the nearest neighbor distance (d) is given below:

$$d = \min(dist(x_i, x_j)) \tag{2}$$

where d is the nearest neighbor distance, x_i is the location of the i^{th} point in the pattern, and x_j is the location of the nearest neighbor to the i^{th} point. The G function is used to quantify the degree of clustering or dispersion in a point pattern [44]. It is defined as the cumulative distribution function of the nearest neighbor distances. The formula for the G function is given below:

$$G(d) = P(d(x_i) \leq d) \tag{3}$$

where $G(d)$ is the value of the G function at distance d, $d(x_i)$ is the nearest neighbor distance of point i, and $P(d(x_i) \leq d)$ is the probability that the nearest neighbor distance is less than or equal to d. The expected value of the G function for a random spatial distribution is given by the following equation:

$$G(r) = 1 - \exp\left(-\frac{\pi r^2}{\lambda}\right) \tag{4}$$

where $G(r)$ is the expected value of the G function at distance r, λ is the intensity of the point pattern (i.e., the number of points per unit area), and $\exp\left(-\frac{\pi r^2}{\lambda}\right)$ is the Poisson probability of finding a point within a distance r of a randomly chosen point in a Poisson point process with intensity λ. The ratio of the observed G function to the expected G function is called the K function, which is the same as the K function used in point pattern analysis. We also use an envelope approach to create a confidence interval around the nearest neighbor histogram. The confidence interval is calculated by simulating 1000 point patterns and calculating the nearest neighbor distances for each simulated pattern.

The point pattern analysis and nearest neighbor approach are appropriate methods to understand the distribution of green recreation spaces in cities. These methods allow us to determine if these spaces are randomly distributed or if they are clustered or dispersed. This information can help urban planners and policymakers make informed decisions about the location and distribution of green spaces in the city. Additionally, this approach can be applied to other cities to compare the distribution of green spaces and identify areas that may require additional green spaces.

The final workflow deals with understanding the accessibility of green recreation spaces in cities by calculating the shortest walking distance from the nearest node to the points of interest (POI). The POI here is the location data of the green recreation space. The methodology using POI data has been widely used in several application domains such as air quality analysis, land-use analysis, finding crime hot spots, etc. [45], [46]. To perform this, we first calculate the nearest node to each POI using the nearest neighbor search algorithm. We then select the nearest node that is part of the pedestrian network. This is followed by the calculation of the shortest walking distance from the nearest node to the POI using the Dijkstra algorithm. The Dijkstra algorithm finds the shortest path between nodes in a graph, in this case, the pedestrian network. This is represented mathematically as:

$$D(i, j) = \min_{v \in V} dist(i, v) + D(v, j) \tag{5}$$

TABLE 1. The number of green recreation spaces in cities analyzed in this article.

City	Number of green recreation spaces
Medellin	816
Milan	790
Chicago	870
Singapore	432
Mumbai	963

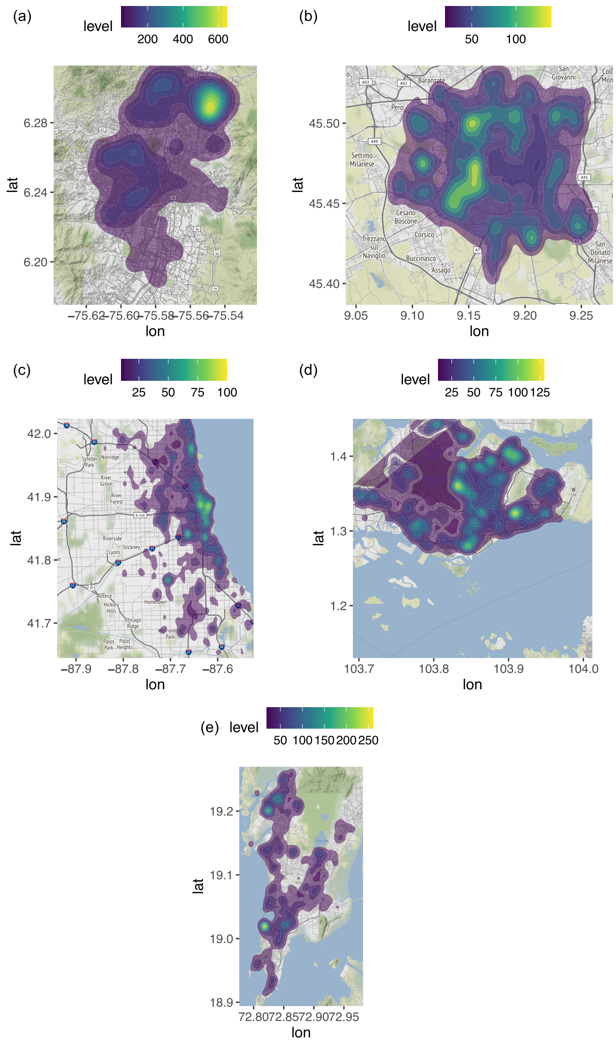

FIGURE 3. Density maps showing the distribution of green recreation spaces in (a) Medellin, (b) Milan, (c) Chicago, (d) Singapore, and (e) Mumbai. The legend "level" refers to the density level of the green recreation spaces.

where $D(i, j)$ is the shortest path between node i and node j, v is the set of all nodes in the graph, $dist(i, v)$ is the distance between node i and node v, and $D(v, j)$ is the shortest path between node v and node j.

After calculating the shortest walking distance from each POI to the nearest node on the pedestrian network, we aggregate this information to understand the accessibility of green recreation spaces in the city. It is important to consider here that our analysis did not take into consideration

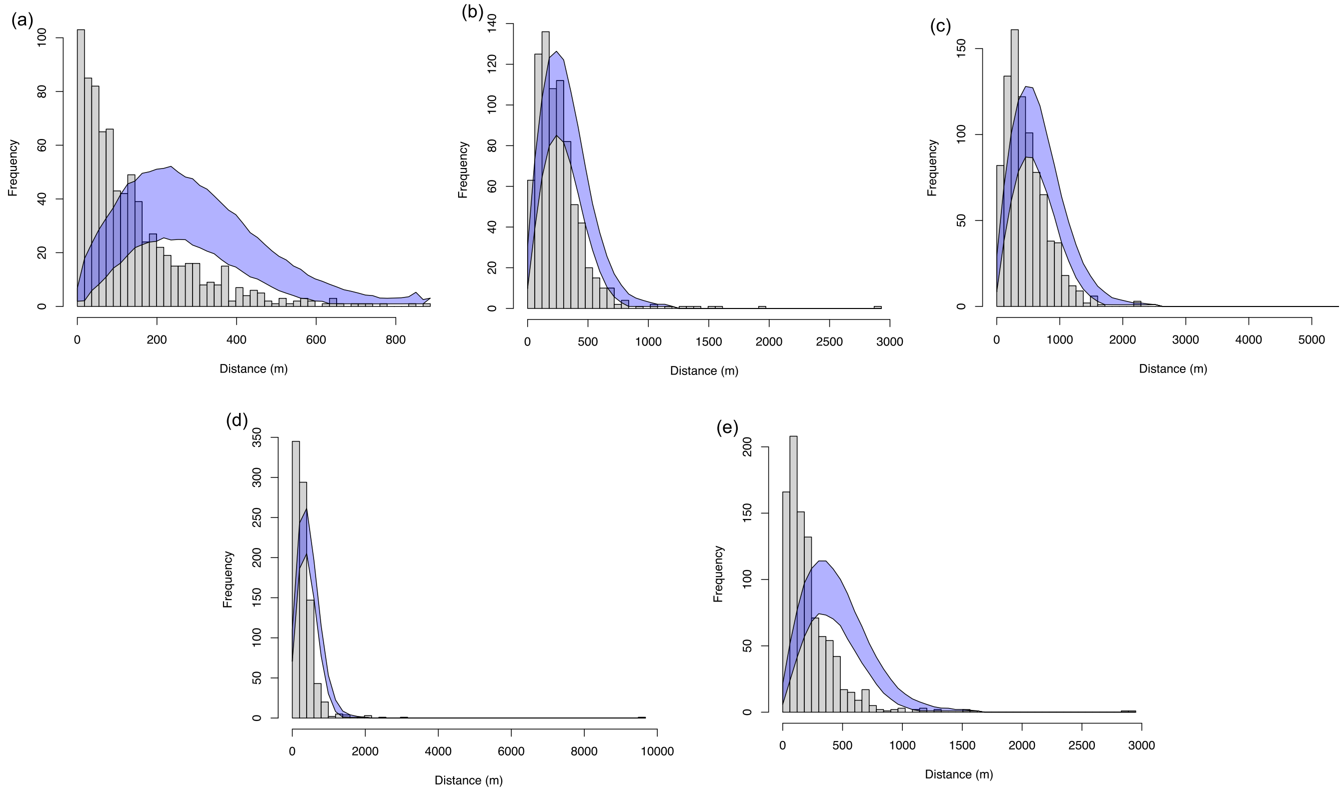

FIGURE 4. **Histograms of observed nearest neighbor distances, with envelopes of expected frequencies for (a) Medellin, (b) Milan, (c) Chicago, (d) Singapore, and (e) Mumbai.**

the total number of green recreation spaces, but rather the proximity of the nearest green recreation space to each node. This technique is based on the convention that accessibility measures are often linked to the number of amenities at specific places, such as POI numbers or other POI features [47].

IV. RESULTS

In this section, we will discuss the results obtained after testing the proposed framework in five global cities: Medellin, Milan, Chicago, Singapore, and Mumbai. The five global cities chosen for this study - Medellin, Milan, Chicago, Singapore, and Mumbai - were selected due to their diverse geographical locations, cultural contexts, and varying stages of urban development. This diversity allows us to test the applicability of our framework across different contexts and draw more generalizable conclusions. Medellin, a South American city noted for its modern urban planning projects, is included to provide unique insights into how green recreation spaces are dispersed and utilized in a developing urban setting. Milan, a European city known for its commitment to sustainability, represents a blend of historical and modern urban planning, serving as a model for green space distribution in a densely populated metropolitan setting. Chicago, a major North American city with a well-planned grid system, offers insight into green space accessibility in a densely populated and diverse metropolis. Singapore, a Southeast Asian city-state famed for its rigorous urban planning, exemplifies how a dense urban setting may provide universal access to green recreational spaces. Lastly, Mumbai, one of the world's most populous cities, presents a unique case study on the challenges of providing equitable access to green areas in a densely populated, fast-expanding urban environment. Each city's unique characteristics and levels of urbanization contribute to a comprehensive understanding of green space accessibility across different urban contexts. By selecting these diverse cities, we aimed to capture a broad range of urban landscapes and examine the distribution and accessibility of green recreation spaces across different contexts. Table 1 shows the cities and the number of green recreation spaces. This is based on the data till December 2022. Figure 3 shows the density of green recreation spaces based on their spatial distribution in different cities. The regions are divided into bins and the density of green recreation spaces is calculated in each bin. The intensity or concentration of green recreation spaces in various parts of the map is represented by these density maps. The maps show kernel density surfaces, with colored values representing a condensed representation of the spatial variance in the density of green recreation spaces across the study areas. Figure 3 shows that the distribution of green recreation places is not uniform across cities.

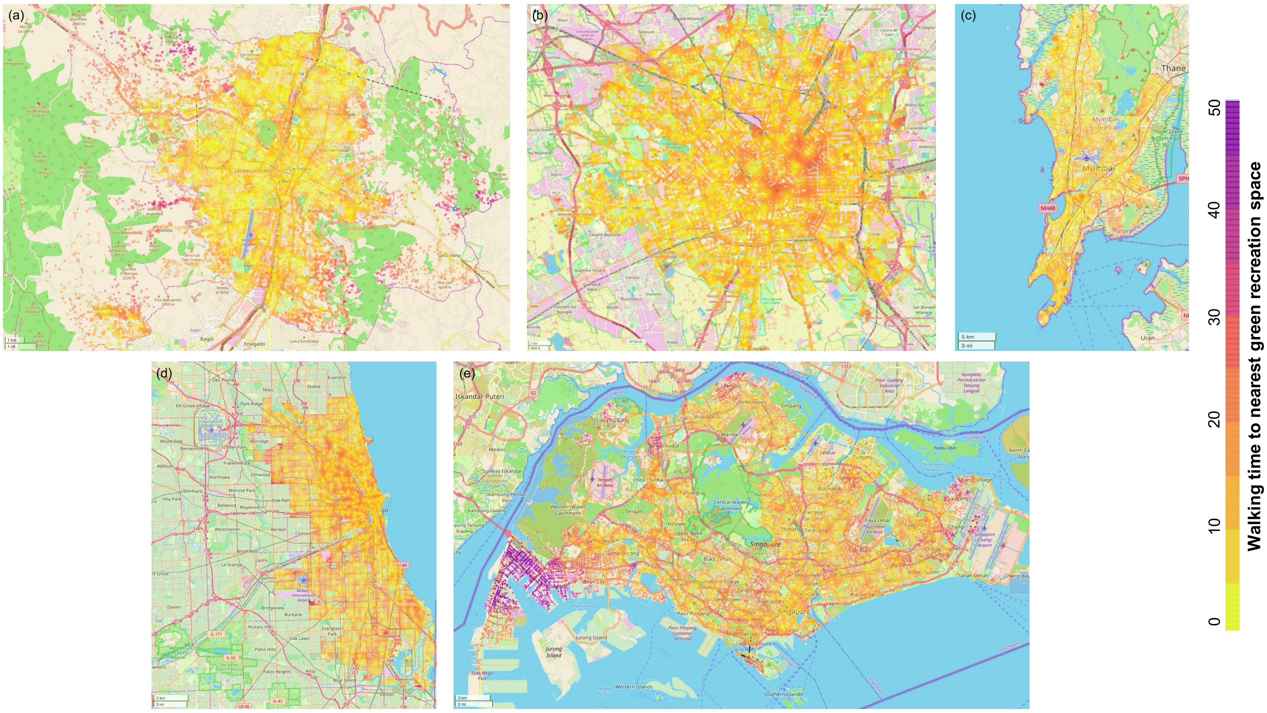

FIGURE 5. The heatmaps display the walking time required to reach the closest green recreation space for each of the five cities analyzed in this study, including (a) Medellin, (b) Milan, (c) Mumbai, (d) Chicago, and (e) Singapore.

The heat maps reveal distinct patterns in different cities. In Figure 3, the color intervals are determined by KDE. This calculates the density of green spaces at each point on the map, and the color of each area corresponds to this estimated density. The color scale is continuous and based on the range of these density levels, rather than being divided into quantiles. Medellin and Mumbai show notable clusters of green spaces in specific areas within the city. In comparison to the other cities, Milan has a less concentrated distribution of green spaces. In Chicago, there is a large clustering of green recreation places around the city center, whereas, in Singapore, significant clustering is evident in the city's eastern part. While the heat maps provide a visual understanding of the distribution of green recreation spaces, further statistical analysis is required to comprehensively understand the distribution patterns. Therefore, our next step involves conducting a point pattern analysis.

Following up on the previous analysis, we performed a point pattern analysis to get a more detailed picture of the distribution of green recreation spaces in the selected cities. Our primary focus was on investigating distribution patterns using the nearest-neighbor approach, as shown in Figure 4. To do this, we created histograms of the actual data's distances to the nearest neighbor, as well as an envelope of expected values acquired through simulation. This involved creating 1000 random point patterns with a Poisson distribution, the intensity of which was determined by the density function produced from the original data. Then, for each bar in the histogram, we generated the 95% confidence interval and superimposed it on the original histogram. This methodology draws inspiration from the work of Bevan et. al in [43]. The primary goal here is to establish if the observed distribution of nearest-neighbor distances matches our expectations or is completely random. This could be done using the Clark and Evans test [48] as it has been used in the past, but we used a more robust alternative. We generate random point patterns through Monte Carlo simulation [49] to simulate a comparable number of random points in the study areas. These simulated point patterns are used to establish an expected distribution and create an envelope of expected values for the nearest-neighbor distances. By comparing the observed distribution of nearest-neighbor distances to the simulated distribution, we can determine whether the observed pattern deviates significantly from randomness. There are three ways to analyze the relationship between the bars and the envelope (Figure 4).

1) If the bars are constantly within the envelope, it indicates that the observed distribution of nearest neighbor distances is consistent with a random pattern. In this scenario, the green recreation spaces are dispersed in the manner that would be expected if they were randomly placed within the study area.
2) If the bars continuously surpass the upper bounds of the envelope, this suggests a clustering pattern. In this case, it suggests that the green recreation spaces tend to be closer to each other than would be expected by

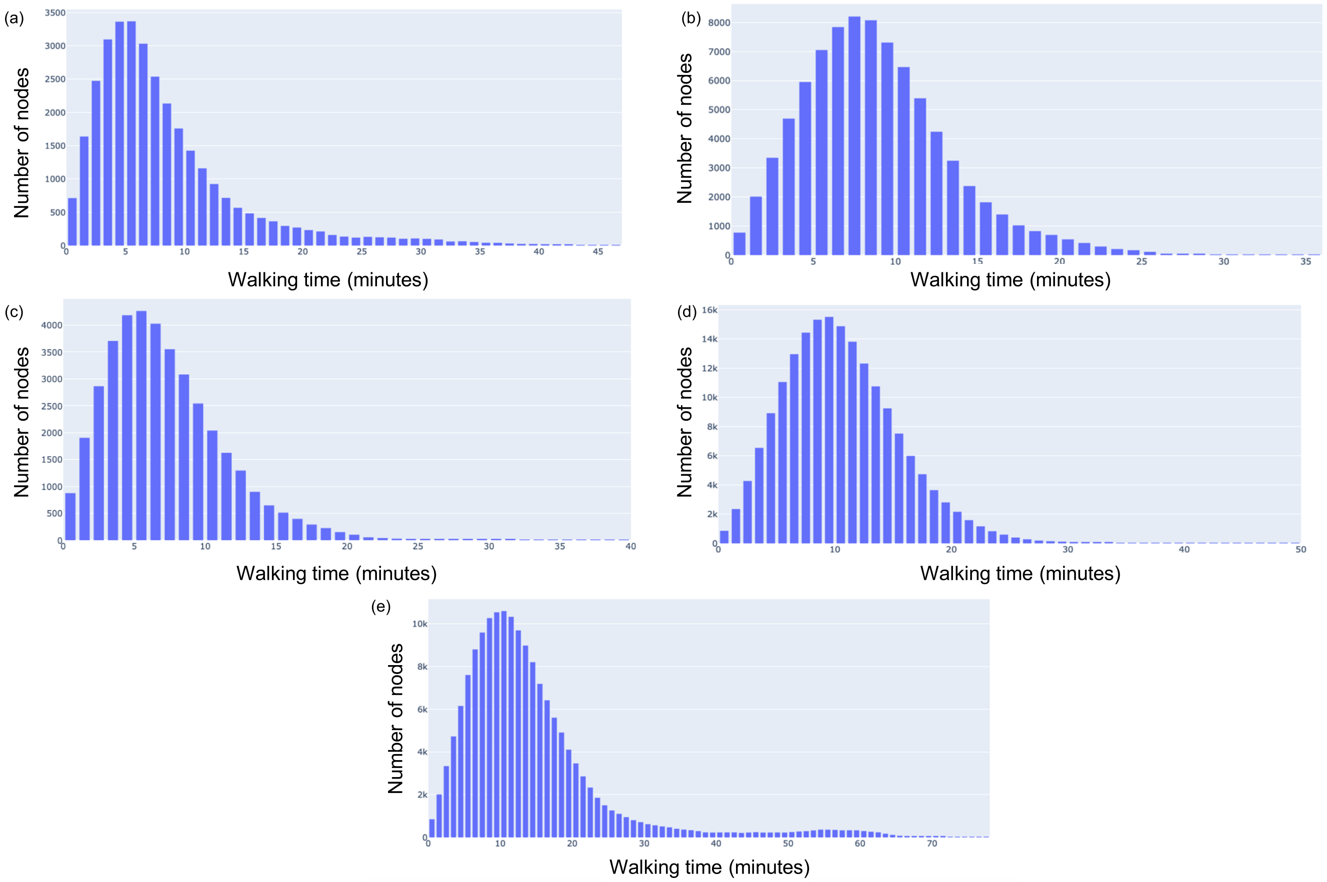

FIGURE 6. The histograms show the count of nodes against the walking time required to access the nearest green recreation spaces in: (a) Medellin, (b) Milan, (c) Mumbai, (d) Chicago, and (e) Singapore.

chance, indicating a non-random spatial organization with significant clustering.

3) If the bars continually fall below the envelope's lower bounds, this indicates a dispersed pattern. In this scenario, this indicates that the green recreation spaces are more equally spaced apart than would be expected by random chance, showing a non-random spatial organization with a dispersed distribution.

As observed in Figure 4, in the case of Medellin, variations from the expected nearest neighbor distances for a Poisson process with similar density are seen. The graph demonstrates a clustering pattern, with a larger concentration of green recreation spaces, than expected observed within 100 meters. Beyond around 200 meters, however, fewer spaces are concentrated, indicating a more dispersed distribution. For Milan, little clustering is observed within a distance of 150 meters, indicating a tendency for green recreation spaces to be located in closer proximity to each other within this range. Above this threshold, however, the observed distribution more closely resembles a random pattern, as indicated by the bars falling within the envelope. In the case of Chicago, a clustering pattern for green recreation spaces within a 300-meter distance is seen. After 500 meters, the distribution falls below the expected values, indicating a deviation from a random spatial layout. Similarly, for Singapore, clustering can be observed within approximately 400 meters, followed by a dispersed distribution of green recreation spaces beyond that point. Lastly, for Mumbai, a similar pattern emerges, with a strong clustering observed within 200 meters, followed by dispersed distribution of green recreation spaces. One takeaway from the analysis of all the cities is that a high frequency of green recreation spaces is clustered within an average distance of 200 meters (approximately). There are higher densities at shorter distances and there is a gradual decrease in density of green recreation spaces as the distance increases. There are some occasional fluctuations in counts that suggest the existence of distinct clusters or zones with varying densities.

While the nearest-neighbor analysis has provided valuable information about the distribution and clustering patterns of green recreation spaces within the study areas, understanding the proximity of these spaces alone is not enough in assessing their true accessibility. In the next step, we aimed to analyze the accessibility of green recreation spaces by calculating the shortest distance from the nearest node to the POI i.e. a green recreation space. The results of our analysis are presented in Figure 5, which shows heatmaps for each city that illustrate the walking time (in minutes) from all nodes in the pedestrian network to the nearest green recreation space. Additionally, Figure 6 presents a histogram for all five cities, showing the number of nodes, and walking time (in minutes) required to reach the nearest green recreation space. Our findings show that the average walking time to the nearest green space is 7.8 minutes for Medellin, 8.3 minutes for Milan, 6.7 minutes for Mumbai, 10 minutes for Chicago, and 14 minutes for Singapore. Despite the reported differences in walking time to the nearest green in these cities, it is critical to account for changes in node density and distribution within each city network. For example, Chicago, with almost 16k nodes, and Singapore, with more than 10k nodes, have a more vast and complicated network structure than Medellin, which has only 3.5k nodes, and Mumbai which has only 4k nodes. The accessibility and closeness of green recreation areas may be influenced by differences in network complexity.

V. DISCUSSION

A. THE IMPORTANCE OF GREEN SPACES IN SMART CITIES

The notion of smart cities is still developing, and many elements are taken into account while creating and implementing smart city programs. While access to nature and green spaces is crucial in urban development, it is not necessarily the major emphasis of smart city projects. There is a growing recognition of the importance of access to nature and green areas for human health and well-being, and this should be reflected in the development of smart city programs. Access to green recreation spaces has been shown to have numerous benefits for mental and physical health. Studies have shown that spending time in nature can reduce stress and improve overall well-being [50]. In addition, green space can help to improve air and water quality, reduce urban heat island effects [51], and provide habitat for urban wildlife [52]. Furthermore, green space can serve as an important public gathering space for communities. Parks, gardens, and other green spaces can provide a place for people to come together and socialize, promoting a sense of community and improving social cohesion. While there have been a lot of ongoing discussions about nature in the cities, potential benefits as well as frameworks to create more inclusive cities [53], [54], there is still a lot to do when it comes to access to green recreation spaces in cities. The uneven spatial distribution of green spaces in cities can have a significant impact on the city's sustainability, environment, and quality of life for city residents [55]. One of the major issues is a lack of precise data and sufficient tools for understanding the distribution and accessibility of these areas. Despite considerable advances in recent years, utilizing this data for informed decision-making remains difficult due to limited technological infrastructure and complexity. Addressing these challenges requires a coordinated effort to create and implement tools and technology that enable data-informed decision-making, not just for the decision-makers but also for city residents. Open-source tools and technology have demonstrated significant promise in facilitating collective intelligence and participatory resilience [56], [57]. These tools can aid in democratizing information availability and encouraging public engagement and feedback in decision-making processes.

B. INTEGRATING GREEN SPACE ACCESSIBILITY INTO SMART CITY DISCOURSE

Despite the growing body of literature on green space accessibility, its integration into the smart city discourse remains limited. Smart city initiatives often prioritize technological solutions, such as sensor networks, data analytics, and intelligent transportation systems, while overlooking the importance of equitable access to green spaces [58], [59]. The existing research gaps highlight the critical need for a comprehensive understanding of accessibility within the context of smart cities, as well as innovative techniques to enhance the availability and distribution of green recreation spaces. Furthermore, it is critical to use digital infrastructure to develop user-friendly frameworks and tools that allow decision-makers as well as other stakeholders to make informed decisions about urban planning initiatives. If done the right way, smart cities have the ability to seamlessly integrate technology, community, and nature, and to create equitable access to resources while increasing communal well-being.

C. DATA LIMITATIONS

While OSM has shown to be a great resource for understanding urban landscapes, is also important to acknowledge the limitations of using OSM data. Its reliance on volunteer contributions may result in data gaps or discrepancies, especially in areas with low levels of community engagement or technical expertise. OSM data may suffer from coverage and data quality biases. Certain geographic areas, particularly those with a more active OSM community or higher technical expertise, might have better representation compared to others [60]. This could potentially lead to inaccuracies or incomplete information in the analysis. OSM data is continuously updated by volunteers, which means the data's temporal variability could impact the analysis. For instance, new green spaces might be added or existing ones removed after the data was collected, which could affect the results. Also, the lack of a formal validation process for OSM data means that some data might be inaccurate or outdated. This could potentially impact the reliability of the findings.

Despite these limitations, we believe that OSM remains a valuable resource for urban data analysis. Its open-source nature, wide coverage, and vast amount of geospatial information it provides make it a useful tool for understanding the spatial distribution and accessibility of green recreation spaces. To further enhance the comprehensiveness and accuracy of OSM data, other data sources can be integrated with it. These include street view data [61] for detailed visual information, remote sensing data for land use and vegetation cover, as well as government datasets for accurate records.

Each of these sources can complement OSM data, addressing some of its limitations while providing a richer and more nuanced understanding of green space accessibility.

VI. CONCLUSION AND FUTURE WORK

In this study, we proposed an open-source and extendable framework that gives statistical insights and visualizations of the distribution and accessibility of green recreation places in cities. The proposed framework was applied to five global cities: Medellin, Milan, Chicago, Singapore, and Mumbai, to analyze the distribution of green recreation spaces and assess their pedestrian accessibility. While the cities represented different geographical and cultural settings, we found that for all of them, most of the green recreation spaces were clustered in small areas resulting in uneven distribution. In terms of accessibility, we found that the walking time to the nearest green space was highest in Chicago and Singapore. Given that our framework utilizes OSM data, it offers adaptability and simplifies the process of testing it for different amenities and locations. Furthermore, the open-source structure allows collaboration and encourages community participation in the tool's development and enhancement. The flexibility for developers and researchers to add new functionality and features to the existing code base considerably saves the time and effort required to build equivalent tools from the ground up.

In terms of future work, we envision expanding our framework to further support the development of smart cities that prioritize accessibility to nature. Specifically, we aim to incorporate predictive modeling capabilities into our framework. This enhancement would allow for the exploration of various urban planning scenarios and the assessment of potential outcomes of different policy interventions, providing a dynamic tool for the strategic development of green recreation spaces.

ACKNOWLEDGMENT

(Jenny Martinez and Sachit Mahajan contributed equally to this work.)

REFERENCES

[1] *68% of the World Population Projected to Live in Urban Areas by 2050, Says UN*, UN, New York, NY, USA, May 2018.

[2] M. Mayer, ''The 'right to the city' in the context of shifting mottos of urban social movements,'' *City*, vol. 13, nos. 2–3, pp. 362–374, Jun. 2009.

[3] S. L. Harlan, A. J. Brazel, L. Prashad, W. L. Stefanov, and L. Larsen, ''Neighborhood microclimates and vulnerability to heat stress,'' *Social Sci. Med.*, vol. 63, no. 11, pp. 2847–2863, Dec. 2006.

[4] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, ''Urban computing: Concepts, methodologies, and applications,'' *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 1–55, Sep. 2014.

[5] S. Mahajan, J. Gabrys, and J. Armitage, ''AirKit: A citizen-sensing toolkit for monitoring air quality,'' *Sensors*, vol. 21, no. 12, p. 4044, Jun. 2021.

[6] S. Mahajan, C.-H. Luo, D.-Y. Wu, and L.-J. Chen, ''From do-it-yourself (DIY) to do-it-together (DIT): Reflections on designing a citizen-driven air quality monitoring framework in Taiwan,'' *Sustain. Cities Soc.*, vol. 66, Mar. 2021, Art. no. 102628.

[7] V. Albino, U. Berardi, and R. M. Dangelico, ''Smart cities: Definitions, dimensions, performance, and initiatives,'' *J. Urban Technol.*, vol. 22, no. 1, pp. 3–21, Jan. 2015.

[8] E. Talen and L. Anselin, ''Assessing spatial equity: An evaluation of measures of accessibility to public playgrounds,'' *Environ. Planning A, Economy Space*, vol. 30, no. 4, pp. 595–613, Apr. 1998.

[9] J. Maas, ''Green space, urbanity, and health: How strong is the relation?'' *J. Epidemiology Community Health*, vol. 60, no. 7, pp. 587–592, Jul. 2006.

[10] A. Chiesura, ''The role of urban parks for the sustainable city,'' *Landscape Urban Planning*, vol. 68, no. 1, pp. 129–138, May 2004.

[11] B. Chen, S. Wu, Y. Song, C. Webster, B. Xu, and P. Gong, ''Contrasting inequality in human exposure to greenspace between cities of global north and global south,'' *Nature Commun.*, vol. 13, no. 1, p. 4636, Aug. 2022.

[12] J. R. Wolch, J. Byrne, and J. P. Newell, ''Urban green space, public health, and environmental justice: The challenge of making cities 'just green enough,''' *Landscape Urban Planning*, vol. 125, pp. 234–244, May 2014.

[13] C. Wan, G. Q. Shen, and S. Choi, ''Underlying relationships between public urban green spaces and social cohesion: A systematic literature review,'' *City, Culture Soc.*, vol. 24, Mar. 2021, Art. no. 100383.

[14] A. C. K. Lee and R. Maheswaran, ''The health benefits of urban green spaces: A review of the evidence,'' *J. Public Health*, vol. 33, no. 2, pp. 212–222, Jun. 2011.

[15] P. J. Villeneuve, M. Jerrett, J. G. Su, R. T. Burnett, H. Chen, A. J. Wheeler, and M. S. Goldberg, ''A cohort study relating urban green space with mortality in Ontario, Canada,'' *Environ. Res.*, vol. 115, pp. 51–58, May 2012.

[16] S. Wan, D. Rojas-Rueda, J. Pretty, C. Roscoe, P. James, and J. S. Ji, ''Greenspace and mortality in the U.K. biobank: Longitudinal cohort analysis of socio-economic, environmental, and biomarker pathways,'' *SSM, Population Health*, vol. 19, Sep. 2022, Art. no. 101194.

[17] F. Li, R. Wang, J. Paulussen, and X. Liu, ''Comprehensive concept planning of urban greening based on ecological principles: A case study in Beijing, China,'' *Landscape Urban Planning*, vol. 72, no. 4, pp. 325–336, May 2005.

[18] *Smarter Cities for a Better Future*, NTU, Singapore, Aug. 2022.

[19] F. Li, A. Nucciarelli, S. Roden, and G. Graham, ''How smart cities transform operations models: A new research agenda for operations management in the digital economy,'' *Prod. Planning Control*, vol. 27, no. 6, pp. 514–528, Apr. 2016.

[20] Y.-L. Lin, M.-F. Yen, and L.-C. Yu, ''Grid-based crime prediction using geographical features,'' *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 8, p. 298, Jul. 2018.

[21] K. Hopf, ''Mining volunteered geographic information for predictive energy data analytics,'' *Energy Informat.*, vol. 1, no. 1, pp. 1–21, Dec. 2018.

[22] X. Ding, H. Fan, and J. Gong, ''Towards generating network of bikeways from mapillary data,'' *Comput., Environ. Urban Syst.*, vol. 88, Jul. 2021, Art. no. 101632.

[23] D. Liu, M.-P. Kwan, and Z. Kan, ''Analysis of urban green space accessibility and distribution inequity in the city of Chicago,'' *Urban Forestry Urban Greening*, vol. 59, Apr. 2021, Art. no. 127029.

[24] J. Yang, C. Li, Y. Li, J. Xi, Q. Ge, and X. Li, ''Urban green space, uneven development and accessibility: A case of Dalian's Xigang district,'' *Chin. Geograph. Sci.*, vol. 25, no. 5, pp. 644–656, Oct. 2015.

[25] L. Nicoletti, M. Sirenko, and T. Verma, ''Disadvantaged communities have lower access to urban infrastructure,'' *Environ. Planning B, Urban Anal. City Sci.*, vol. 50, no. 3, pp. 831–849, Mar. 2023.

[26] H. E. Wright Wendel, R. K. Zarger, and J. R. Mihelcic, ''Accessibility and usability: Green space preferences, perceptions, and barriers in a rapidly urbanizing city in Latin America,'' *Landscape Urban Planning*, vol. 107, no. 3, pp. 272–282, Sep. 2012.

[27] W. Luo and F. Wang, ''Measures of spatial accessibility to health care in a GIS environment: Synthesis and a case study in the Chicago region,'' *Environ. Planning B, Planning Design*, vol. 30, no. 6, pp. 865–884, Dec. 2003.

[28] J. Wolch, J. P. Wilson, and J. Fehrenbach, ''Parks and park funding in Los Angeles: An equity-mapping analysis,'' *Urban Geogr.*, vol. 26, no. 1, pp. 4–35, Feb. 2005.

[29] A. Rigolon, ''A complex landscape of inequity in access to urban parks: A literature review,'' *Landscape Urban Planning*, vol. 153, pp. 160–169, Sep. 2016.

[30] L.-J. Chen, Y.-H. Ho, H.-C. Lee, H.-C. Wu, H.-M. Liu, H.-H. Hsieh, Y.-T. Huang, and S. C. Lung, ''An open framework for participatory $PM_{2.5}$ monitoring in smart cities,'' *IEEE Access*, vol. 5, pp. 14441–14454, 2017.

[31] P. Hamel et al., ''Mapping the benefits of nature in cities with the invest software,'' *NPJ Urban Sustainability*, vol. 1, no. 1, p. 25, Jun. 2021.

[32] S. Mahajan, W.-L. Wu, T.-C. Tsai, and L.-J. Chen, ''Design and implementation of IoT-enabled personal air quality assistant on instant messenger,'' in *Proc. 10th Int. Conf. Manage. Digit. EcoSyst.*, Sep. 2018, pp. 165–170.
[33] S. Mahajan, ''Design and development of an open-source framework for citizen-centric environmental monitoring and data analysis,'' *Sci. Rep.*, vol. 12, no. 1, p. 14416, Aug. 2022.
[34] T. Dogan, Y. Yang, S. Samaranayake, and N. Saraf, ''Urbano: A tool to promote active mobility modeling and amenity analysis in urban design,'' *Technol., Archit. Des.*, vol. 4, no. 1, pp. 92–105, Jan. 2020.
[35] S. Law, B. Paige, and C. Russell, ''Take a look around: Using street view and satellite images to estimate house prices,'' *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 5, pp. 1–19, Sep. 2019.
[36] S. Mahajan and J. Martinez, ''Water, water, but not everywhere: Analysis of shrinking water bodies using open access satellite data,'' *Int. J. Sustain. Develop. World Ecology*, vol. 28, no. 4, pp. 326–338, May 2021.
[37] Y. Liu, F. Wang, Y. Xiao, and S. Gao, ''Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai,'' *Landscape Urban Planning*, vol. 106, no. 1, pp. 73–87, May 2012.
[38] G. Boeing, ''OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks,'' *Comput., Environ. Urban Syst.*, vol. 65, pp. 126–139, Sep. 2017.
[39] C. C. Weiss, M. Purciel, M. Bader, J. W. Quinn, G. Lovasi, K. M. Neckerman, and A. G. Rundle, ''Reconsidering access: Park facilities and neighborhood disamenities in New York City,'' *J. Urban Health*, vol. 88, no. 2, pp. 297–310, Apr. 2011.
[40] A. C. Gatrell, T. C. Bailey, P. J. Diggle, and B. S. Rowlingson, ''Spatial point pattern analysis and its application in geographical epidemiology,'' *Trans. Inst. Brit. Geographers*, vol. 21, no. 1, pp. 256–274, 1996.
[41] T. Wiegand and K. A. Moloney, *Handbook of Spatial Point-Pattern Analysis in Ecology*. Boca Raton, FL, USA: CRC Press, 2013.
[42] P. Haase, ''Spatial pattern analysis in ecology based on Ripley's K-function: Introduction and methods of edge correction,'' *J. Vegetation Sci.*, vol. 6, no. 4, pp. 575–582, Aug. 1995.
[43] A. Bevan, E. Jobbová, C. Helmke, and J. J. Awe, ''Directional layouts in central lowland maya settlement,'' *J. Archaeol. Sci.*, vol. 40, no. 5, pp. 2373–2383, May 2013.
[44] G. L. W. Perry, B. P. Miller, and N. J. Enright, ''A comparison of methods for the statistical analysis of spatial point patterns in plant ecology,'' *Plant Ecol.*, vol. 187, no. 1, pp. 59–82, Oct. 2006.
[45] A. de Araujo, J. Marcos do Valle, and N. Cacho, ''Geographic feature engineering with points-of-interest from OpenStreetMap,'' in *Proc. KDIR*, 2020, pp. 116–123.
[46] A. Psyllidis, S. Gao, Y. Hu, E.-K. Kim, G. McKenzie, R. Purves, M. Yuan, and C. Andris, ''Points of interest (POI): A commentary on the state of the art, challenges, and prospects for the future,'' *Comput. Urban Sci.*, vol. 2, no. 1, p. 20, Jun. 2022.
[47] W. G. Hansen, ''How accessibility shapes land use,'' *J. Amer. Inst. Planners*, vol. 25, no. 2, pp. 73–76, May 1959.
[48] B. D. Ripley, ''Tests of 'randomness' for spatial point patterns,'' *J. Roy. Stat. Soc. B, Methodol.*, vol. 41, no. 3, pp. 368–374, 1979.
[49] C. P. Robert, G. Casella, and G. Casella, *Monte Carlo Statistical Methods*, vol. 2. New York, NY, USA: Springer, 1999.
[50] J. Roe, C. Thompson, P. Aspinall, M. Brewer, E. Duff, D. Miller, R. Mitchell, and A. Clow, ''Green space and stress: Evidence from cortisol measures in deprived urban communities,'' *Int. J. Environ. Res. Public Health*, vol. 10, no. 9, pp. 4086–4103, Sep. 2013.
[51] K. Van Ryswyk, N. Prince, M. Ahmed, E. Brisson, J. D. Miller, and P. J. Villeneuve, ''Does urban vegetation reduce temperature and air pollution concentrations? Findings from an environmental monitoring study of the central experimental farm in Ottawa, Canada,'' *Atmos. Environ.*, vol. 218, Dec. 2019, Art. no. 116886.
[52] M. B. Haverland and J. A. Veech, ''Examining the occurrence of mammal species in natural areas within a rapidly urbanizing region of Texas, USA,'' *Landscape Urban Planning*, vol. 157, pp. 221–230, Jan. 2017.
[53] Q. Liu, H. Ullah, W. Wan, Z. Peng, L. Hou, T. Qu, and S. Ali Haidery, ''Analysis of green spaces by utilizing big data to support smart cities and environment: A case study about the city center of Shanghai,'' *ISPRS Int. J. Geo-Inf.*, vol. 9, no. 6, p. 360, Jun. 2020.
[54] D. D'Alessandro, M. Buffoli, L. Capasso, G. Fara, A. Rebecchi, and S. Capolongo, ''Green areas and public health: Improving wellbeing and physical activity in the urban context,'' *Epidemiologia Prevenzione*, vol. 39, no. 5, pp. 8–13, 2015.
[55] M. Triguero-Mas, P. Dadvand, M. Cirach, D. Martínez, A. Medina, A. Mompart, X. Basagaña, R. Gražulevičienė, and M. J. Nieuwenhuijsen, ''Natural outdoor environments and mental and physical health: Relationships and mechanisms,'' *Environ. Int.*, vol. 77, pp. 35–41, Apr. 2015.
[56] T. W. Malone and M. Klein, ''Harnessing collective intelligence to address global climate change,'' *Innov., Technol., Governance, Globalization*, vol. 2, no. 3, pp. 15–26, Aug. 2007.
[57] S. Mahajan, C. I. Hausladen, J. A. Sánchez-Vaquerizo, M. Korecki, and D. Helbing, ''Participatory resilience: Surviving, recovering and improving together,'' *Sustain. Cities Soc.*, vol. 83, Aug. 2022, Art. no. 103942.
[58] R. Kitchin, ''The real-time city? Big data and smart urbanism,'' *GeoJournal*, vol. 79, no. 1, pp. 1–14, Feb. 2014.
[59] M. Angelidou, ''The role of smart city characteristics in the plans of fifteen cities,'' *J. Urban Technol.*, vol. 24, no. 4, pp. 3–28, Oct. 2017.
[60] F. Biljecki, Y. S. Chow, and K. Lee, ''Quality of crowdsourced geospatial building information: A global assessment of OpenStreetMap attributes,'' *Building Environ.*, vol. 237, Jun. 2023, Art. no. 110295.
[61] Y. Zhang, P. Liu, and F. Biljecki, ''Knowledge and topology: A two layer spatially dependent graph neural networks to identify urban functions with time-series street view image,'' *ISPRS J. Photogramm. Remote Sens.*, vol. 198, pp. 153–168, Apr. 2023.

JENNY MARTINEZ received the M.E. degree in systems engineering from Universidad Nacional de Colombia, in 2013, where she is currently pursuing the Ph.D. degree. Her current research interests include data science applied to the environment and urban areas.

SACHIT MAHAJAN received the Ph.D. degree in social networks and human-centered computing. He is currently a Postdoctoral Researcher with the Professorship of Computational Social Science, ETH Zürich. His research interests include the intersection of ubiquitous systems, data science, and citizen science.