

ARTICLE OPEN



Exploring residential built-up form typologies in Delhi: a grid-based clustering approach towards sustainable urbanisation

Aviral Marwal¹ and Elisabete A. Silva¹

Previous studies have established a significant link between urban form and sustainability. However, the diversity of micro-scale urban forms in cities in the global south has received limited attention, hindered by the lack of neighbourhood-level spatial data and maps, which poses challenges in exploring micro-urban form features. The study addresses this gap using a grid-based k-means clustering algorithm to identify residential built-up form typologies in Delhi and assess their impact on sustainable urbanisation. The algorithm clusters 100×100 metre grid cells based on their attributes of accessibility, built-up density, and street design. The results show six distinct built-up form typologies in Delhi. However, only 19% of residential areas meet the criteria for sustainable urbanisation, highlighting the need for planning interventions in most areas. The study methodology can be applied to analyse micro-scale urban form features in other cities in the global south, providing a fresh perspective on urbanisation research.

npj Urban Sustainability (2023)3:40; <https://doi.org/10.1038/s42949-023-00112-1>

INTRODUCTION

Cities have significantly impacted human and environmental well-being throughout modern civilisation. Currently, more than half of the global population, which is 4.4 billion people, resides in urban areas, and this is projected to rise to 68% by 2050. According to the UN DESA report¹, Delhi is currently the second most populous urban agglomeration (UA) globally and is anticipated to surpass all other UAs in terms of population with an estimated 37.2 million individuals by 2028. As the world becomes more urban, human interaction with a city's built-up environment is bound to increase, and thus important to study. The built-up of a city, referred to as the urban form, provides an objective tool to understand this human-city relationship². Previous studies have emphasised the significance of comprehending urban form as a fundamental component of urban sustainability³. Research shows that urban form influences a city's land use pattern and has a widespread impact on residents' lifestyle choices⁴ and the urban environment, such as residential location and commuting⁵, social well-being⁶, environmental well-being^{7,8}, and energy use⁹. Thus, a more comprehensive understanding of urbanisation can be achieved by characterising it through urban form, which is today largely driven by population density and the nature of employment¹⁰.

While urban morphology has been a subject of academic enquiry for a considerable period, it has regained emphasis since the 1990s with the advancement of geographical information systems (GIS) and remote sensing². Over the last 20 years, studies have used different classification methods to categorise the morphological elements of urban form at different spatial scales, such as neighbourhoods and cities, applying both quantitative and qualitative categorisation tools¹¹. While these studies enrich our understanding of the relationships between different urban configurations and their impact on urban sustainability, the bulk of our comprehension regarding urban morphology has come from the cities in the global north. However, in recent years, a discernible shift towards understanding urban morphology in cities in the global south has been noticed¹².

In recent decades, cities in developing countries have seen a very high influx of migrants, leading to changes in land use

pattern¹³ and the mushrooming of residential settlements, many of which are often in unregularised neighbourhoods and can be termed illegal settlements or slums¹⁴. In cities with such diverse settlement patterns, neighbourhoods differ not only in terms of their socio-economic indicators but also in their built-up structure¹⁵. In such cases, characterising a city with a particular urban form can be misleading. Thus, to study the extent of sustainable urbanisation in such cities, one needs to explore the variation in neighbourhood built-up types. Although a topic of great importance, few studies have analysed variations in urban form within a city¹⁶.

In this context, this study raises some important questions: Do residential areas in Delhi have diverse built-up forms? If so, how can we visualise and measure them? Moreover, what impact does the built-up form have on sustainable urbanisation in Delhi? In this regard, the study has two objectives: first, to cluster the neighbourhoods in Delhi using the k-means clustering algorithm and characterise them with their dominant built-up form typology; second, to analyse how these different built-up form typologies affect sustainable urbanisation. It is important to note that, as the study aims to explore residential built-up form typologies, it considers only the physical aspects of residential areas. Other aspects of residential areas related to socioeconomic and demographic features have not been explicitly addressed in this study.

In the context of urban morphology, to the best of our knowledge, this study is one of the earliest to examine and map the variations in the built-up form of residential areas in Delhi. The study holds significance on two grounds. First, the study uses the grid-based clustering method, which provides a methodological tool for urban planners to delineate the different built-up forms in a city in a more dynamic and adaptive manner, in contrast to relying on administratively defined boundaries. This provides more flexibility in mapping, as it can easily adapt to changes in a city's physical layout and demographics. The method also offers greater precision as it considers the actual physical layout of the city in the mapping¹⁷. With flexibility and precision, this method also offers consistency in mapping residential areas across

¹Interdisciplinary Lab of Spatial Analysis, Department of Land Economy, University of Cambridge, Cambridge, UK. ✉email: am2839@cam.ac.uk

different cities and regions, which makes it generalisable and thus significant for cities in the global south that lack micro-scale spatial maps.

Second, and more importantly, by addressing the sustainable cities paradigm, as specified under sustainable development goal 11: sustainable cities and communities, the study provides another perspective to study urbanisation through the lens of urban form. While urbanisation is seen as a synonym for economic growth, if poorly planned, it can have adverse implications for individual and environmental well-being¹⁸. In this context, we argue that the study of urbanisation should include elements of urban morphology¹⁹. By doing so, we can have a more informed understanding of how future urbanisation will impact the neighbourhood's living and what planning interventions can be made to achieve sustainable urbanisation. This can help ensure that urbanisation leads to sustainable cities rather than just economic growth.

We now provide a brief review of the literature and highlight the important gaps. Our examination of existing literature on urban form reveals three primary issues. First, studies conducting systematic exploration to capture the heterogeneity of spatial patterns at the neighbourhood level are found to be limited²⁰. While previous studies have analysed cities based on their dominant urban forms^{21,22}, few have applied quantitative methods to study the variation in urban forms within a city^{16,23}. Cities, especially in developing countries, have diverse settlement patterns that result due to inadequate zoning laws and weak regulations^{24,25}. Unplanned urbanisation can result in the proliferation of urban sprawl, slums, and unauthorised colonies in a city, which has distinct urban forms compared to more affluent areas²⁶. Thus, it is crucial to understand the possible urban form typologies within a city for effective localised land use planning^{27,28}.

A few recent studies have investigated urban form typologies at the neighbourhood level. For example, Braulio et al.²⁹ developed a taxonomy of the city of Castellón de la Plana, Spain, using elements of residential buildings, and analysed variations in urban form patterns at different geographical scales. Lu et al.³⁰ measured urban form in different neighbourhoods of Chengdu City, China, using indicators of density, accessibility, shape, and diversity. Fleischmann et al.²⁰ developed a numerical taxonomy for urban form to classify urban types using street networks and building footprints, which they applied to generate a hierarchical classification of urban form in Prague and Amsterdam. On similar lines, Fusco et al.³¹ built a taxonomy of contemporary urban forms in France using indicators of street design and building types. With limited studies analysing urban form in the global south, our study contributes to the growing body of literature on built-up form typologies at the neighbourhood level in cities in the global south.

The second issue concerns the geographical scale for measuring urban form, an inappropriate selection of which can lead to the modified area unit problem (MAUP). The Modified Area Unit Problem (MAUP) refers to the phenomenon wherein the results of statistical analysis vary based on the scale or size of the geographic units used³². It is a common issue in spatial analysis, where geographical data can be aggregated into different levels of spatial resolution, such as census tracts, counties, states, or countries. Studies find that based on the choice of spatial resolution, results can vary³³.

To minimise the MAUP, data should be aggregated at the most appropriate spatial scale as per the research objective. In urban micro-planning, residential blocks/ neighbourhoods can serve as a unit of spatial analysis. However, if block-level spatial maps are unavailable, a grid cell approach can be used to define the study area^{34,35}. Under this approach, the geographic space is divided into a mesh of identically sized cells that are commonly square shaped, known as a grid. Each cell contains a numerical value that represents a specific geographic attribute, such as density or

elevation, for that unit of space³⁶. One of the key advantages of the grid cell approach is that it allows for a high degree of precision and granularity in spatial analysis¹⁷. By breaking down a geographic area into small cells, researchers can identify subtle spatial patterns and relationships that might be missed with other methods³⁷. Aggregating data under grid cells has been found to yield better results than aggregating data under administrative boundaries^{38,39}. Considering the lack of administratively defined neighbourhood boundary maps in Delhi and the advantages of the grid cell approach in minimising MAUP, this study relies on the grid cell approach as a spatial unit of analysis.

Finally, there is a challenge in clustering spatial data in urban form analysis. Clustering techniques in spatial science refer to a group of methods used to identify and group spatially related data points or objects. Clustering involves grouping objects based on their similarity in terms of geographic attributes such as distance, spatial density, or other spatially relevant features⁴⁰. The basic idea behind clustering techniques is to divide a dataset into subsets or clusters, such that the objects within each cluster are more similar to each other than to objects in other clusters. This allows researchers to identify spatial patterns or groupings in the data that may not be immediately apparent through visual inspection.

Recent studies on urban morphology have applied various clustering techniques to group spatial units with similar attributes and to identify dominant urban form typologies. These techniques include k-means, hierarchical agglomerative clustering^{41–43}, density-based clustering⁴⁴, Bayesian clustering⁴⁵, Gaussian mixture model⁴⁶, and spatial clustering methods like local indicators of spatial association (LISA) and local indicators of network-constrained clusters (ILINCS)⁴⁷. Advanced methods like self-organising maps, which combine statistical and machine learning methods, are also being used^{2,48}.

Although the study used different clustering algorithms, k-means was found to be more suitable in terms of model construction and execution. The model uses few input parameters and can be executed using different open-source software^{49,50}. The model results are also easy to interpret compared with hierarchical clustering⁵¹. We also noticed the model requires less computational time and is more efficient in processing large multivariate datasets in comparison to density-based and agglomerative clustering^{52,53}. More importantly, we found that among different clustering techniques, k-means clustering has been widely used in recent studies^{19,54–57}. Thus, owing to its widespread applicability, it made it easier for us to connect our model results with those of studies that have employed similar clustering techniques. However, one of the limitations of the K-means algorithm is that it requires the number of clusters to be specified in advance, which can be difficult if the data does not have a clear structure⁵². Incorrectly specifying the number of clusters in k-means clustering can produce oversimplified or overcomplicated results. A low number of clusters may miss important distinctions between data points and result in the loss of information, whereas an excessive number of clusters may lead to meaningless clusters, obscuring the underlying structure of the data⁵⁸. To overcome this limitation, this study uses a cluster optimisation method, which is discussed in the methodology section.

This section concludes by highlighting the challenge of interpreting the cluster results. Although machine learning (ML)-based clustering algorithms can cluster multivariable big datasets, they may not provide a clear understanding of how to interpret the cluster results⁵⁹. With the increase in the usage of machine learning models in data analysis, there has been growing concern about how to efficiently interpret the results of ML models⁶⁰. This becomes particularly challenging when multiple features are significant predictors of a cluster, making it difficult to determine how a particular feature affects cluster prediction⁶¹. ML-based

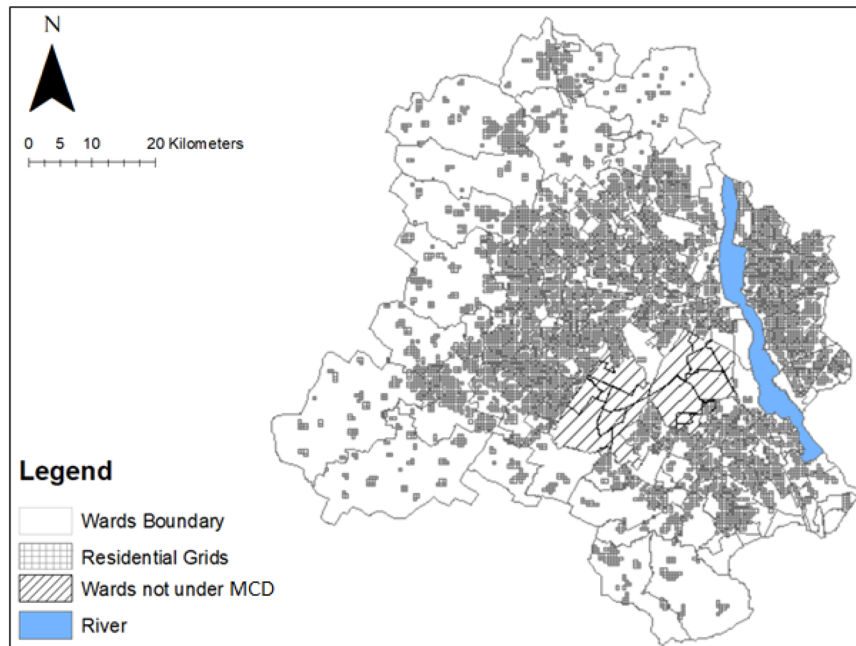


Fig. 1 Map of Delhi with residential areas shown under 100×100 m grid size. Source—Author, mapped using ArcGIS and Google Earth Software. The grid cells are 100×100 m in size and represent the residential areas, which come under the wards managed by the municipal corporation of Delhi (MCD). The residential areas in the city's outer areas are spatially non-contiguous as the majority of land use in these locations is under cropland or fallow land.

clustering algorithms typically employ unsupervised learning techniques, meaning that there is no predetermined outcome. As a result, the interpretation of the results is subjective and relies heavily on analysts' understanding of the data and research question. Thus, unfamiliarity with the working of ML models or inadequate understanding of the study context can lead users to wrongly interpret the model results, which can affect the study findings⁶².

To overcome this issue, a recent development has been the rise of explanatory methods, such as SHAP, to enhance the interpretability and transparency of ML models⁶³. SHAP (SHAPley Additive exPlanations) is a method that explains the output of a machine learning model by computing the contribution of each feature to the final prediction by averaging the marginal contribution of each feature over all possible coalitions of features⁶⁴. The SHAP method is effective in providing insights into the inner workings of complex machine learning models⁶⁵. In the last few years, the use of the SHAP tool has been seen in studies from different disciplines, such as health⁶⁶, engineering⁶⁷, and finance⁶⁸, which shows the growing acceptance of the SHAP tool in interpreting cluster results.

The remainder of the paper is organised as follows: The study results are presented in the next section. After that the discussion section discusses the study results and their implications for promoting sustainable urbanisation in Delhi. At last, the methods section provides the data preparation and research methods.

RESULTS

Study Context

Delhi, the capital of India, is the second most populous city globally, with over 28 million residents. It is expected to become the densest city in the world by 2030¹. Spread over 1483 sq. km, Delhi is divided into 11 districts and 250 wards that come under the Municipal Corporation of Delhi (MCD), with almost 3000 residential colonies. Figure 1 shows the map of Delhi with residential areas under the 100×100 metre grid size. The city is a hub for social, economic, and cultural activities, attracting

migrants from across the country and abroad. In the last decade, Delhi's built-up area has almost doubled, making housing and transportation management challenging for the government⁶⁹. According to Delhi's 2018–19 socio-economic survey, 85% of the population requires affordable housing options, 11% live in slums, and 60% of households (size of 5) are congested.

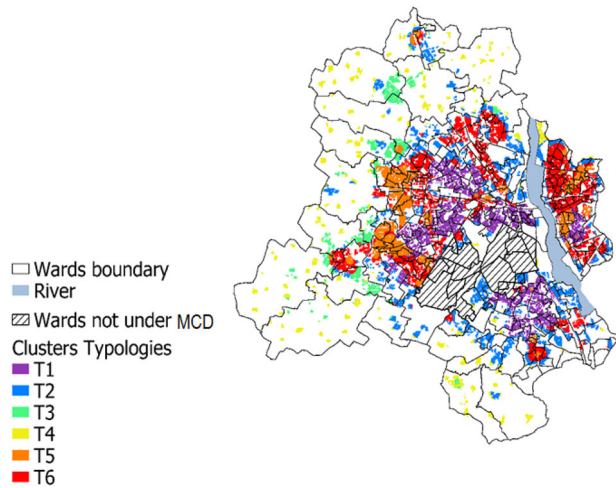
The 2011 Census estimates that Delhi will need 34.5 lakh dwelling units by 2041. As per the Delhi Development Authority⁷⁰, the mismatch between housing demand and supply and unaffordable prices has led to the growth of over 1700 unauthorised colonies, which are home to over four million people in the city. The Delhi economic survey 2019–20 reports that there are 643 vehicles per 1000 population, double the 2005–06 number, leading to traffic congestion, road accidents, and parking space shortages. The Master Plan Delhi-2041 projects that Delhi will have over 46 million daily trips, with a per capita trip rate of 1.58. These are some of the issues that Delhi is facing due to massive urbanisation, which may be exemplified further if suitable planning interventions are not made. In this context, we consider Delhi as our case study to understand what local planning interventions can be performed to achieve sustainable urbanisation.

Clusters characteristics

One of the study objectives was to cluster the neighbourhoods in Delhi and characterise them with their dominant built-up forms. After scaling the data and determining the optimum number of clusters, we ran the k-means clustering algorithm with five urban form elements. The algorithm clustered the grid cells into six distinctive clusters, designated T1 to T6. Table 1 shows the percentage share of the grid cells in the six clusters. Cluster T6 has the maximum share of the total grid cells, i.e., 28%, while cluster T3 has the least share, i.e., 7%. Figure 2 shows the spatial distribution of the grid cells under the six clusters. We find cluster T1 is located in the centre and south of the city. Cluster T2 is spread in the south and north of the city, while cluster T3 is visible in small pockets in the north and west of the city. Cluster T4 is

Table 1. Percentage share of grid cells in the six clusters

| Cluster | T1 | T2 | T3 | T4 | T5 | T6 |
|-------------------------|----|----|----|----|----|----|
| Share of grid cells (%) | 19 | 23 | 7 | 10 | 13 | 28 |

**Fig. 2** Spatial distribution of grid cells under the six clusters. The figure shows the spatial distribution of six clusters across Delhi. Cluster T6 has the maximum share of the total grid cells while cluster T3 has the least share.

primarily located in the city's outer areas, and cluster T5 can be seen in locations lying towards the central west of the city. Lastly, cluster T6 is primarily clustered in the northeast of the city.

To determine the statistical disparity among the clusters, we performed the multivariate analysis of variance and covariance (MANOVA) test. MANOVA is a statistical test used to measure the impact of one or more independent variables (factor variables) on two or more dependent variables. In other words, the MANOVA test determines whether the mean value of the dependent variable changes for different groups in the independent variable. The null hypothesis assumes that there is no statistical difference in the mean values of the chosen dependent variables across different groups⁷¹. We ran the MANOVA test using the urban form elements as dependent variables and the assigned clusters as independent variables. We used four different test parameters to determine statistical significance. All four tests computed in the MANOVA rejected the null hypothesis based on p-value significance. Table 2 presents the result of the MANOVA test. The test revealed substantial disparities in the average values of urban form elements among the six clusters, indicating that each cluster possesses a different composition of urban form features.

To graphically analyse the differences among the clusters, two data visualisation techniques were employed: box plots and parallel coordinate plots. Figure 3 displays the box plots of various urban form elements across the six clusters. For example, Fig. 3(a) shows that the median accessibility score is highest in cluster T1, whereas Fig. 3(b) shows that the median street intersection density is highest in cluster T5.

Figure 4 depicts the parallel coordinate plot of the multivariate data for each cluster, offering a visual illustration of the disparities between the clusters. As shown in Fig. 4, the urban form variables are marked on the x-axis, and the clusters are shown with different coloured line segments. For example, cluster T1 (marked in green) has the highest normalised score for accessibility, whereas cluster T3 (marked in red) has the highest normalised score for growth in the built-up area.

Table 2. MANOVA test result.

| Test Name | Statistic | df | df1 | df2 | F | Prob>F |
|------------------------|-----------|----|--------|----------|----------------|----------------|
| Wilks' lambda | 0.0145 | 5 | 25 | 137755 | 11707.02 | 0 ^a |
| Pillai's trace | 2.4759 | 25 | 185430 | 7275.75 | 0 ^a | |
| Lawley-Hotelling trace | 8.8455 | 25 | 185402 | 13119.79 | 0 ^a | |
| Roy's largest root | 4.8066 | 5 | 37086 | 35651.56 | 0 ^u | |
| Residuals | 37086 | | | | | |
| Total | 37091 | | | | | |

e exact, *a* approximate, *u* upper bound on F.

Cluster typology

The SHAP (Shapley Additive exPlanations) tool was employed to visualise the variations between the clusters using the SHAP module in Python. As described in the Introduction section, the SHAP tool is a machine learning technique that provides a way to explain the contribution of each feature in a prediction made by a model and how that feature affects the output. The SHAP tool generates a plot called a "summary plot", which displays the most important features and how they affect the prediction. Figure 5 displays the summary plot of the SHAP tool.

The summary plot shows the nature and magnitude of the impact of urban form elements on cluster prediction. The element that has the strongest impact on cluster predictability is considered the dominant element and is used to frame the cluster typology. In the case of Cluster T1, we find that high values of accessibility to services, built-up density, and block size have a positive SHAP value. Moderate values of street intersection density and low values of the growth rate also have a positive SHAP value. This means that cluster T1 is more likely to be predicted by high values of accessibility to services, built-up density, and block size, along with moderate values of street intersection density, and low values of growth. However, to construct the cluster typology, we use the cluster feature that has the strongest impact, which is a high value of accessibility to services in the case of cluster T1. Thus, we label cluster T1 as an area with high accessibility to services.

Based on the analysis of cluster T1, the typologies of the remaining clusters can be similarly framed. Cluster T2 has a moderate built-up density as its dominant feature, and it is also marked by a moderate growth rate, a low accessibility value, and a large block size. With such characteristics, cluster T2 is labelled as a moderate built-up density area. The dominant feature in cluster T3 is a very high growth rate, with a moderate built-up density and street intersection density also contributing to cluster characterisation. Thus, cluster T3 is labelled as an area with high growth rate.

In cluster T4, low built-up density is the dominant feature. Other important features include a high growth rate and low accessibility. We find that residential areas in this cluster are mostly located on the city's periphery. Considering the high growth rate and spatial location of residential areas lying in cluster T4, the cluster is referred to as an urbanisable area with a rural landscape. The dominant feature in cluster T5 is high street intersection density. High built-up density and small block size also contribute to the cluster characterisation, which makes the cluster densely populated. Thus, cluster T5 is labelled as a compact and congested area. Finally, in cluster T6, we find that high built-up density has the highest positive impact, followed by moderate to low accessibility and moderate to low street intersection density. Due to this, cluster T6 is referred to as a high density area.

Table 3 tabulates the clusters' characteristics based on the boxplot and multivariate plot analyses, and their dominant urban form or typology framed using the SHAP tool. The next section

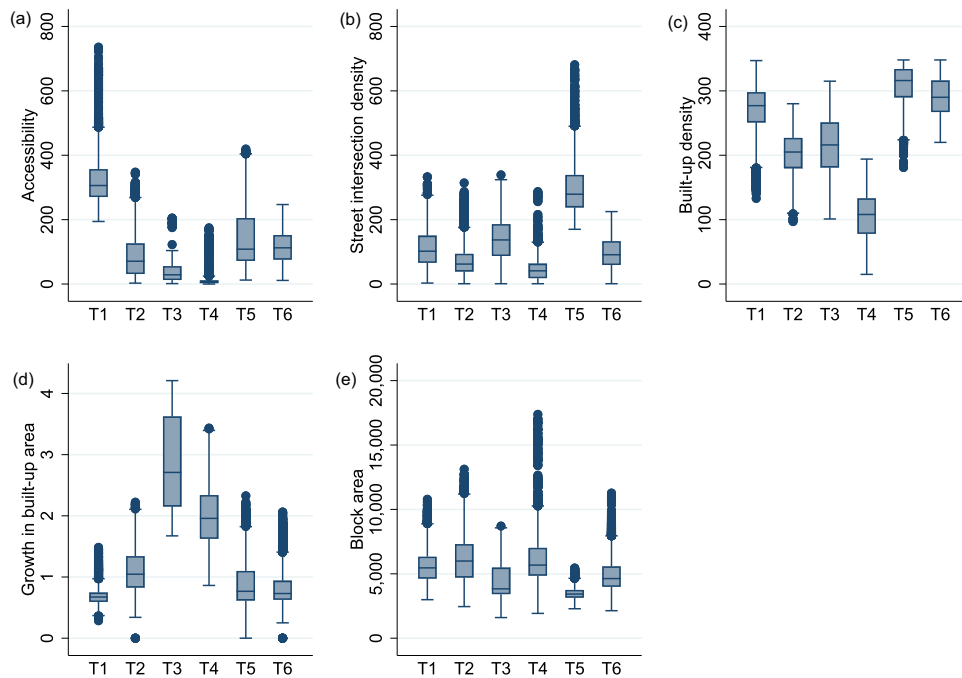


Fig. 3 Box plots showing the data distribution of urban form elements across different clusters. The urban form elements shown in the figure are as follows: (a) accessibility, (b) Street intersection density, (c) Built-up density, (d) Growth in the built-up area, and (e) Block area (sq. metres). In a box plot, the distribution of the data is represented using a box and a set of whiskers. The box in a box plot represents the interquartile range (IQR) of the data, which is the range between the 25th and 75th percentiles of the data. The median value of the data is represented by a line inside the box. The whiskers extend from the box to the minimum and maximum values of the data, excluding any outliers.

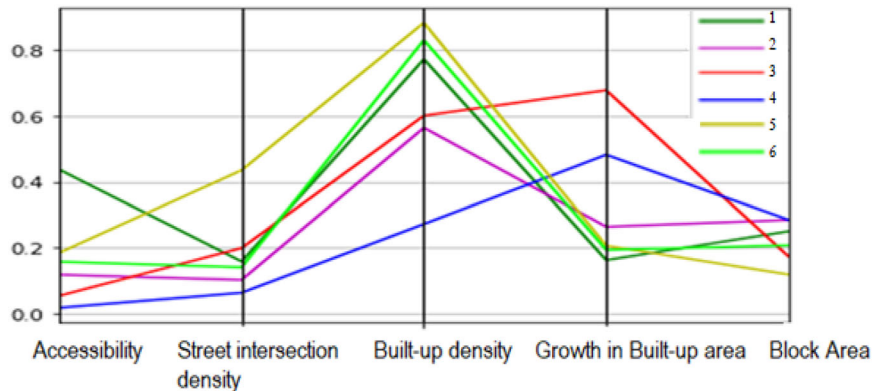


Fig. 4 Parallel coordinate plot. The variables are shown on the x-axis, and their normalised mean value score for every cluster is shown on the y-axis. In a parallel plot, each variable in the dataset is represented by a separate axis, which is arranged in parallel to each other. The data points are then plotted as a set of connected line segments across the different axes, with each line segment representing the value of a particular variable for a specific data point¹¹⁰. The lines marked with numbers 1 to 6 represent the clusters from T1 to T6 in the same order.

discusses the clusters' characteristics and their impact on sustainable urbanisation.

Before concluding this section, we highlight a few recent studies that have used k-means clustering to analyse neighbourhood typologies. For example, Vogiazides and Mondani⁷² used k-means clustering to cluster neighbourhoods in Sweden to analyse the variation in neighbourhood status and found ten different neighbourhood types. Wu et al.⁷³ used k-means clustering to identify four neighbourhood typologies for London, Paris, and Amsterdam. Similarly, Lynge et al.⁷⁴ used k-means clustering to build eight neighbourhood typologies for different cities in South Africa. Although, due to different study contexts

and the choice of variables, the results from these studies cannot be directly compared with our study results, the fact that k-means clustering has been used successfully in these studies reinforces the reliability and robustness of this technique. The demonstration of consistent findings in previous studies lends additional support to the validity of the current research.

DISCUSSION

The cluster analysis identified six distinct residential built-up form typologies in Delhi. This section develops the study's second objective, which is to analyse how these different built-up form

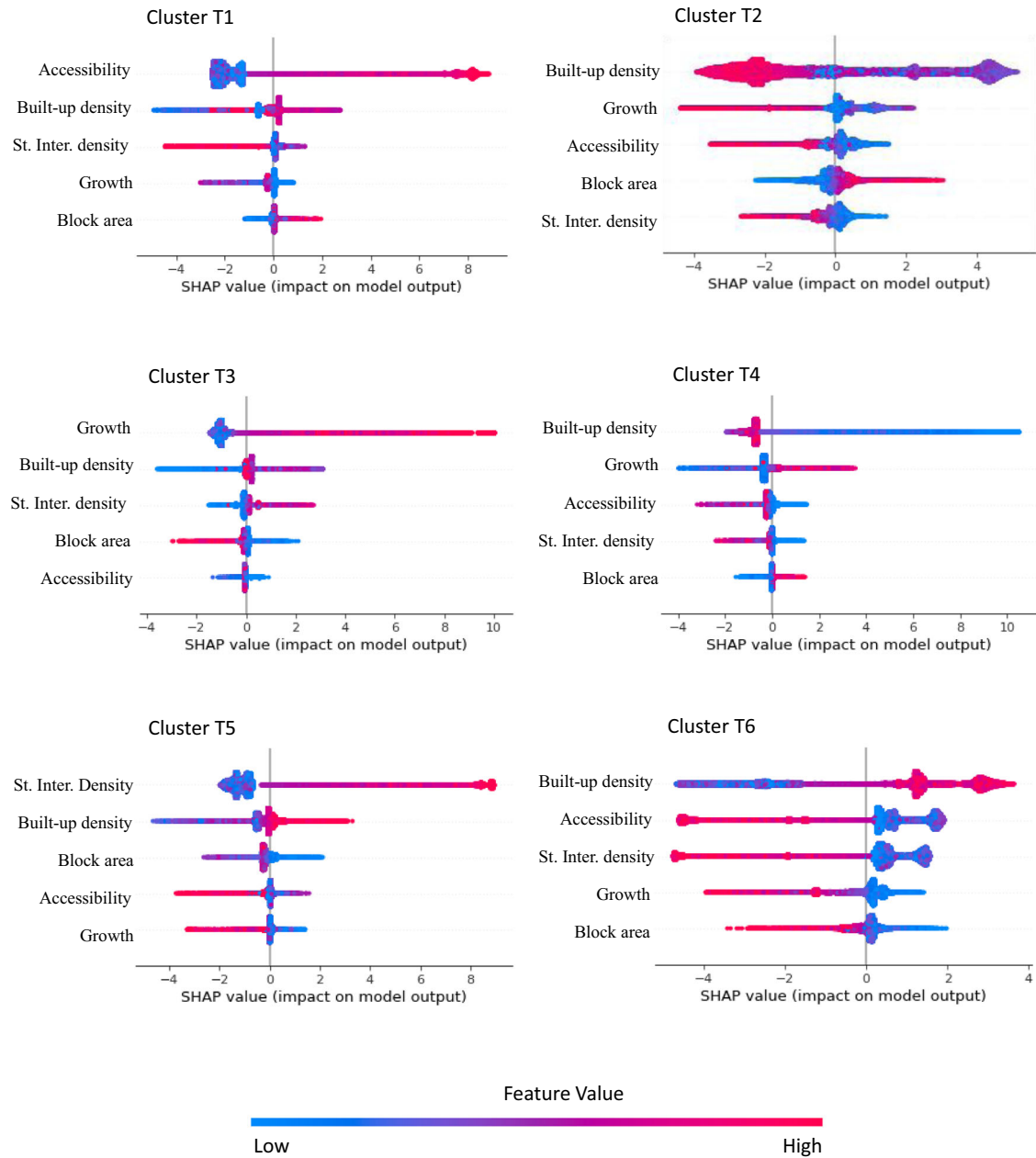


Fig. 5 SHAP value summary plots for all six clusters. The SHAP value summary plot ranks features based on their contribution to cluster characterisation, with the most important features at the top. Each feature is represented by a horizontal bar where the colour of the bar indicates the value of the variable or feature, with blue indicating a low value and red indicating a high value⁶⁴. The direction of the SHAP value denotes how effectively a feature value can characterise the cluster. A positive (negative) SHAP value denotes a feature that is more likely (less likely) to characterise the cluster.

typologies affect sustainable urbanisation in Delhi. Before discussing further, we first list the parameters used in this study to comprehend sustainable urbanisation. Sustainability is a comprehensive concept examined in the literature from different perspectives^{75,76}. In this study, we analysed sustainability from the perspective of the urban form. Previous studies have analysed sustainability for different urban forms/city models, such as compact cities⁷⁷, urban sprawl⁷⁸, green city framework⁷⁹, transit-oriented development⁸⁰, and smart cities⁸¹. While all of these frameworks have some advantages and limitations, there is no consensus as to what constitutes the best urban form from a sustainability perspective⁸². Moreover, the applicability of an

urban form to a city's planning is influenced by diverse factors, including the city's existing land use pattern and resource availability.

Using the sustainability indicators from the different city models and based on our understanding of urbanisation and urban form in Delhi, we identify the following urban form characteristics that can contribute to sustainable urbanisation: high accessibility to services and transit stations, moderate or high built-up density with open spaces, moderate street network density, and large block sizes. Moderate to high built-up density, along with high accessibility to services, ensures spatial equity in the distribution of public services across neighbourhoods. Moderate street

Table 3. Cluster characterisation and typology.

| Cluster | Accessibility | Built-up density | Growth rate | Street intersection density | Block area | Dominant Urban Form or Cluster Typology |
|---------|---------------|------------------|-------------|-----------------------------|------------|---|
| T1 | + | + | – | +/- | + | High accessibility to services |
| T2 | – | +/- | +/- | – | + | Moderate built-up density |
| T3 | – | +/- | ++ | +/- | +/- | High growth rate |
| T4 | – | – | + | – | + | Urbanisable area with a rural landscape |
| T5 | +/- | + | – | ++ | – | Compact and congested |
| T6 | +/- | + | – | +/- | +/- | High density |

Note: '++' denotes very high/ very big, '+' denotes high/big, '+/-' denotes moderate, and '–' denotes low/small.

network density in areas of high density and high accessibility makes the area compact, which encourages active forms of transportation and reduces street traffic congestion and vehicular emissions^{83–85}. The study also recognises the potential benefits of large block sizes in the context of Delhi, where block sizes are generally smaller than standard norms. Therefore, the inclusion of large block sizes is deemed necessary for sustainable urbanisation in Delhi.

The first typology (T1) is of high accessibility to services. Along with high accessibility, cluster T1 also has moderate street intersection density and larger block sizes, making it less congested despite its high built-up density. Such characterisation resembles the features of a planned neighbourhood, which is known to provide a better quality of life to its residents^{86,87}. In this context, we classify T1 under sustainable urbanisation. From a different perspective, high accessibility to services in planned neighbourhoods also leads to increased housing demand, establishing upscale gated communities and excluding low-income households^{88,89}. In Delhi as well, such areas have a very high residential land price per sq. km, approximately four times higher than the city's average land price⁹⁰. As these areas of high accessibility are primarily populated by high-income households, there is a need for inclusive policies such as (a) affordable housing subsidies for low-income households; (b) inclusive zoning, where a certain percentage of new housing developments is reserved for low- and middle-income households; and (c) building community land trusts, which are non-profit organisations that hold land and make it available for affordable housing or community development purposes⁹¹. Such measures can ensure sustainable urbanisation with social equity.

The second urban form typology (T2) exhibits a moderate level of built-up density and features large block sizes, indicating its potential to facilitate sustainable urbanisation. However, this cluster is also characterised by low street intersection density and limited accessibility, which degrade its sustainability levels. In urban planning, low street intersection density is associated with low walkability and limited access to public transportation, as well as decreased social interaction and community connectivity, which studies have found to have an adverse impact on individual physical and mental health^{92,93}. Thus, better street design and improved access to transit services are crucial for sustainable urbanisation⁸¹. Some policy and planning measures to enhance street intersection density and walkability can be: (a) tactical urbanism, which involves low-cost interventions like street paintings, parklets, and pedestrian plazas that can help test new intersections and street designs before committing to permanent changes; (b) transit-oriented development (TOD) to promote the development of mixed-use, high-density developments around public transit stations, which can help increase the density of intersections in those areas; and (c) narrower streets: reducing the width of streets can encourage slower traffic and can also create space for new intersections and crossings⁹⁴.

The third urban form typology (T3) is of newly urbanised areas with a rapid growth rate over the last ten years. Such residential areas also have moderate building density and moderate street intersection density, which aid in sustainable urbanisation. However, low access to services in such areas is a cause for concern. Low accessibility to services diminishes growth opportunities and degrades the quality of life, as previous studies show^{95,96}. To sustain growth and promote sustainable urbanisation in such areas, accessibility to different services needs to be enhanced. One of the important planning interventions in this regard can be encouraging land use diversity in this cluster, which can provide a variety of services and amenities within a single building or block⁹⁷.

The fourth urban form typology (T4) is of urbanisable areas. These settlements are located on the city's outskirts in isolated pockets and have rural characteristics. Despite their growth over the past decade, they have a low building density. Due to unplanned street networks and block design, features such as large block sizes and low street intersection density are common here. Access to services is also low due to their peripheral location. As these settlements transform from rural to urban, there is an opportunity to improve sustainability through strategic land use and accessibility planning interventions, such as compact and mixed land use development and designing the streets and block size in a manner that promotes walkability and reduces traffic congestion⁹⁸.

The fifth urban form typology (T5) is of compact and congested areas with high street intersection density, small block sizes, and high built-up density. Neighbourhoods with such characteristics, despite having a high degree of interconnected street networks that enhance walkability, also face issues of traffic congestion and environmental pollution⁹⁹. To achieve sustainable urbanisation in such areas, the following planning interventions can be carried out: (a) promoting the efficient use of land through land conversion policies to create more green spaces; (b) promoting smart mobility through the use of technology to get real-time traffic information; and (c) other measures such as encouraging carpooling and road pricing can reduce the number of cars on the road and alleviate traffic congestion¹⁰⁰.

The final urban form typology (T6) can be considered similar to that of cluster T5 in terms of high built-up density. However, cluster T6 has a higher street intersection density and block size, which makes it less compact and congested as compared to cluster T5. To achieve sustainable urbanisation, planning interventions of similar nature as highlighted for cluster T5 can be carried out.

After discussing the different typologies, we are now in a position to evaluate the level of sustainable urbanisation in Delhi. The above analysis shows that while a certain urban form characteristic may contribute to sustainable urbanisation in isolation, sustainability within a specific cluster must be evaluated by considering the contribution of each element. In many clusters,

Table 4. Summary of services and their sources

| Services | Number of observations | Variables | Source |
|--------------------------|------------------------|--|--|
| Schools | 2403 | Location of private and government schools of all levels. | Directorate of Education, Government of Delhi. |
| Hospitals | 969 | Location of public and private hospitals offering tertiary care | Department of Health and Family Welfares, Govt. of Delhi |
| Entertainment Facilities | 827 | Location of shopping malls, movie theatres, and registered restaurants | Department of Excise, Entertainment & Luxury Tax, Govt. of Delhi |
| Metro Stations | 185 | Location of all metro stations | Delhi Metro Rail Corporation |
| Commercial areas | 23 | District centres and Sub-district centres | Delhi Master Plan 2020 |

there are urban form characteristics that support sustainable urbanisation, such as a large block size in cluster T2 or moderate street intersection density in cluster T3. However, to achieve sustainability in a given cluster, all the urban form elements must have a positive impact. As our results show, only in cluster T1 do all urban form characteristics contribute to sustainable urbanisation. Considering that the total area of grid cells that come under cluster T1 is only 19% of the total residential area in Delhi (refer to Table 1), we conclude that only 19% of the residential area in Delhi can be considered under sustainable urbanisation. The rest of the area requires different forms of intervention to make urbanisation sustainable, as noted in the above paragraphs.

As highlighted in the Introduction section, this study is important considering the rapid pace of urbanisation in cities in the global south. Given the nature of urbanisation in cities like Delhi, as they become more urbanised, they face various challenges for sustainable development. Noting the impact of urban forms on sustainability, as widely recognised in the literature, this study argues for incorporating the built-up form into the characterisation and measurement of urbanisation. Such an approach, as demonstrated in this study, can provide a more accurate assessment of urbanisation and help bring localised planning interventions to areas that have unsustainable urban form features.

We conclude the study by highlighting the major findings. The study aimed to explore residential built-up form typologies and assess their impact on sustainable urbanisation in Delhi. Only a few studies have explored variations in the urban form at the neighbourhood level, and none exist specifically for cities in the Indian subcontinent. The study used a grid-based technique to divide residential areas into 100 x 100 metre grid cells and assigned attributes of accessibility, built-up density, and street design. The grid cells were then clustered using the k-means clustering algorithm, which showed the presence of six built-up form clusters in Delhi. Using the MANOVA test statistics and graphical visualisations, these clusters were analysed for variation in their urban form elements and were found to be significantly different from one another. Using the SHAP tool, the clusters were analysed for their dominant urban form, using which cluster typologies were framed. These typologies can be listed as: (1) areas with high accessibility to services; (2) areas with moderate built-up density; (3) areas with a high growth rate; (4) urbanisable areas with a rural landscape; (5) compact and congested areas; and (6) high density areas. The study then discussed how the different built-up form elements in these clusters contribute to sustainable urbanisation in Delhi. Based on the results, the study concludes that only 19% of residential areas in Delhi can be classified under sustainable urbanisation, while the remaining areas require different planning interventions to achieve sustainable urbanisation.

We note here the limitations of the study. First, the study considers only the physical elements of urban form and their association with sustainable urbanisation. The understanding of

Table 5. Data summary of urban form elements

| Variable | Mean | Std. dev. | Min | Max |
|-----------------------------|---------|-----------|---------|----------|
| Accessibility | 134.8 | 115.31 | 0 | 735.63 |
| Built-up area | 245.58 | 69.32 | 15 | 348 |
| Growth in built-up area | 10.33 | 35.44 | 0 | 314 |
| Street intersection density | 116.73 | 91.47 | 1 | 681 |
| Block area (sq. metres) | 5187.11 | 1682.68 | 1602.36 | 17379.24 |

sustainable urbanisation can be further enriched by including socioeconomic and demographic indicators of neighbourhoods, such as population density, economic status, and age and caste-wise composition. Furthermore, sustainability can be analysed using local environmental and ecological indicators, such as the air quality index. In this manner, one can develop a more comprehensive understanding of sustainable urbanisation in different neighbourhoods, and appropriate localised policies can be developed. Second, the study has used the cumulative measure of accessibility, which is a potential measure. Other accessibility measures based on time/ distance or demand and supply of services, such as the 2-step floating catchment area method, can be used. This can provide a more realistic measure of accessibility. Finally, with the advancements in computationally efficient learning algorithms, future studies can perform a comparative analysis among different clustering techniques to examine the method with the highest clustering efficiency and utilise it to cluster the urban form at the neighbourhood level.

Despite the shortcomings, the study offers a viewpoint for understanding urbanisation in rapidly urbanising cities like Delhi, which are characterised by spatial heterogeneity in their urban form. Our study shows that urban form at the neighbourhood level can show significant spatial variation, and thus characterising a city with a particular urban form can be misleading. Our study methodology is generalisable to other cities and can be utilised to create development zones that are based on the dominant built-up types and are defined by adaptive and dynamic boundaries. Based on the dominant characteristics of the zones or clusters, the required planning interventions can be sought. Thus, by factoring in the neighbourhood's built-up form in the analysis of sustainable urbanisation, this study provides another perspective to study urbanisation, on which future studies can build by considering non-physical characteristics of neighbourhoods.

METHODS

Data preparation

The first step in the data preparation process involved mapping residential areas in Delhi using Google Earth. The mapped

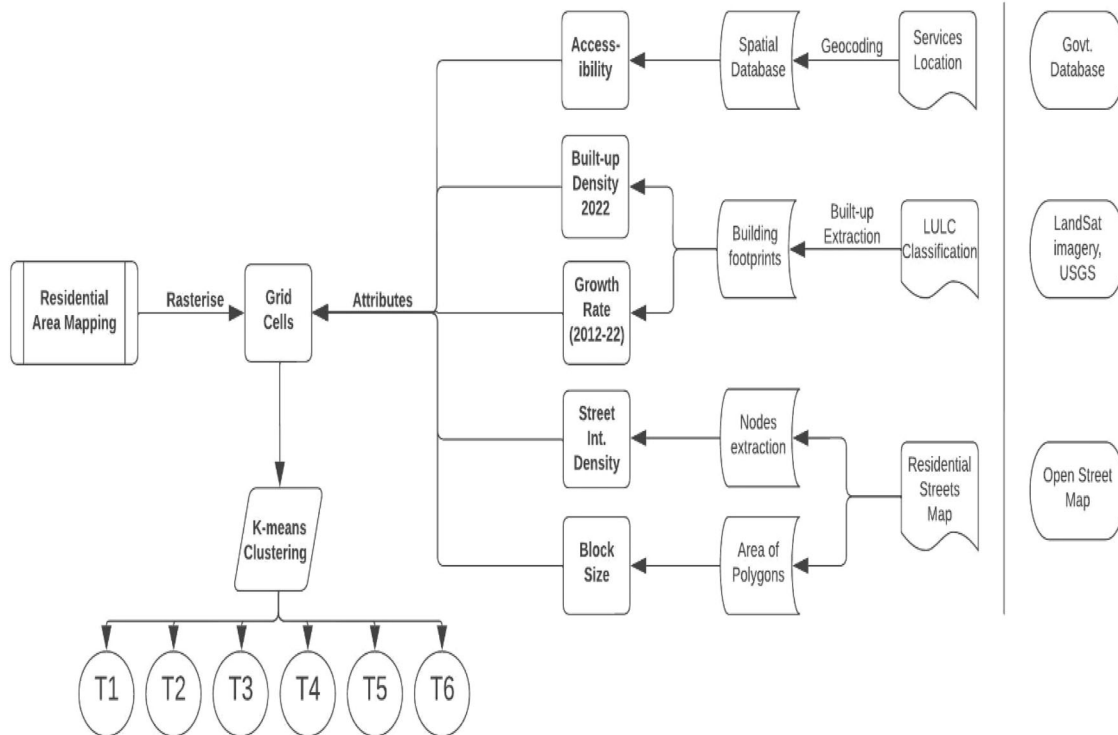


Fig. 6 A flowchart of research methodology. The flowchart summarises the research methodology under three parts: residential area mapping, assigning the attributes of urban form elements to the grid cells, and k-means clustering.

residential area was then exported to ArcGIS software and transformed into a raster file with a cell size of 100×100 metre, resulting in 37,092 grid cells. These grid cells served as the spatial unit of analysis and were assigned different urban form elements.

Urban form refers to the physical and spatial characteristics of urban areas, including the arrangement and distribution of buildings, streets, open spaces, and other features that shape the built environment¹⁰¹. Elements of the urban form commonly found in previous studies belong to the 5D framework developed by Ewing and Cervero¹⁰², which includes population density, land use diversity, street design, destination accessibility, and distance to transit stations^{3,103}. Despite its widespread use, there has yet to be a consensus in the literature on what constitutes urban form. As Fleischmann et al.¹¹ note, the term has many interpretations, leading to the need for an objective system of measuring urban form features. In this study, we focus on commonly used elements of urban form: (a) density, including built-up density and growth in built-up density from 2012 to 2022; (b) street design, including street intersection density and block size; and (c) accessibility to the five services. Our choice of elements was based on the research aim, study area context, data availability, and ease of result interpretation.

The first element is accessibility to services. While accessibility can be computed using different measures such as time/distance or supply-demand of services, we stick to the cumulative measure of accessibility because it is easy to measure and interpret, and provides a good indication of the spatial distribution of services in a neighbourhood¹⁰⁴. A buffer radius of 2 km was used to represent services lying in the immediate neighbourhood that can be accessed with a non-motorised travel mode. We considered five types of services: schools, hospitals, entertainment facilities, commercial areas, and metro rail stations. Good access to schools and hospitals is crucial for human development, as highlighted in Sustainable Development Goals 3 and 4. High accessibility to metro stations provides quick and convenient transportation

options to people, reduces their travel time, and improves overall mobility. Studies show that high access to entertainment facilities and commercial areas enhances the quality of life²².

Table 4 lists the services, the number of observations for each service, and their sources. To compute the accessibility to each of these services, we first obtained the location addresses of all the observations from their respective sources and then created a spatial database by geocoding the addresses in Google Maps. The geocoded addresses of these observations for every service were then mapped in ArcGIS in a point shapefile format. Grid cells were assigned a total count of observations lying in a circular radius of 2 km for every service. Finally, accessibility was calculated as the sum of the normalised value of the cumulative count of observations for all five services, as shown in Eq. (1):

$$A_i = K \sum_{j=1}^5 \left[\frac{x_{ij} - \min_j}{\max_j - \min_j} \right] \quad (1)$$

where A_i represents the accessibility of the grid cell i , x_{ij} represents the total count of observations belonging to service j and lying in the circular radius of 2 km from the grid cell i , \min_j and \max_j represents the minimum and maximum count of observations, respectively, belonging to service j and lying within a radius of 2 km across the grid cells. K was used as a constant to keep the accessibility values within a reasonable limit.

The second and third factors were built-up density and growth in built-up density, respectively. The built-up density was calculated in three sequential steps. First, land satellite imagery of Delhi for the year 2022 was acquired from the USGS (United States Geological Survey) and exported in ArcGIS. Second, land use classification was done using a supervised classification tool to classify land use under built-up, bare soil, cultivated area, wasteland, and water. The accuracy of land use classification was verified by cross-checking the land use of randomly sampled 350 data points from the classified image with the actual land use as visible in the historical imagery tool

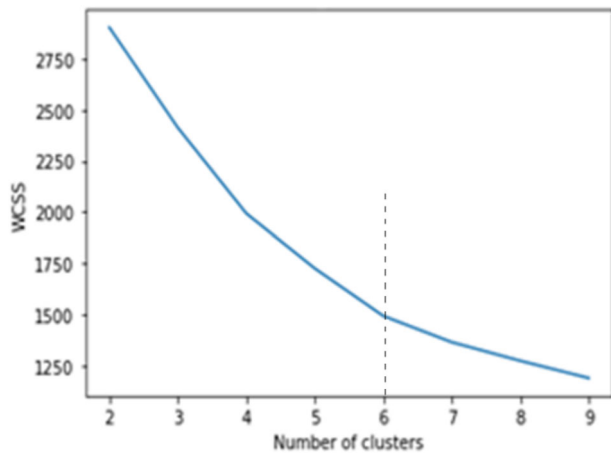


Fig. 7 Elbow method to figure out the optimum number of clusters. The figure shows the decrease in the value of WCSS with increase in number of clusters. The dotted line shows the value of k at which the marginal change in the value of the WCSS with an increase in the value of k is at its minimum.

of the Google Earth software. The classification accuracy measured using the kappa statistics was 0.86, which denotes a high accuracy of classification¹⁰⁵. Third, the area under the built-up category was extracted from the classified image and vectorised into points. The built-up density of a grid cell was calculated as the number of built-up points lying in a 500 metre buffer radius from each grid cell. The same process was repeated to calculate the built-up density for each grid cell in 2012. Finally, we calculated the growth in built-up density as the percentage change in built-up density from 2012 to 2022 for each grid cell.

The fourth and fifth elements of our study were street intersection density and block size, respectively, which were computed using the open street map (OSM) database. The OSM is a collaborative open-source mapping platform that provides information on roads, buildings, landmarks, and other geographic features¹⁰⁶. OSM can be accessed through various applications, such as QGIS, which was used in this study. To calculate street intersection density, we used the residential street layout from the OSM database, which provides a map of all residential streets in Delhi. Using the line intersection tool in QGIS, we measured the number of residential street intersection nodes within a 500 metre radius for each grid cell. To measure the block size, we first calculated the area of the polygons formed by enclosing residential streets. For a grid cell, the block size represented the average area of all polygons lying inside a buffer radius of 500 metre. A data summary of all the elements of the urban form used in this study is shown in Table 5. The research methodology described in this section is summarised in the flowchart in Fig. 6.

K-means clustering

We used k-means clustering to classify grid cells into similar urban form attributes. K-means clustering is a machine learning algorithm used for clustering or grouping data points in a dataset. The algorithm partitions the data into 'k' non-overlapping clusters, where k is a predefined number chosen by the user. The algorithm works by iteratively assigning each data point to the closest cluster centre (centroid) and then recalculating the centroid of each cluster based on the newly assigned points. This process continues until the centroids no longer move significantly or a specified maximum number of iterations is reached¹⁰⁷. We used the *scikit-learn* library¹⁰⁸ in Python to execute the algorithm. First, we scaled the data using the min-max scaler to a range of 0 to 1.

We then determined the optimum number of clusters using the widely used elbow method¹⁰⁹. The elbow method calculates the total variation within a cluster using the WCSS (within-cluster sum of squares) and plots the results to determine the optimum number of clusters (k).

As the number of clusters increases, the variation within every cluster, i.e., the value of the WCSS, is expected to decline. The optimum value of k is one where the marginal decrease in the value of the WCSS by adding one more cluster is minimal. This can be visualised by plotting the values of the WCSS against the number of clusters (k). As shown in Fig. 7, when the value of k is six, a sharp bend or an elbow-shaped curve occurs in the graph. At this point, the marginal change in the value of the WCSS with an increase in the value of k is at its minimum. The elbow method showed that the optimum number of clusters was six, which was used in the k-means clustering. Each cluster was then assigned a built-up form typology using the SHAP tool.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

The dataset used in the study are available in the Github repository, <https://github.com/aviralmarwal/aviral>. This work is licenced under a Creative Commons Attribution 4.0 International (CC BY 4.0) licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The authors would like to clarify that all the datasets used in this study have been acquired from secondary sources, as mentioned in the methods section. We have no ethical considerations to declare concerning data confidentiality, consent, or security.

CODE AVAILABILITY

The code used in the study are available in the Github repository, <https://github.com/aviralmarwal/aviral>.

Received: 1 February 2023; Accepted: 1 June 2023;
Published online: 26 June 2023

REFERENCES

- United Nations, Department of Economic and Social Affairs, Population Division. *World Urbanization Prospects: The 2018 Revision* Online Edition (United Nations, 2018).
- Abrantes, P. et al. Modelling urban form: a multidimensional typology of urban occupation for spatial analysis. *Environ. Plan. B: Urban Anal. City Sci.* **46**, 47–65 (2019).
- Sharifi, A. Urban form resilience: a meso-scale analysis. *Cities* **93**, 238–252 (2019).
- Alam, T. & Banerjee, A. Deciphering urbanization and spatial disparity in south 24 Parganas district of West Bengal, India. In: *Proc. 58th ISOCARP World Planning Congress*, Brussels, Belgium (2022).
- Engelfriet, L. & Koomen, E. The impact of urban form on commuting in large Chinese cities. *Transportation* **45**, 1269–1295 (2018).
- Mouratidis, K. Built environment and social well-being: How does urban form affect social life and personal relationships? *Cities* **74**, 7–20 (2018).
- Zhou, C., Li, S. & Wang, S. Examining the impacts of urban form on air pollution in developing countries: A case study of China's megacities. *Int. J. Environ. Res. Public Health* **15**, 1565 (2018).
- Hankey, S. & Marshall, J. D. Urban form, air pollution, and health. *Curr. Environ. Health Rep.* **4**, 491–503 (2017).
- Zhang, J. et al. Impact of urban block typology on building solar potential and energy use efficiency in tropical high-density city. *Appl. Energy* **240**, 513–533 (2019).
- Eurostat. *Methodological Manual on Territorial Typologies* (Publications Office of the European Union, 2018).
- Fleischmann, M., Romice, O. & Porta, S. Measuring urban form: Overcoming terminological inconsistencies for a quantitative and comprehensive morphologic analysis of cities. *Environ. Plan. B: Urban Analytics City Sci.* **48**, 2133–2150 (2021).

12. Peimani, N. & Kamalipour, H. Assembling transit urban design in the global South: urban morphology in relation to forms of urbanity and informality in the public space surrounding transit stations. *Urban Sci.* **6**, 18 (2022).
13. Alam, T. & Banerjee, A. Characterizing land transformation and densification using urban sprawl metrics in the South Bengal region of India. *Sustain. Cities Soc.* **89**, 104295 (2023).
14. Sandoval, V. & Sarmiento, J. P. A neglected issue: Informal settlements, urban development, and disaster risk reduction in Latin America and the Caribbean. *Disaster Prev. Manag. Int. J.* **29**, 731–745 (2020).
15. Kraff, N. J., Wurm, M. & Taubenböck, H. The dynamics of poor urban areas—analyzing morphologic transformations across the globe using Earth observation data. *Cities* **107**, 102905 (2020).
16. Mahtta, R., Mahendra, A. & Seto, K. C. Building up or spreading out? Typologies of urban growth across 478 cities of 1 million+. *Environ. Res. Lett.* **14**, 124077 (2019).
17. Leasure, D. R., Jochem, W. C., Weber, E. M., Seaman, V. & Tatem, A. J. National population mapping from sparse survey data: A hierarchical Bayesian modeling framework to account for uncertainty. *Proc. Natl. Acad. Sci.* **117**, 24173–24179 (2020).
18. Nieuwenhuijsen, M. J. Urban and transport planning pathways to carbon neutral, liveable and healthy cities; A review of the current evidence. *Environ. Int.* **140**, 105661 (2020).
19. Schirmer, P. M. & Axhausen, K. W. A multiscale clustering of the urban morphology for use in quantitative models. In: D'Acci, L. (eds) *The Mathematics of Urban Morphology. Modeling and Simulation in Science, Engineering and Technology* 355–382 (Birkhäuser, Cham., 2019).
20. Fleischmann, M., Feliciotti, A., Romice, O. & Porta, S. Methodological foundation of a numerical taxonomy of urban form. *Environ. Plan. B: Urban Analytics City Sci.* **49**, 1283–1299 (2022).
21. Su, H., Han, G., Li, L. & Qin, H. The impact of macro-scale urban form on land surface temperature: an empirical study based on climate zone, urban size and industrial structure in China. *Sustainable Cities Soc.* **74**, 103217 (2021).
22. Li, F., Zhou, T. & Lan, F. Relationships between urban form and air quality at different spatial scales: A case study from northern China. *Ecol. Indic.* **121**, 107029 (2021).
23. Masoumi, H. E., Terzi, F. & Serag, Y. M. Neighborhood-scale urban form typologies of large metropolitan areas: Observations on Istanbul, Cairo, and Tehran. *Cities* **85**, 170–186 (2019).
24. Debray, H., Kraff, N. J., Zhu, X. X. & Taubenböck, H. Planned, unplanned, or in-between? A concept of the intensity of plannedness and its empirical relation to the built urban landscape across the globe. *Landsc. Urban Plan.* **233**, 104711 (2023).
25. Ahmad, S. & Choi, M. J. The context of uncontrolled urban settlements in Delhi. *Asien* **118**, 75–90 (2011).
26. Abascal, A. et al. Identifying degrees of deprivation from space using deep learning and morphological spatial analysis of deprived urban areas. *Comput. Environ. Urban Syst.* **95**, 101820 (2022).
27. Tang, B. S., Wong, K. K., Tang, K. S. & Wai Wong, S. Walking accessibility to neighbourhood open space in a multi-level urban environment of Hong Kong. *Environ. Plan. B: Urban Analytics City Sci.* **48**, 1340–1356 (2021).
28. Palaiologou, G., Larimian, T. & Vaughan, L. The use of morphological description in neighbourhood planning: form-based assessment of physical character and design rules. *J. Urbanism* **14**, 490–514 (2021).
29. Braulio-Gonzalo, M., Ruá, M. J. & Bovea, M. D. Exploring residential urban form patterns: a Spanish case study. *Int. Plann. Stud.* **23**, 120–138 (2018).
30. Lu, S., Huang, Y., Shi, C. & Yang, X. Exploring the associations between urban form and neighborhood vibrancy: a case study of Chengdu, China. *ISPRS Int. J. Geo-Inf.* **8**, 165 (2019).
31. Fusco, G., Araldi, A., Emsellem, D. & Overal, D. Taxonomy of contemporary urban forms in France: Towards an Urban Atlas through Multiple Fabric Assessment. In: *XXIX International Seminar on Urban Form ISUF2022: Urban Redevelopment and Revitalisation. A Multidisciplinary Perspective-Proceedings, Łódź - Kraków* (HAL, 2022).
32. Openshaw, S. *The Modifiable Areal Unit Problem (Concepts and Techniques in Modern Geography: 38)* (Geo-Books, 1984).
33. Viegas, J. M., Martinez, L. M. & Silva, E. A. Effects of the modifiable areal unit problem on the delineation of traffic analysis zones. *Environ. Plan. B: Plann. Des.* **36**, 625–643 (2009).
34. Hou, H., Estoque, R. C. & Murayama, Y. Spatiotemporal analysis of urban growth in three African capital cities: A grid-cell-based analysis using remote sensing data. *J. Afr. Earth Sci.* **123**, 381–391 (2016).
35. Nedovic-Budic, Z., Knaap, G. J., Shahumyan, H., Williams, B. & Slaev, A. Measuring urban form at community scale: Case study of Dublin, Ireland. *Cities* **55**, 148–164 (2016).
36. Tollefsen, A. F., Strand, H., & Buhaug, H. PRIO-GRID: A unified spatial data structure. *J. Peace Res.* **49**, 363–374.
37. Brown, D., Japa, A. & Shi, Y. A fast density-grid based clustering method. In: *IEEE 9th Ann. Comput. Commun. Workshop Conf. (CCWC) 0048-0054* (IEEE, 2019).
38. Rothlisberger, V., Zischg, A. P. & Keiler, M. Identifying spatial clusters of flood exposure to support decision making in risk management. *Sci. Total Environ.* **598**, 593–603 (2017).
39. Ahmed, S. & Bramley, G. How will Dhaka grow spatially in the future?—modelling its urban growth with a near-future planning scenario perspective. *Int. J. Sustain. Built Environ.* **4**, 359–377 (2015).
40. Jiawei, H., Micheline, K., & Jian, P. *Data Mining: Concepts and Techniques*. 3rd edn. (Morgan Kaufmann, 2012).
41. Su, S., Zhang, H., Wang, M., Weng, M. & Kang, M. Transit-oriented development (TOD) typologies around metro station areas in urban China: a comparative analysis of five typical megacities for planning implications. *J. Transp. Geogr.* **90**, 102939 (2021).
42. Mehrotra, S., Bardhan, R. & Ramamritham, K. Outdoor thermal performance of heterogeneous urban environment: An indicator-based approach for climate-sensitive planning. *Sci. Total Environ.* **669**, 872–886 (2019).
43. Oke, J. B. et al. A novel global urban typology framework for sustainable mobility futures. *Environ. Res. Lett.* **14**, 095006 (2019).
44. Pilehforooshha, P. & Karimi, M. A local adaptive density-based algorithm for clustering polygonal buildings in urban block polygons. *Geocarto Int.* **35**, 141–167 (2020).
45. Guyot, M., Araldi, A., Fusco, G. & Thomas, I. The urban form of Brussels from the street perspective: The role of vegetation in the definition of the urban fabric. *Landsc. Urban Plan.* **205**, 103947 (2021).
46. Jochem, W. C. et al. Classifying settlement types from multi-scale spatial patterns of building footprints. *Environ. Plan. B: Urban Analytics City Sci.* **48**, 1161–1179 (2020).
47. Perez, J., Fusco, G., Araldi, A. & Fuse, T. Identifying building typologies and their spatial patterns in the metropolitan areas of Marseille and Osaka. *Asia Pac. J. Reg. Sci.* **4**, 193–217 (2020).
48. Li, Z. et al. Transit oriented development among metro station areas in Shanghai, China: Variations, typology, optimization and implications for land use planning. *Land Use Policy* **82**, 269–282 (2019).
49. Yuan, C. & Yang, H. Research on K-value selection method of K-means clustering algorithm. *J.* **2**, 226–235 (2019).
50. Fränti, P. & Sieranoja, S. How much can k-means be improved by using better initialization and repeats? *Pattern Recognit.* **93**, 95–112 (2019).
51. Govender, P. & Sivakumar, V. Application of k-means and hierarchical clustering techniques for analysis of air pollution: A review (1980–2019). *Atmos. Pollut. Res.* **11**, 40–56 (2020).
52. Patel, E. & Kushwaha, D. S. Clustering cloud workloads: K-means vs gaussian mixture model. *Proc. Comput. Sci.* **171**, 158–167 (2020).
53. Ahmed, M., Seraj, R. & Islam, S. M. S. The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics* **9**, 1295 (2020).
54. Iqbal, N. et al. Linkages between Typologies of Existing Urban Development Patterns and Human Vulnerability to Heat Stress in Lahore. *Sustainability* **14**, 10561 (2022).
55. Shi, Z., Fonseca, J. A. & Schlueter, A. A parametric method using vernacular urban block typologies for investigating interactions between solar energy use and urban design. *Renew. Energy* **165**, 823–841 (2021).
56. Xu, S. et al. A method of calculating urban-scale solar potential by evaluating and quantifying the relationship between urban block typology and occlusion coefficient: a case study of Wuhan in Central China. *Sustain. Cities Soc.* **64**, 102451 (2021).
57. Bobkova, E., Berghauer Pont, M. & Marcus, L. Towards analytical typologies of plot systems: Quantitative profile of five European cities. *Environ. Plan. B: Urban Analytics City Sci.* **48**, 604–620 (2021).
58. Fahim, A. K. and starting means for k-means algorithm. *J. Comput. Sci.* **55**, 101445 (2021).
59. Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R. & Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. USA* **116**, 22071–22080 (2019).
60. Rai, A. Explainable AI: From black box to glass box. *J. Acad. Market. Sci.* **48**, 137–141 (2020).
61. Brandsæter, A., & Glad, I. K. Shapley values for cluster importance. *Data Min. Knowl. Discov.* 1–32 (2022).
62. Wang, J. & Biljecki, F. Unsupervised machine learning in urban studies: a systematic review of applications. *Cities* **129**, 103925 (2022).
63. Khadem, H., Nemat, H., Eissa, M. R., Elliott, J. & Benaissa, M. COVID-19 mortality risk assessments for individuals with and without diabetes mellitus: Machine learning models integrated with interpretation framework. *Comput. Biol. Med.* **144**, 105361 (2022).
64. Lundberg, S. M., & Lee, S. I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **30** (2017).
65. Ekanayake, I. U., Meddage, D. P. P. & Rathnayake, U. A novel approach to explain the black-box nature of machine learning in compressive strength predictions

- of concrete using Shapley additive explanations (SHAP). *Case Stud. Constr. Mater.* **16**, e01059 (2022).
66. Zheng, B. et al. An interpretable model-based prediction of severity and crucial factors in patients with COVID-19. *BioMed Res. Int. Biomed. Res. Int.* **2021**, 8840835 (2021).
 67. Meddage, P. et al. Interpretation of machine-learning-based (black-box) wind pressure predictions for low-rise gable-roofed buildings using Shapley additive explanations (SHAP). *Buildings* **12**, 734 (2022).
 68. Mokhtari, K. E., Higdon, B. P. & Başar, A. Interpreting financial time series with SHAP values. In: *Proc. 29th Annual International Conference on Computer Science and Software Engineering* 166–172 (IBM Corp., 2019).
 69. Naikoo, M. W., Rihan, M. & Ishtiaque, M. Analyses of land use land cover (LULC) change and built-up expansion in the suburb of a metropolitan city: Spatio-temporal analysis of Delhi NCR using landsat datasets. *J. Urban Manage.* **9**, 347–359 (2020).
 70. Delhi Development Authority. *List of Unauthorised colonies in Delhi* (2022). Accessed from <https://dda.org.in/planningmap.aspx>.
 71. Johnson, R. A., & Wichern, D. W. *Applied Multivariate Statistical Analysis* 6th edn. (Prentice-Hall, 2007).
 72. Vogiazides, L. & Mondani, H. Neighbourhood trajectories in Stockholm: Investigating the role of mobility and in situ change. *Appl. Geogr.* **150**, 102823 (2023).
 73. Wu, Y. et al. Heat-prone neighbourhood typologies of European cities with temperate climate. *Sustain. Cities Soc.* **87**, 104174 (2022).
 74. Lynge, H. et al. Developing neighbourhood typologies and understanding urban inequality: a data-driven approach. *Reg. Stud. Reg. Sci.* **9**, 618–640 (2022).
 75. Lowe, M. et al. City planning policies to support health and sustainability: an international comparison of policy indicators for 25 cities. *Lancet Glob. Health* **10**, e882–e894 (2022).
 76. Sodiq, A. et al. Towards modern sustainable cities: review of sustainability principles and trends. *J. Clean. Prod.* **227**, 972–1001 (2019).
 77. Bibri, S. E., Krogstie, J. & Kärholm, M. Compact city planning and development: emerging practices and strategies for achieving the goals of sustainability. *Dev. Built Environ.* **4**, 100021 (2020).
 78. Egidi, G., Cividino, S., Vinci, S., Sateriano, A. & Salvia, R. Towards local forms of sprawl: a brief reflection on Mediterranean urbanization. *Sustainability* **12**, 582 (2020).
 79. Debrah, C. et al. A factor analysis of the key sustainability content underpinning green cities development in Ghana. *Int. J. Constr. Manag.* 1–10, <https://doi.org/10.1080/15623599.2022.2068786> (2022).
 80. Knowles, R. D., & Ferbrache, F. (eds.). *Transit Oriented Development and Sustainable Cities* (Edward Elgar Publishing, 2019).
 81. Bibri, S. E. Data-driven smart sustainable cities of the future: an evidence synthesis approach to a comprehensive state-of-the-art literature review. *Sustain. Futures* **3**, 100047 (2021).
 82. Mobaraki, A. & Vehbi, B. O. A conceptual model for assessing the relationship between urban morphology and sustainable urban form. *Sustainability* **14**, 2884 (2022).
 83. Zhao, P., Li, S. & Liu, D. Unequable spatial accessibility to hospitals in developing megacities: new evidence from Beijing. *Health Place* **65**, 102406 (2020).
 84. Gul, Y., Jokhio, G. A. & Bibi, T. Walk towards sustainability: improved neighbourhood street connectivity helps. In *E3S Web Conf.* **211**, 01004 (2020).
 85. Aram, F., Garcia, E. H., Solgi, E. & Mansournia, S. Urban green space cooling effect in cities. *Heliyon* **5**, e01339 (2019).
 86. Sharifi, A., Roosta, M. & Javadpoor, M. Urban form resilience: a comparative analysis of traditional, semi-planned, and planned neighborhoods in Shiraz, Iran. *Urban Sci.* **5**, 18 (2021).
 87. Türkoğlu, H., Terzi, F., Salihoğlu, T., Bölen, F., & Okumuş, G. Residential satisfaction in formal and informal neighborhoods: the case of Istanbul, Turkey. *Int. J. Architectural Res. ArchNet-IJAR* **13**, 112–132 (2019).
 88. Silva, C. A., Giannotti, M. & de Almeida, C. M. Dynamic modeling to support an integrated analysis among land use change, accessibility and gentrification. *Land Use Policy* **99**, 104992 (2020).
 89. Padeiro, M., Louro, A. & da Costa, N. M. Transit-oriented development and gentrification: a systematic review. *Transport Rev.* **39**, 733–754 (2019).
 90. Government of Delhi. Notification on circle rate. *Revenue Department, National Capital Territory of Delhi* (2014). Retrieved from https://srams.delhi.gov.in/docs_avail/c_28ab375a72f800406822233e3497e649_ircle1.pdf.
 91. Lowe, J. S., Prochaska, N. & Keating, W. D. Bringing permanent affordable housing and community control to scale: The potential of Community Land Trust and land bank collaboration. *Cities* **126**, 103718 (2022).
 92. Baobeid, A., Koç, M. & Al-Ghamdi, S. G. Walkability and its relationships with health, sustainability, and livability: elements of physical environment and evaluation frameworks. *Front. Built Environ.* **7**, 721218 (2021).
 93. Xue, H., Cheng, X., Jia, P. & Wang, Y. Road network intersection density and childhood obesity risk in the US: a national longitudinal study. *Public Health* **178**, 31–37 (2020).
 94. U.S. Environmental Protection Agency. *Smart Growth Zoning Codes* (2017). Retrieved from https://www.epa.gov/sites/default/files/2017-06/documents/sm_growth_guide.pdf.
 95. Guida, C. & Carpentieri, G. Quality of life in the urban environment and primary health services for the elderly during the Covid-19 pandemic: an application to the city of Milan (Italy). *Cities* **110**, 103038 (2021).
 96. Mouratidis, K. Urban planning and quality of life: a review of pathways linking the built environment to subjective well-being. *Cities* **115**, 103229 (2021).
 97. Pozoukidou, G. & Chatziyiannaki, Z. 15-Minute City: decomposing the new urban planning utopia. *Sustainability* **13**, 928 (2021).
 98. Alawadi, K., Khanal, A., Mouselly, A. & Aletaywi, A. B. Planning in the age of pandemics: renewing suburban design. *Sustainable Cities Soc.* **87**, 104261 (2022).
 99. Lu, J., Li, B., Li, H. & Al-Barakani, A. Expansion of city scale, traffic modes, traffic congestion, and air pollution. *Cities* **108**, 102974 (2021).
 100. Dulal, H. B. Making cities resilient to climate change: Identifying “win-win” interventions. *Local Environ.* **22**, 106–125 (2017).
 101. Anderson, W. P., Kanargoglou, P. S. & Miller, E. Urban form, energy and the environment: a review of issues, evidence and policy. *Urban Stud.* **33**, 17–35 (1996).
 102. Ewing, R. & Cervero, R. Travel and the built environment: a meta-analysis. *J. Am. Plann. Assoc.* **76**, 265–294 (2010).
 103. Bourdic, L., Salat, S. & Nowacki, C. Assessing cities: a new system of cross-scale spatial indicators. *Build. Res. Inf.* **40**, 592–605 (2012).
 104. Kelobonye, K. et al. Relative accessibility analysis for key land uses: a spatial equity perspective. *J. Transp. Geogr.* **75**, 82–93 (2019b).
 105. Girma, R., Fürst, C. & Moges, A. Land use land cover change modeling by integrating artificial neural network with cellular Automata-Markov chain model in Gidabo river basin, main Ethiopian rift. *Environ. Challenges* **6**, 100419 (2022).
 106. OpenStreetMap contributors. *OpenStreetMap: Delhi*. Retrieved 23 Nov 2022. from <https://www.openstreetmap.org>.
 107. Jain, A. K., Murty, M. N. & Flynn, P. J. Data clustering: a review. *ACM Comput. Surv.* **31**, 264–323 (1999).
 108. Pedregosa, F. et al. Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* **12**, 2825 (2011).
 109. Syakur, M. A., Khotimah, B. K., Rochman, E. M. S. & Satoto, B. D. Integration k-means clustering method and elbow method for identification of the best customer profile cluster. *IOP Conf. Ser.: Mater. Sci. Eng.* **336**, 012017 (2018).
 110. Moustafa, R. E. Parallel coordinate and parallel coordinate density plots. *Wiley Interdiscip. Rev.: Comput. Stat.* **3**, 134–148 (2011).

ACKNOWLEDGEMENTS

This work was supported by the European Union’s Horizon 2020 Research and Innovation Programme under grant agreement No 945307 (eMOTIONAL Cities). We would like to thank H2020 Project “Emotional Cities” for the support during preparation of the paper regarding the discussion of methods and metrics to the development of spatial analysis. This included the transferable methods discussed at the Cambridge Michaelmas’ Workshop: methods and metrics for health and wellbeing – transferable skills for graduate students and early career researchers. The authors would also like to acknowledge the feedback given by the reviewers in enriching the study.

AUTHORS CONTRIBUTION

Study conceptualisation, Data preparation and Modelling, Analysis, and Discussion—A.M.; Analysis, and Review—E.S.

COMPETING INTEREST

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42949-023-00112-1>.

Correspondence and requests for materials should be addressed to Aviral Marwal.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023