Characterizing The Evolution Of Indian Cities Using Satellite Imagery And Open Street Maps

Chahat Bansal
chahat.bansal@cse.iitd.ac.in
Indian Institute of Technology
Delhi, India

Aditi Singla
aditisksingla@gmail.com
Indian Institute of Technology
Delhi, India

Ankit Kumar Singh
cs1170328@iitd.ac.in
Indian Institute of Technology
Delhi, India

Hari Om Ahlawat
mcs182093@iitd.ac.in
Indian Institute of Technology
Delhi, India

Mayank Jain
mcs182011@iitd.ac.in
Indian Institute of Technology
Delhi, India

Prachi Singh
prachi.singh220@gmail.com
Indian Institute of Technology
Delhi, India

Prashant Kumar
mcs182021@iitd.ac.in
Indian Institute of Technology
Delhi, India

Ritesh Saha
cs1170481@iitd.ac.in
Indian Institute of Technology
Delhi, India

Sakshi Taparia
mt1170748@iitd.ac.in
Indian Institute of Technology
Delhi, India

Shailesh Yadav
cs1170489@iitd.ac.in
Indian Institute of Technology
Delhi, India

Aaditeshwar Seth
aseth@cse.iitd.ernet.in
Indian Institute of Technology
Delhi, India

ABSTRACT

With growing urbanization, being able to track urban change is important to plan cities better. Land-use classification of satellite data has been actively used for this purpose. We augment this analysis through the use of crowd-sourced data about the road network in cities, obtained through the Open Street Maps platform. We develop several indicators to quantify the spatial layout of cities and how different localities have changed over time. We apply our methods to study seven Indian cities (Bangalore, Chennai, Delhi, Gurgaon, Hyderabad, Kolkata, and Mumbai) and relate our findings with that of other studies. Our contribution lies in synthesizing two freely available datasets of satellite imagery and road information to develop a series of standardized indicators for different aspects of urbanization, which can serve to compare various cities with one another and to track change happening in the cities over time.

CCS CONCEPTS

• **Computing methodologies → Machine learning**; • **Applied computing** → *Computing in government*.

COMPASS '20, June 15–17, 2020, , Ecuador

ACM ISBN 978-1-4503-7129-2/20/06…$15.00
https://doi.org/10.1145/3378393.3402258

KEYWORDS

Urbanization, Road layout, Built-up areas, Urban Mapping, Satellite Imagery, Open Street Maps

ACM Reference Format:
Chahat Bansal, Aditi Singla, Ankit Kumar Singh, Hari Om Ahlawat, Mayank Jain, Prachi Singh, Prashant Kumar, Ritesh Saha, Sakshi Taparia, Shailesh Yadav, and Aaditeshwar Seth. 2020. Characterizing The Evolution Of Indian Cities Using Satellite Imagery And Open Street Maps. In *ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS) (COMPASS '20), June 15–17, 2020, , Ecuador*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3378393.3402258

1 INTRODUCTION

The World Urbanization Report issued by the United Nations expects the population percentage living in urban areas in India to grow from 34% in 2019 to over 50% in the next three decades, largely driven by an influx of migrants from rural areas [50]. This can create tremendous stress on the urban infrastructure, causing cities to either expand spatially at their peripheries or satellite towns, or have denser construction emerge within cities through high-rises [31, 60, 77]. These dynamics have varying effects on socio-economic development [27, 28, 73]. Growth at the peripheries of cities may increase the commute time and transport expenses for people [2, 64], whereas growth within cities may require an overhaul of the support infrastructure for basic amenities and bring a need for intelligent systems like in smart-cities [34, 51]. It is, therefore, important to understand the urbanization patterns of cities to improve future urban planning.

Urbanization patterns can be quantified using different indicators like the density of construction, area under construction [26], formally Vs. informally developed settlements [40, 58], etc. Such indicators have traditionally been computed through data obtained from field surveys, censuses, topographic maps, city master plans, etc. Such datasets are, however, not uniformly available, which makes it difficult to conduct standardized comparisons between cities. The use of satellite imagery has consequently become quite popular to consistently monitor areas at fine spatio-temporal scales [4, 54]. Crowd-sourced information on roads, building, railways, etc. through platforms like OSM (Open Street Maps) has also been used to understand urban regions [19], although with caveats on the accuracy and completeness of the data [72, 74]. Our contribution lies in the combined use of satellite images and OSM data to build standardized indicators that can help compare cities with one another, to support urban planners, government authorities, and citizens in answering questions such as the following:

- What is the spatial footprint of built-up areas in different cities? Which cities have undergone rapid spatial expansion of their built-up areas?
- What are the central hubs around which different cities are organized?
- How do cities differ in terms of the construction density of their urban settlements? Which cities have the most densely packed settlements?
- How are different urban settlements within a city changing over time?

We use land-use classifiers applied on satellite imagery obtained from the Sentinel-2 satellite system, at a 30m granularity, to track changes in built-up areas over time. We then divide built-up areas into grids of approximately 1 sq.km each and use OSM data of the road network in each grid to build indicators like the density of 3-way intersections, 4-way intersections, and walkability measures, as developed in [38]. A combination of the extent of built-up area and road network indicators helps us to classify these grids into meaningful labels, track how each grid changes over time, and compare cities in terms of their constituent grids. These steps are described in Figure 1.

We apply this method to track changes between 2016 and 2019 in seven Indian cities: Bangalore, Chennai, Delhi, Gurgaon, Hyderabad, Kolkata, and Mumbai. These cities are among the most populous cities in India and represent diversity in terms of new and old cities, and different types of industrialization taking place in these cities. Out of all the state capitals and major industrial cities in India, we choose only those cities for which OSM annotation had peaked at least two years ago and sees only infrequent changes now.

The paper is organized as follows: Section 2 describes related work on methods to study urbanization, Section 3 describes the dataset and pre-processing performed on the satellite imagery and OSM data, Section 4 describes our methodology to compute different urbanization indicators, and section 5 presents the results.

2 RELATED WORK

The study of urbanization includes many topics [9, 17]. Researchers have studied the reasons behind urbanization, such as industrialization and a better provision of social amenities in cities, which lead to inbound migration from rural and other less-developed areas [30, 61, 77]. Studies on the effects of urbanization include

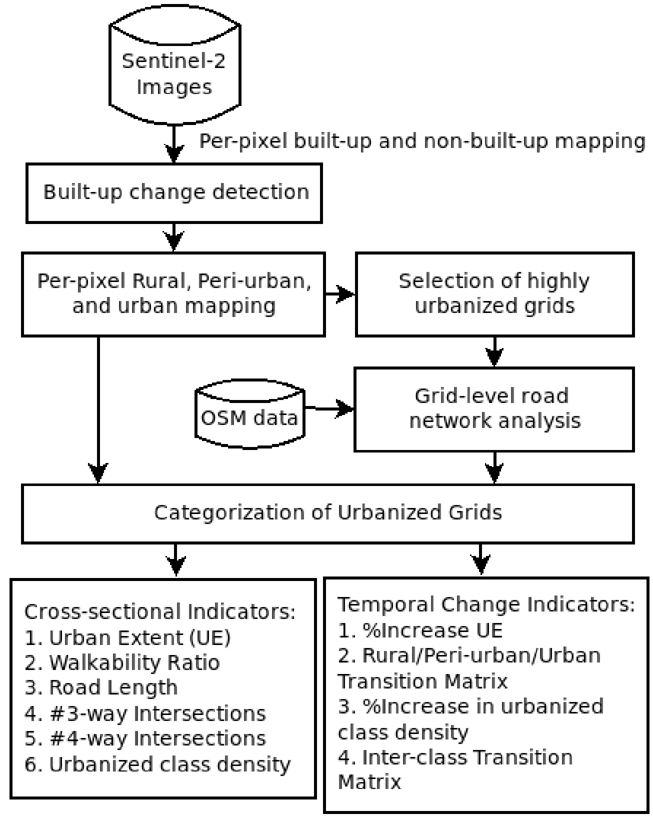


Figure (1) Summary of methodology

growth in urban poverty when large incoming populations engage in low-income work available in urban areas [35], the emergence of segregated neighborhoods of slums, and well-developed settlements [63], and poor living conditions in slums [45, 66]. Urban planning looks at whether cities have adequate resources to handle incoming populations [67], choices between whether to expand spatially to develop suburbs and satellite towns or to create more dense settlements [37], understand the changes that have taken place in cities over time [69], and how these changes relate to economic growth [20, 41, 70].

Our work is a measurement study to quantify the changes taking place in cities in terms of their spatial growth and road infrastructure [38, 39, 57]. It relates most closely with a study by Lamson et al. [38] who compares 200 global cities based on similar indicators. Our approach, however, goes beyond this work: We formulate a standardized methodology that covers the entire city rather than sampling neighborhoods randomly, and we use freely available data sources such as satellite imagery from the Sentinel-2 system and road network information from Open Street Maps. Our methods can, therefore, be used to easily study other cities as well.

Like many other studies, we rely on the use of satellite imagery to understand the spatial layout in cities. With advances in machine learning, a rich body of literature has emerged in the use of this data. This includes the use of deep learning-based models for land-use and land-cover classification [4, 8, 11], and other types of classification methods [22, 44], to gain insights into land-use patterns in cities [55]. These have also been linked to citizen perceptions of livability for different types of spatial layouts [14, 18, 47, 53]. Several individual studies have also been conducted for Indian cities like for Bangalore [46], Kolkata [12, 13], Mumbai [44, 62], Chennai [3],and Pune [33, 56]. However, these city-specific studies make it difficult to compare different cities with one another. Our approach to having developed a standardized method makes it straightforward to run such comparisons across different cities. Further, these studies look into the transition of cities over longer timescales (ten years or more), while our approach works at finer temporal-granularity. Since the temporal transferability of machine learning models is a known issue [10], we take steps to ensure that our methods lead to a robust inference of changes over time.

Although publicly available satellite data can be used for land-use classification, it is not of a sufficiently high resolution to detect roads [15]. The road infrastructure in different neighborhoods can, however, provide useful information about how well planned and developed these neighborhoods are [38]. Our novel contribution lies in building a method to use data from Open Street Maps [25], to develop road-based indicators of urban living. This is a relatively new data source that has mostly been used to map land-use classes [6, 7], identify public properties [32], and construct urban transportation-network models [21]. Our work differs in using the road network information from OSM to develop standardized indicators related to urban living conditions. Although reservation has been expressed about the accuracy and completeness of OSM data [74], we take care to only study cities for which the data seems to be reliable, as explained in Section 3.

3 DATASET

Next, we describe the data acquisition and pre-processing methods we applied to the Sentinel-2 and OSM datasets.

3.1 Mapping Of Changes In Built-Up Areas

The Sentinel-2 system has been capturing thirteen spectral bands at a 10m pixel resolution from across the globe, every 14 days since 2016. We obtained freely available Sentinel-2 data from the Google Earth Engine and applied a land-use classifier developed as part of an on-going study [1]. This classifier is trained on a dataset of 364,000 pixels and generates a 30m resolution map with four land-use classes: water body, green land, barren land, and built-up area. The classifier produces a single classification for each year but takes images from the entire year into account to apply error correcting rules for cases where water bodies may grow or shrink in different months due to seasonal rainfall, or similarly farmland may appear to be barren during non-farming seasons. A robust accuracy of 97% has been reported for the classifier. We aggregate the four land-use classes into two classes, built-up and non-built-up, at a 30m pixel resolution, for each of the four years 2016-2019, and for each of the seven cities that we study. We then develop a 3-class temporal mapping for each pixel:

- *Constantly Built-Up (CBU):* Pixels which remained built-up throughout the four years
- *Constantly Non-Built-Up (CNBU):* Pixels which remained non-built-up throughout the four years
- *Changed:* Pixels which changed from non-built-up to built-up in these four years

Obtaining this 3-class mapping was, however, not as straightforward as just considering a difference between the 2019 value and the 2016 value for each pixel. Through a manually curated dataset of 164 pixels from the district of Gurgaon, we found several cases where errors in the land-use classifier for either of the boundary years could lead to an inaccurate inference of change. Through careful observation of the nature of errors, we found two common sources. First, incorrect classification of pixels could happen because of recurrent shadows of nearby buildings, or spillovers at boundary pixels between different land-use classes, often resulting in salt-and-pepper like noise. This is a known problem with pixel-level classification [75]. Second, sometimes images for an entire year for a large regional footprint (at the scale of an entire large state) could look very different from other regions, possibly because of uncorrected calibration errors in the satellite sensors for that year. We, therefore, developed an error-correcting layer over the temporal mapping obtained for each pixel, as explained next.

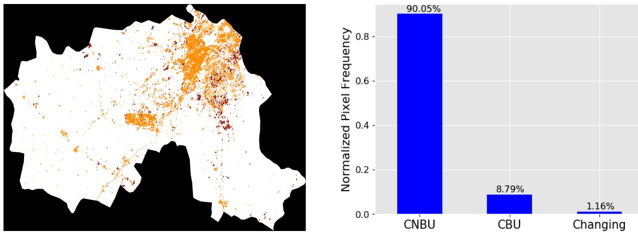
Attaching a value of 1 to built-up pixels and 0 to non-built-up pixels, for each pixel, for each year, we first obtain a new integral value for the pixel in the range [0, 25] by applying a convolution filter using a kernel of size 5x5. For each pixel, we then fit a linear regression across its values for the four years. If the mean squared error of this fit is less than a particular threshold, we mark the pixel as having been constant (CNBU if was non-built-up in 2016 and CBU if it was built-up in 2016), else we mark it as having changed. This method helps us deal with both the salt-and-pepper noise type errors by essentially running a spatial smoothing over each pixel, and also eliminate the effect of a sporadic incorrect classification due to satellite calibration problems in some year. The details of this method is covered in the supplementary material [16]. It gave us an overall accuracy of 93.90% over the ground-truth of 164 pixels (manually created for Gurgaon), and a class-level accuracy as follows: (95.74 precision, 95.74 recall) for CBU over 47 pixels, (92.22 precision, 98.80 recall) for CNBU over 84 pixels, and (96.30 precision, 78.78 recall) for 33 changing pixels. This is shown in Figure 2 for Gurgaon. We then applied this same method to other cities.

3.2 Processing Of OSM Data For Road-Network Indicators

Data from OSM is organized in terms of nodes and ways:

- *Nodes:* These are (latitude, longitude) markings that are labeled on the ways.
- *Ways:* These are an ordered list of nodes to mark railways, roads, intersections on roads, bounding areas of buildings, etc.

OSM tags are used to identify *ways*, which are roads, and *nodes* that are common on multiple roads and are identified as intersections. After downloading the entire OSM data for each city, we are thus able to build an undirected graph of roads and intersections of the city's road network. Since OSM data can be incomplete, we choose only those cities for our analysis for which OSM annotation had peaked at least two years ago and sees only infrequent changes



(a) (b)

Class Label	Precision	Recall
CBU	95.74	95.74
CNBU	92.22	98.80
Changing	96.30	78.78
Accuracy: 93.90%		

(c)

Figure (2) (a) Pixel map, (b) Frequency distribution, and (c) Accuracy of pixel-level temporal mapping in Gurgaon: Constantly built-up, Constantly non-built-up, Changed

now. We are able to estimate this by checking the last-modified-time on each node and way annotation in the OSM data.

4 METHODOLOGY

Next, we describe the different indicators we develop based on the satellite and OSM data.

4.1 Built-Up Footprint In A City

The amount of built-up pixels in a city's satellite image is an indicator of human development activities (like the construction of residential buildings, commercial structures, roads, etc.) in the city [23]. We use the CBU/CNBU/Changed pixel classification to develop built-up maps for 2016 and 2019. CBU pixels denote built-up areas as of 2016, and CBU along with the Changed pixels denote built-up areas as of 2019. A map of the built-up pixels for Delhi in 2019 is shown in Figure 3b.

We next follow the method developed by Angel et al. [5] and Subasinghe et al. [65] to reclassify the 2016 and 2019 pixel maps into urban (high density of built-up structures), peri-urban (moderate density of built-up structures), or rural (low density or no built-up structures) pixels. This is done by considering a *walking distance circle* (WDC) around each pixel - a circle of 1 Km^2, which comes to an approximately 584-meter radius, denoting a ten-minute walking distance. Pixels with at least 50% built-up pixels in their WDC are said to be urban pixels, those with 25-50% built-up pixels in their WDC are said to be peri-urban pixels, and rural pixels are those with less than 25% built-up pixels in their WDC. This is shown in Figure 3c for Delhi. The urban and peri-urban pixels together define the urban extent (UE) in a city [9, 59], to develop the following indicator:

$$UE = \frac{\#Urban\_pixels + \#Periurban\_pixels}{\#Total\_pixels} \quad (1)$$

4.2 Identification Of Urbanized Grids

We then divide the city into grids of 0.01° latitudinal and 0.01° longitudinal widths. Each grid ends up containing around 1000

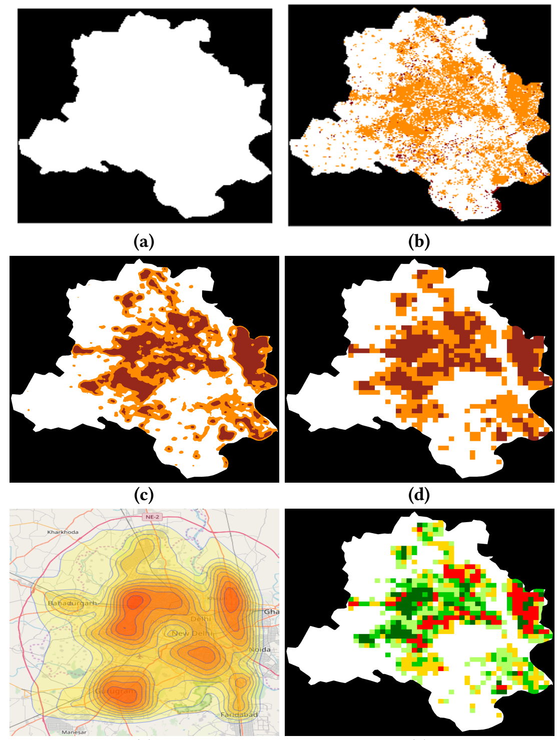
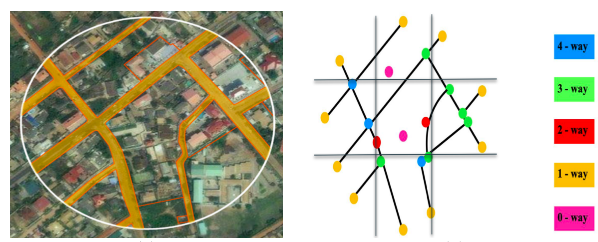


Figure (3) Analysis pipeline shown for Delhi 2019 as an example: (a) Administrative boundary, (b) CBU/CNBU/Changed pixels, (c) Urban/peri-urban/rural pixels, (d) Urban and peri-urban grids, (e) heatmap of road length, (f) categorization of different types of urban grids



(a) (b)

Figure (4) Graph representation of an OSM road network [38], with different types of intersections

pixels with a combined area of approximately 1Km^2, which is one-third of the average size of a village in India. A grid is selected for further analysis of its road network only if more than 50% of its pixels are either urban or peri-urban.

4.2.1 ***Road-based indicators in urbanized grids.*** For each urbanized grid as identified above, we then obtain the road network graph in that grid using the OSM data. Figure 4a shows a sample graph representation. The graph in each grid is used to develop several road-based indicators.

4.2.2 *Type of road intersections.* Road intersections are junctions where at least two roads cross each other and can be identified based on the degree of the nodes. Nodes with degrees 0, 1, and 2 are ignored as they represent isolated points, dead ends, and points defining the path of a single road, respectively. Nodes of interest are those with degrees three or four. 4-way intersections are considered as representing formal development in a city, typically arising from a block-type urban layout [60]. 3-way intersections, on the other hand, typically denote informal development and are said to lead to traffic congestion [38]. We are able to calculate for each grid the number of 3-way and 4-way intersections in the grid.

4.2.3 Road length. Central parts of a city to which many roads converge, tend to have a dense road presence. Being able to calculate the sum of lengths of roads in a grid can thus help determine whether the grid lies in a central part of the city or not. We call this value the road-length in a grid, and visualizing a heat-map of road-lengths can reveal whether a city is monocentric or polycentric, i.e., whether the city is organized around one central hub or multiple hubs. Figure 3e shows the example of Delhi as organized around multiple hubs.

To compute the road-length using the undirected road-network graph within each grid, we assign to each edge a length equal to the distance between the two nodes that it connects. For edges having nodes in adjacent grids, a linear interpolation is done to calculate the length of the road falling within the grid under consideration. The sum of these lengths is then taken as the road-length within a grid, and heat-map visualizations are used to determine different hubs in a city. Geodesic distances are used for all road lengths.

4.2.4 Walkability ratio. We use the method by Lamson et al. [38] to calculate a walkability ratio for each grid. The walkability ratio indicates the ease with which people can reach different parts of an area by walking - a high walkability ratio is seen in well developed residential colonies with a good road network, and a low walkability ratio is seen in industrial areas.

We calculate the walkability ratio by picking 50 pairs of points randomly within a grid. For each pair, a beeline distance is calculated between them through a straight line, and a street travel distance is calculated by traversing the shortest-path on the road network from the roads closest to these points. The road network in the eight surrounding grids is also considered for the shortest-path computation, to allow the traversal of roads that may go outside the grid as well. The mean of the ratios of beeline to street travel distances for all the pairs is then calculated to obtain the walkability ratio for the grid. This is shown in Figure 5.

4.3 Categorization Of Urbanized Grids In Indian Cities

Since we are able to build indicators for each grid based on its urban extent, 3-way intersections, 4-way intersections, and walkability ratio, we next attempt to categorize the grids into different coarse types. We perform hierarchical clustering on vectors of normalized values of these indicators. Instead of using the urban extent indicator, we use the ratio of urban pixels to total pixels in a grid, which we call the urban footprint of a grid, because this variable gives us more distinctive classes. We eventually obtain 5 clusters, as shown with their box-plots in Figure 6. We attach the following interpretation to these clusters:

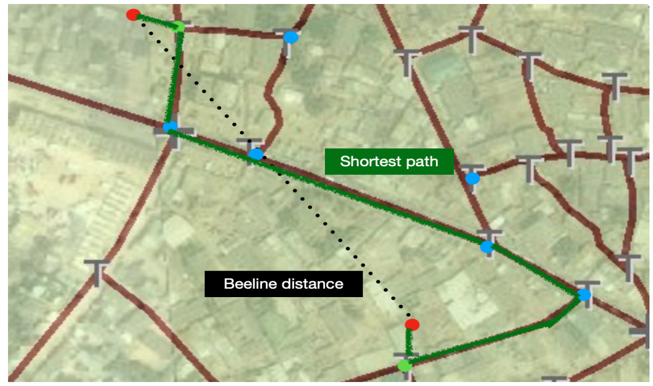


Figure (5) Walkability ratio computation: Beeline distance and street travel distance

(1) *Class I:* Sparse settlements with less road infrastructure, which seem to include non-residential areas like industrial zones and airports, or upcoming residential areas.
(2) *Class II:* Sparse settlements with sufficient road infrastructure, which also seem to be upcoming residential areas.
(3) *Class III:* Moderately dense settlements with proportionately sufficient road infrastructure, which seem to be well developed residential areas.
(4) *Class IV:* Dense settlements with a proportionately sufficient road infrastructure, which too seem to be well developed residential areas but more formally planned than Class III areas.
(5) *Class V:* Very dense settlements with inadequate road infrastructure, which seem to be slum resettlement areas.

Figure 3f shows the five types of grids in Delhi. Further, we can also calculate the Class Density (CD) for each class in a city:

$$CD_i = \frac{\#C_i\_Grids}{\sum_{n=1}^{5} \#C_i\_Grids} \tag{2}$$

4.4 Change Based Indicators

All the indicators we have listed above can be computed separately for 2016 and 2019, and thus allow us to track the change between these years. The different change-based indicators we compute are as follows:

4.4.1 Percentage increase in the urban extent of a city. If UE_{2016} and UE_{2019} denote the urban extent of a city in 2016 and 2019, respectively, the rate of increase denoting the rise in construction activity can be calculated as:

$$Increase\_UE = \frac{UE_{2019} - UE_{2016}}{UE_{2016}} \tag{3}$$

Further disaggregation into studying the transition from rural to peri-urban or peri-urban to urban pixels can lead to more nuanced indicators as well.

4.4.2 Transition matrix for class I-V urbanized grids. We can similarly study the transition between the five types of urbanized grids. For example, a conversion of C1 or C2 into C3 or C4 grids could indicate an increase in the density of residential construction in the area. Similarly, an increase in the density of C5 grids could

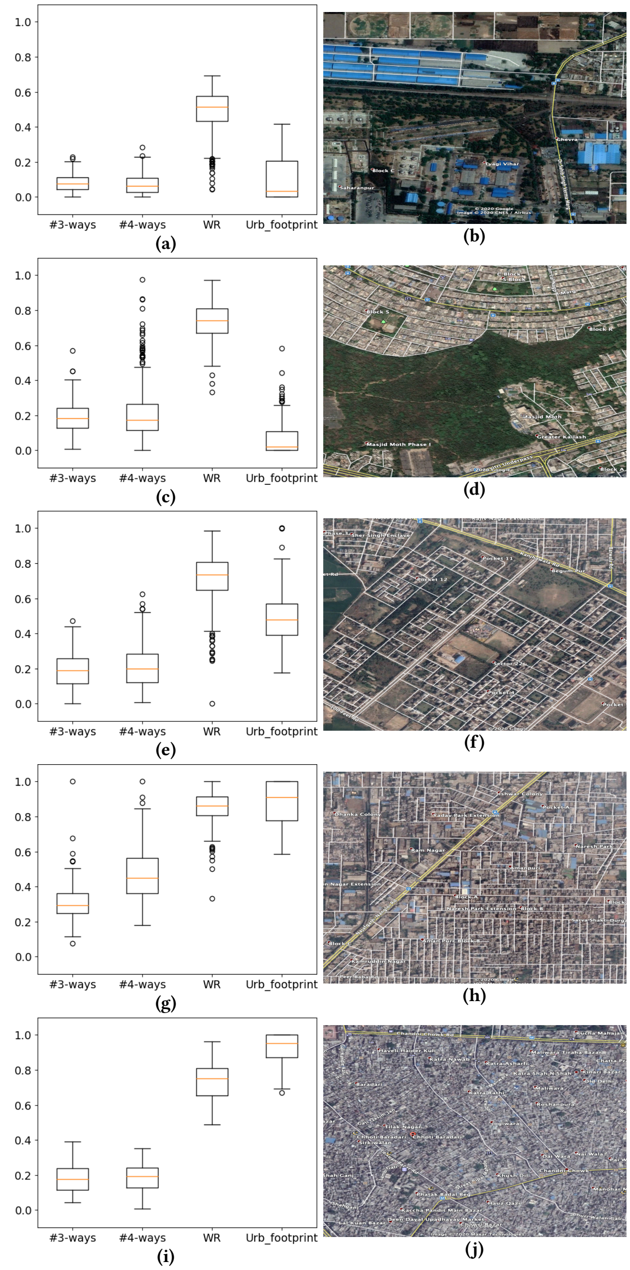


Figure (6) Boxplots for different types of urban grids, alongside images of representative examples from Delhi in 2019. (a)(b) - Class I, (c)(d) - Class II, (e)(f) - Class III, (g)(h) - Class IV, and (i)(j) - Class V

raise warnings about an increase in areas with inadequate road infrastructure and thereby poor living conditions. The increase in the class density (CD), of class 'i', can be calculated as:

$$Increase\_CD_i = \frac{CD_{i2019} - CD_{i2016}}{CD_{i2016}} \tag{4}$$

To dig deeper, we also create a transition matrix for all grids in a city. Each cell C_{ij} of this matrix denotes the number of grids converting from class C_i to class C_j, between 2016 and 2019. We also add a "NULL" row for 2016, to include grids that were not counted as urbanized grids in 2016 but were counted in 2019.

Note that the OSM data does not carry any timestamps of when the roads were developed, therefore for any urbanized grid that we count in 2016, we assume that the road network was developed at the same time as the built-up structures came up. This is not a bad assumption at least for most residential areas because roads and plots are typically cut in advance when areas are opened up for purchase by future homeowners. The actual housing structures come up later. Therefore, by the time a grid acquires a sufficient density of built-up pixels for it to be counted as an urbanized grid, the road network is likely to be present there already.

5 RESULTS AND ANALYSIS

Next, we use the indicators developed in the previous section to answer several questions about the seven cities that we examine. We try to triangulate our observations with other studies, and present an analysis at two levels: (a) spatial layout of the cities in 2019, and (b) changes in the layout between 2016 and 2019.

5.1 Urban patterns in 2019

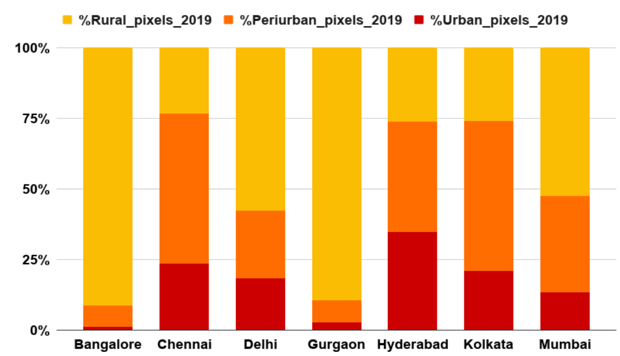


Figure (7) Distribution of rural, peri-urban, and urban pixels for cities, in 2019

5.1.1 ***What is the urban extent of different cities?*** We defined urban extent as the fraction of land in a district that has a reasonably high density of construction for it to be considered as supporting an urban or peri-urban settlement. From Figure 7, we can see that Chennai, Kolkata, and Hyderabad have the greatest urban extent, followed by Mumbai, Delhi, Gurgaon, and Bangalore. On visualizing Figure 8h,8k,8j, &8n, it is evident that this second set of districts are surrounded by forests or agricultural land, or have large forests and parks in the city itself. Partly this is also because these districts are much larger in area (see Supplementary Material [16]).

5.1.2 ***How do cities differ in terms of the density of their urban settlements?*** Figure 9 shows the distribution of areas within a city into the five types of urbanized grids. We look at the aggregate density of C3, C4, and C5 grids as those which have high density of built-up infrastructure. We can see that Hyderabad is highly dense, followed by Delhi, Chennai, Kolkata, Mumbai, Gurgaon, and Bangalore. It is interesting to note that Hyderabad not only has a high urban extent, it also has a high density of settlements, indicating

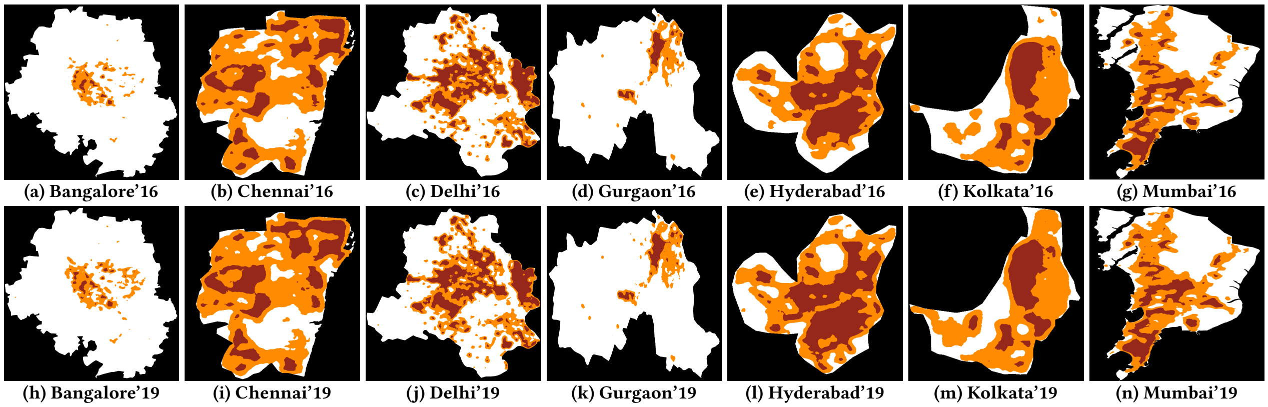


Figure (8) City-wise urban extent in 2016 and 2019

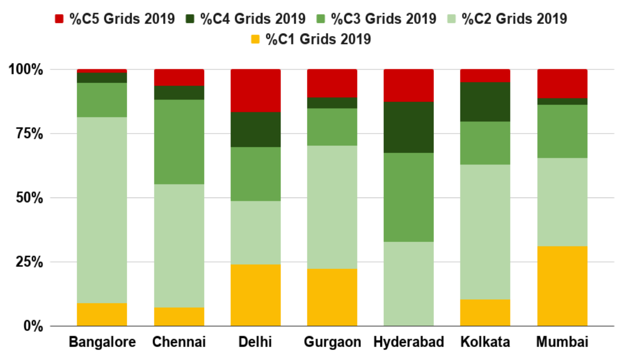


Figure (9) C1-C5 class density for cities, in 2019

that the city has limited room for expansion. Chennai also has a high urban extent, but not a very high density in its settlements.

5.1.3 ***What are the central hubs around which cities are organized?*** Based on a heatmap visualization of road-lengths, Figure 12 shows that Delhi is highly polycentric having multiple urban hubs, followed by Mumbai and Hyderabad. This is confirmed by Taubenböck et al. [68], who explains that Mumbai became polycentric due to a shortage of land and led to the development of Navi Mumbai through land reclamation. In the case of Delhi, the emergence of satellite cities (Gurgaon, Noida, Ghaziabad, and Faridabad) led to the development of new areas in the direction of these cities, and several slum resettlement projects also led to the urbanization of peripheral areas of Delhi, leading to its highly polycentric structure. In comparison, Chennai and Kolkata seem to have developed in two distinct parts, while Bangalore and Gurgaon have mostly grown around a center [42].

5.1.4 ***Which cities have the most formally laid road infrastructure?*** Urbanized grids with a high number of 4-way intersections indicate a formally developed road network [38]. We, therefore, look at the density of C4 grids in the cities. Figure 9 shows that Hyderabad has a large percentage of area under C4 grids, in concurrence with observations about its ongoing planned development [29]. This is followed by Kolkata and Delhi, both of which moved towards privatization of city planning and efforts like Special Economic Zones (SEZ) for fast-tracked infrastructural development [36]. Cities like Chennai, Bangalore, Gurgaon, and Mumbai lie at the lowest end of the spectrum.

5.1.5 ***Which cities have a large presence of densely packed areas that lack adequate road infrastructure?*** The density of C5-grids is an indicator of areas that are densely packed and also lack an adequate road infrastructure. Figure 9 shows that Delhi has a high density of these grids, and is closely followed by Mumbai and Hyderabad. This is probably why Delhi lies on the bottom of the Mercer's list of Quality of Living [43] because living conditions deteriorate in extremely congested areas with poor infrastructural facilities. These parts of Delhi include areas like Shahdara, Bhalswa, Azadpur, and most parts of old Delhi, which are known to be heavily congested. Mumbai is also heavily congested, having to deal with almost 5-times the traffic as Delhi but on a much lesser road density [71]. Even Gurgaon, despite being a newer city with heavy industrial development, has over 11% of its urbanized grids lacking adequate road infrastructure [49]. This is probably because villages close to the industrial areas have a history of rapidly converting themselves into residential hubs that provide poor quality rental accommodation to industrial workers, most of whom are migrants from rural areas in search of jobs [48].

5.2 Changes In Urban Pattern From 2016-2019

We next look at changes that have happened in the cities between 2016 to 2019.

5.2.1 ***Which cities have undergone rapid spatial expansion in built-up areas?*** We track which rural pixels in 2016 changed into peri-urban or urban pixels in 2019, denoting the spatial expansion of cities. Figure 10 shows that Bangalore and Delhi saw the largest amount of conversion of rural land into urban settlements, in terms of the absolute number of square kilometers converted. Bangalore indeed has recently been ranked as the third fastest growing city in the world [76]. This sprawl is witnessed mainly in the northern, north-eastern, and south-eastern parts of the city (refer Figure 13a & 13h), where the main IT (Information Technology) hubs of Bangalore are located [56].

5.2.2 ***How are different urban settlements within a city changing over time?*** We next study how grids of different types transition over the years. Figure 11 shows the change in the grid density

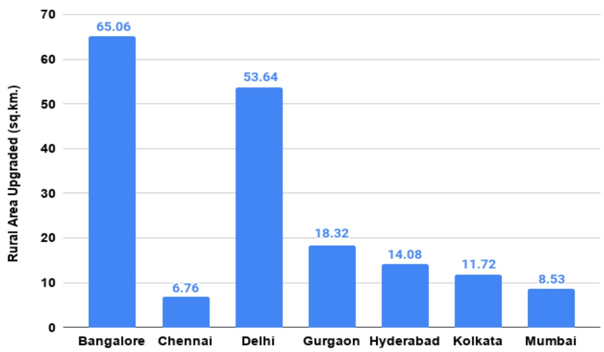


Figure (10) Rural areas (in sq.km) that converted to urban or periurban areas between 2016 to 2019

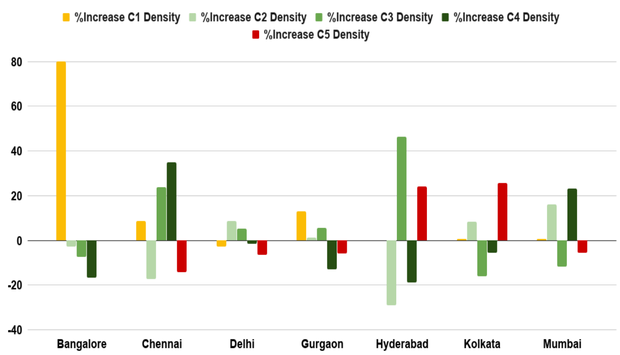


Figure (11) Percentage change in C1-C5 class densities between 2016 and 2019

2016 | 2019	C1	C2	C3	C4	C5
C1	97	26	11	0	0
C2	3	105	17	0	0
C3	5	1	92	5	6
C4	0	2	0	70	4
C5	0	0	3	6	88
NULL	36	13	1	0	0

(a) Delhi

2016 | 2019	C1	C2	C3	C4	C5
C1	4	1	0	0	0
C2	1	69	7	0	0
C3	0	0	12	1	2
C4	0	0	0	5	0
C5	0	0	0	0	0
NULL	8	38	1	0	0

(b) Bangalore

Table (1) Transition matrix for changes in C1 to C5 urbanized grids

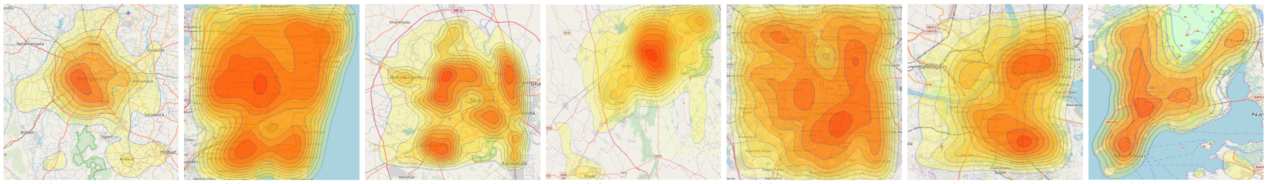
for each type. The cities like Bangalore, Gurgaon, and Chennai have a net increase in their C1-grids, showing that these grids indicate emerging settlements after the year 2016. Table 1a shows the transition matrix for Delhi as an example, revealing that most non-urbanized grids converted into C1 grids over these years. Similarly, the transition matrix for Bangalore(Table 1b) shows that most non-urbanized grids converted directly into C2 grids, implying a faster rate of urbanization than Delhi. A deeper analysis into C1 to C2/C3 transitions shows that Delhi outpaces Bangalore because the C1 areas in Delhi are mostly the ones under new development which seem to grow faster than the C1 areas in Bangalore that are mostly industrial and did not change much over this short period of time. Another interesting pattern is that in cities like Bangalore, Delhi, and Gurgaon, the proportion with which non-urbanized grids are getting urbanized (i.e. NULL to C1/C2), the density of more congested grids like C4 and C5 are going down. This construction activity can be in response to the ongoing efforts of Jawaharlal Nehru National Urban Renewal Mission (JNNURM) of building houses for the economically weaker sections in these cities for the elevation of poor (like in Bhorgarh and PoothKhurd, Delhi [52]). In other cities like Chennai, Mumbai, Hyderabad, and Kolkata, the increase in the density of either C4 or C5 grids reveals an infilling pattern of urbanization which is making them more congested over the years.

6 CONCLUSION AND DISCUSSION

In this paper, we demonstrated how freely available satellite imagery from the Sentinel-2 system and road network information from Open Street Maps can be used to develop useful indicators to study the urbanization of cities. These indicators can be added to the indicator-list related with infrastructural development in different districts of India [24]. Our work goes beyond state of the art in having developed a standardized methodology that makes it possible to compare different cities with one another, track changes at fine spatial scales across the entire city, and derive a nuanced understanding of the nature of these changes. We applied this methodology on seven large cities of India and found interesting observations, most of which we were able to triangulate with observations made in other studies using different methods, lending veracity to our methodology. A limitation of using OSM data is that it may not be very complete or accurate for all cities, and although we took precautions in selecting only those cities for which the data seemed reliable, we may not be able to use OSM data for just about any city. Alternatives like Google Maps have a paid API and are likely to be more complete, and our methods can be easily extended to them as well. Another challenge we faced is in the temporal transferability of land-use classification models on satellite imagery, to apply them on years different from the ones on which they were trained. We developed a method to obtain reliable results from across different years, and we are attempting to evaluate it on more ground-truth data that we are manually curating. As part of future work, we are also attempting to use data from the mass media about news articles in which different areas of the city may have been mentioned, so that plausible qualitative explanation can be drawn about the differences between these areas and changes taking place therein.

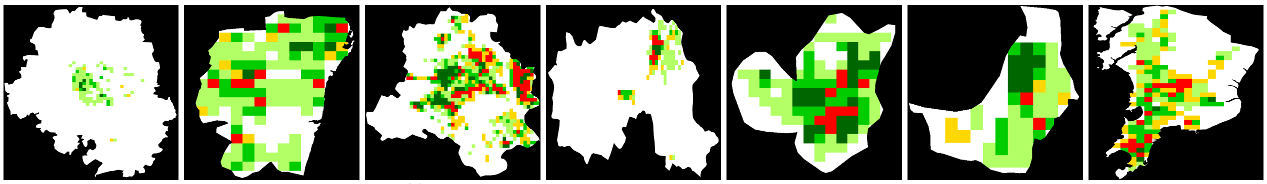
REFERENCES

[1] Hari Om Ahlawat. 2020. An open dataset for landuse classification in India for Sentinel-2. https://github.com/hariomahlawat/An-open-dataset-for-landuse-classification-in-India-for-Sentinel-2

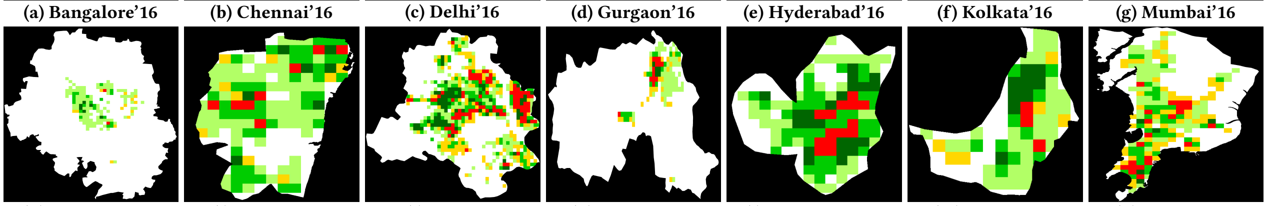


(a) Bangalore (b) Chennai (c) Delhi (d) Gurgaon (e) Hyderabad (f) Kolkata (g) Mumbai

Figure (12) City-wise heat-maps based on road length, in 2019



(a) Bangalore'16 (b) Chennai'16 (c) Delhi'16 (d) Gurgaon'16 (e) Hyderabad'16 (f) Kolkata'16 (g) Mumbai'16



(h) Bangalore'19 (i) Chennai'19 (j) Delhi'19 (k) Gurgaon'19 (l) Hyderabad'19 (m) Kolkata'19 (n) Mumbai'19

Figure (13) City-wise visualizations of different types of grids: (a)-(g) Distribution of C1-C5 urbanized grids in 2016, (h)-(n) Distribution of C1-C5 urbanized grids in 2019. Legend- C1 grids C2 grids C3 grids C4 grids C5 grids

[2] Rumi Aijaz. 2019. India's Peri-Urban Regions: The Need for Policy and the Challenges of Governance. (2019).

[3] Bharath H Aithal and TV Ramachandra. 2016. Visualization of urban growth pattern in Chennai using geoinformatics and spatial metrics. *Journal of the Indian Society of Remote Sensing* 44, 4 (2016), 617–633.

[4] Adrian Albert, Jasleen Kaur, and Marta C Gonzalez. 2017. Using convolutional networks and satellite imagery to identify patterns in urban environments at a large scale. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. ACM, 1357–1366.

[5] Shlomo Angel, Alejandro M Blei, Daniel L Civco, and Jason Parent. 2012. *Atlas of urban expansion*. Lincoln Institute of Land Policy Cambridge, MA.

[6] Jamal Jokar Arsanjani, Peter Mooney, Alexander Zipf, and Anne Schauss. 2015. Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets. In *OpenStreetMap in GIScience*. Springer, 37–58.

[7] Nicolas Audebert, Bertrand Le Saux, and Sébastien Lefèvre. 2017. Joint learning from earth observation and openstreetmap data to get faster better semantic maps. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 67–75.

[8] Nicolas Audebert, Bertrand Le Saux, and Sébastien Lefèvre. 2018. Beyond RGB: Very high resolution urban remote sensing with multimodal deep networks. *ISPRS Journal of Photogrammetry and Remote Sensing* 140 (2018), 20–32.

[9] Deborah Balk, Stefan Leyk, Bryan Jones, Mark R Montgomery, and Anastasia Clark. 2018. Understanding urbanization: A study of census and satellite-derived urban classes in the United States, 1990-2010. *PloS one* 13, 12 (2018).

[10] Chahat Bansal, Arpit Jain, Phaneesh Barwaria, Anuj Choudhary, Anupam Singh, Ayush Gupta, and Aaditeshwar Seth. 2020. Temporal Prediction of Socio-economic Indicators Using Satellite Imagery. In *Proceedings of the 7th ACM IKDD CoDS and 25th COMAD*. 73–81.

[11] Saikat Basu, Sangram Ganguly, Supratik Mukhopadhyay, Robert DiBiano, Manohar Karki, and Ramakrishna Nemani. 2015. Deepsat: a learning framework for satellite imagery. In *Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems*. ACM, 37.

[12] Basu Bhatta. 2009. Analysis of urban growth pattern using remote sensing and GIS: a case study of Kolkata, India. *International Journal of Remote Sensing* 30, 18 (2009), 4733–4746.

[13] Basudeb Bhatta, S Saraswati, and D Bandyopadhyay. 2010. Quantifying the degree-of-freedom, degree-of-sprawl, and degree-of-goodness of urban growth from remote sensing data. *Applied Geography* 30, 1 (2010), 96–111.

[14] Gabriel Cadamuro, Aggrey Muhebwa, and Jay Taneja. 2018. Assigning a Grade: Accurate Measurement of Road Quality Using Satellite Imagery. *arXiv preprint arXiv:1812.01699* (2018).

[15] Gabriel Cadamuro, Aggrey Muhebwa, and Jay Taneja. 2019. Street smarts: measuring intercity road quality using deep learning on satellite imagery. In *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*. 145–154.

[16] Ankit Kumar Singh Hari Om Ahlawat Mayank Jain Prachi Singh Prashant Kumar Ritesh Saha Sakshi Taparia Shailesh Yadav Chahat Bansal, Aditi Singla and Aaditeshwar Seth. 2020. Supplementary Material: Characterizing the Evolution of Indian Cities using Satellite Imagery and Open Street Maps. https://bit.ly/3c4qRUO

[17] Barney Cohen. 2006. Urbanization in developing countries: Current trends, future projections, and key challenges for sustainability. *Technology in society* 28, 1-2 (2006), 63–80.

[18] Marco De Nadai, Radu Laurentiu Vieriu, Gloria Zen, Stefan Dragicevic, Nikhil Naik, Michele Caraviello, Cesar Augusto Hidalgo, Nicu Sebe, and Bruno Lepri. 2016. Are safer looking neighborhoods more lively?: A multimodal investigation into urban life. In *Proceedings of the 24th ACM international conference on Multimedia*. ACM, 1127–1135.

[19] Paolo Fogliaroni, Dominik Bucher, Nikola Jankovic, and Ioannis Giannopoulos. 2018. Intersections of our world. In *10th International Conference on Geographic Information Science*, Vol. 114. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 3.

[20] Hildebrand Frey. 2003. *Designing the city: Towards a more sustainable urban form*. Taylor & Francis.

[21] Jorge Gil. 2015. Building a multimodal urban network model using OpenStreetMap data for the analysis of sustainable accessibility. In *OpenStreetMap in GIScience*. Springer, 229–251.

[22] Ran Goldblatt, Michelle F Stuhlmacher, Beth Tellman, Nicholas Clinton, Gordon Hanson, Matei Georgescu, Chuyuan Wang, Fidel Serrano-Candela, Amit K Khandelwal, Wan-Hwa Cheng, et al. 2018. Using Landsat and nighttime lights for supervised pixel-based image classification of urban land cover. *Remote Sensing of Environment* 205 (2018), 253–275.

[23] Ran Goldblatt, Wei You, Gordon Hanson, and Amit K Khandelwal. 2016. Detecting the boundaries of urban areas in india: A dataset for pixel-based image classification in google earth engine. *Remote Sensing* 8, 8 (2016), 634.

[24] Dibyajyoti Goswami, Shyam Bihari Tripathi, Sansiddh Jain, Shivam Pathak, and Aaditeshwar Seth. 2019. Towards building a district development model for india using census data. In *Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies*. 259–271.

[25] Mordechai Haklay and Patrick Weber. 2008. Openstreetmap: User-generated street maps. *IEEE Pervasive Computing* 7, 4 (2008), 12–18.

[26] Meg Holden. 2006. Urban indicators and the integrative ideals of cities. *Cities* 23, 3 (2006), 170–183.

[27] Bert F Hoselitz. 1957. Urbanization and economic growth in Asia. *Economic Development and Cultural Change* 6, 1 (1957), 42–54.

[28] Ganlin Huang and Yaqiong Jiang. 2017. Urbanization and Socioeconomic Development in Inner Mongolia in 2000 and 2010: A GIS Analysis. *Sustainability* 9, 2 (2017), 235.

[29] Janapriya. 2017. What Makes Hyderabad the Best City to Live in? https://www.janapriya.com/blog/what-makes-hyderabad-the-best-city-to-live-in/

[30] Remi Jedwab, Luc Christiaensen, Marina Gindelsky, et al. 2014. Rural push, urban pull and... urban push? New historical evidence from developing countries. *Institute for International Economic Policy: Washington, DC, USA* (2014).

[31] Ramanath Jha and Sayli Udas-Mankikar. 2019. India's Urban Challenges: Recommendations for the New Government (2019-2024). (2019).

[32] Mohsen Kalantari and Veha La. 2015. Assessing OpenStreetMap as an open property map. In *OpenStreetMap in GIScience*. Springer, 255–272.

[33] Lakshmi N Kantakumar, Shamita Kumar, and Karl Schneider. 2016. Spatiotemporal urban expansion in Pune metropolis, India using remote sensing. *Habitat International* 51 (2016), 11–22.

[34] Jeongseob Kim. 2016. Achieving mixed income communities through infill? The effect of infill housing on neighborhood income diversity. *Journal of Urban Affairs* 38, 2 (2016), 280–297.

[35] Divyani Kohli, Richard Sliuzas, and Alfred Stein. 2016. Urban slum detection using texture and spatial metrics derived from satellite imagery. *Journal of spatial science* 61, 2 (2016), 405–426.

[36] Ashok Kumar. 2006. Trends of planning and governance in metropolitan India. *Institute of Town Planners India* 3, 2 (2006), 10–20.

[37] Danielle Labbé and Julie-Anne Boudreau. 2011. Understanding the causes of urban fragmentation in Hanoi: the case of new urban areas. *International Development Planning Review* 33, 3 (2011), 273–291.

[38] Patrick Lamson-Hall, Shlomo Angel, Alejandro Blei, Manuel Madrid, and Nicolas Galarza. 2016. The Quality of Urban Layouts.

[39] Hongbo Li, Yali Liu, and Kaili Peng. 2018. Characterizing the relationship between road infrastructure and local economy using structural equation modeling. *Transport Policy* 61 (2018), 17–25.

[40] Laura Lutzoni. 2016. In-formalised urban space design. Rethinking the relationship between formal and informal. *City, Territory and Architecture* 3, 1 (2016), 1–14.

[41] Evert Meijers. 2008. Summing small cities does not make a large city: polycentric urban regions and the provision of cultural, leisure and sports amenities. *Urban Studies* 45, 11 (2008), 2323–2342.

[42] Michele Melchiorri, Aneta Florczyk, Sergio Freire, Daniele Ehrlich, Marcello Schiavina, Martino Pesaresi, and Thomas Kemper. 2018. Megacities spatiotemporal dynamics monitored with the Global Human Settlement Layer. In *REAL CORP 2018–EXPANDING CITIES–DIMINISHING SPACE. Are "Smart Cities" the solution or part of the problem of continuous urbanisation around the globe? Proceedings of 23rd International Conference on Urban Planning, Regional Development and Information*. CORP–Compentence Center of Urban and Regional Planning, 285–294.

[43] Mercer. 2019. Hyderabad and Pune retail top rankings amongst Indian cities in Mercer's 21st Quality Of Living. https://www.mercer.co.in/newsroom/2019-quality-of-living-survey.html

[44] Hossein Shafizadeh Moghadam and Marco Helbich. 2013. Spatiotemporal urbanization processes in the megacity of Mumbai, India: A Markov chains-cellular automata urban growth model. *Applied Geography* 40 (2013), 140–149.

[45] Frederick Mugisha. 2006. School enrollment among urban non-slum, slum and rural children in Kenya: Is the urban advantage eroding? *International Journal of Educational Development* 26, 5 (2006), 471–482.

[46] Harini Nagendra, Suparsh Nagendran, Somajita Paul, and Sajid Pareeth. 2012. Graying, greening and fragmentation in the rapidly expanding Indian city of Bangalore. *Landscape and Urban Planning* 105, 4 (2012), 400–406.

[47] Nikhil Naik, Ramesh Raskar, and César A Hidalgo. 2016. Cities are physical too: Using computer vision to measure the quality and impact of urban appearance. *American Economic Review* 106, 5 (2016), 128–32.

[48] Vishal Narain. 2009. Growing city, shrinking hinterland: land acquisition, transition and conflict in peri-urban Gurgaon, India. *Environment and Urbanization* 21, 2 (2009), 501–512.

[49] Sanjoy Narayan. 2015. The 10 fastest-growing cities in the world are all in India. https://www.hindustantimes.com/columns/gurgaon-is-an-example-of-how-not-to-urbanise-india/story-KqAcFBWI8jp62fCvKTEPwK.html

[50] United Nations. 2018. 2018 revision of world urbanization prospects. https://population.un.org/wup/

[51] Peter Newton and Stephen Glackin. 2014. Understanding infill: Towards new policy and practice for urban regeneration in the established suburbs of Australia's cities. *Urban policy and research* 32, 2 (2014), 121–143.

[52] Government of NCT of Delhi. 2020. Economic Survey of Delhi 2016-17. http://delhiplanning.nic.in/content/economic-survey-delhi-2016-17

[53] Barak Oshri, Annie Hu, Peter Adelson, Xiao Chen, Pascaline Dupas, Jeremy Weinstein, Marshall Burke, David Lobell, and Stefano Ermon. 2018. Infrastructure quality assessment in africa using satellite imagery and deep learning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 616–625.

[54] Bhartendu Pandey and Karen C Seto. 2015. Urbanization and agricultural land loss in India: Comparing satellite estimates with census data. *Journal of environmental management* 148 (2015), 53–66.

[55] Sanoj Kumar Patel, Pramit Verma, and Gopal Shankar Singh. 2019. Agricultural growth and land use land cover change in peri-urban India. *Environmental monitoring and assessment* 191, 9 (2019), 600.

[56] TV Ramachandra, Bharath H Aithal, and Barik Beas. 2014. Urbanisation pattern of incipient mega region in india. *Tema. Journal of Land Use, Mobility and Environment* 7, 1 (2014), 83–100.

[57] Yongheng Rao, Jianjun Zhang, Qin Xu, and Shuqing Wang. 2018. Sustainability assessment of road networks: A new perspective based on service ability and landscape connectivity. *Sustainable cities and society* 40 (2018), 471–483.

[58] Ananya Roy. 2005. Urban informality: toward an epistemology of planning. *Journal of the american planning association* 71, 2 (2005), 147–158.

[59] Mehebub Sahana, Haoyuan Hong, and Haroon Sajjad. 2018. Analyzing urban spatial patterns and trend of urban growth using urban sprawl matrix: A study on Kolkata urban agglomeration, India. *Science of the Total Environment* 628 (2018), 1557–1566.

[60] Annemarie Schneider and Curtis E Woodcock. 2008. Compact, dispersed, fragmented, extensive? A comparison of urban growth in twenty-five global cities using remotely sensed data, pattern metrics and census information. *Urban Studies* 45, 3 (2008), 659–692.

[61] Karen C Seto, Roberto Sánchez-Rodríguez, and Michail Fragkias. 2010. The new geography of contemporary urbanization and the environment. *Annual review of environment and resources* 35 (2010), 167–194.

[62] Hossein Shafizadeh-Moghadam and Marco Helbich. 2015. Spatiotemporal variability of urban growth factors: A global and local perspective on the megacity of Mumbai. *International Journal of Applied Earth Observation and Geoinformation* 35 (2015), 187–198.

[63] Kalpana Srivastava. 2009. Urbanization and mental health. *Industrial psychiatry journal* 18, 2 (2009), 75.

[64] Dominic Stead and Stephen Marshall. 2001. The relationships between urban form and travel patterns. An international review and evaluation. *European Journal of Transport and Infrastructure Research* 1, 2 (2001).

[65] Shyamantha Subasinghe and Yuji Murayama. 2017. Urban Growth Evaluation: A New Approach Using Neighborhood Characteristics of Remotely Sensed Land Use Data. In *Spatial Data Handling in Big Data Era*. Springer, 181–196.

[66] Andrew J Tatem and Simon I Hay. 2004. Measuring urbanization pattern and extent for malaria research: a review of remote sensing approaches. *Journal of Urban Health* 81, 3 (2004), 363–376.

[67] Hannes Taubenböck, Thomas Esch, Andreas Felbier, Michael Wiesner, Achim Roth, and Stefan Dech. 2012. Monitoring urbanization in mega cities from space. *Remote sensing of Environment* 117 (2012), 162–176.

[68] Hannes Taubenböck, Martin Wegmann, Christian Berger, Markus Breunig, Achim Roth, and Harald Mehl. 2008. Spatiotemporal analysis of Indian megacities. *Proceedings of the international archives of the photogrammetry, remote sensing and spatial information sciences* 10, Part B (2008), 75–82.

[69] Hannes Taubenböck, Martin Wegmann, Achim Roth, Harald Mehl, and Stefan Dech. 2009. Urbanization in India–Spatiotemporal analysis using remote sensing data. *Computers, environment and urban systems* 33, 3 (2009), 179–188.

[70] CO Tong and SC Wong. 1997. The advantages of a high density, mixed land use, linear urban development. *Transportation* 24, 3 (1997), 295–307.

[71] Intelligent Transport. 2019. Report finds Mumbai's car density is five times that of capital Delhi. https://www.intelligenttransport.com/transport-news/77576/mumbai-car-density/

[72] Evelin Priscila Trindade, Marcus Phoebe Farias Hinnig, Eduardo Moreira da Costa, Jamile Sabatini Marques, Rogério Cid Bastos, and Tan Yigitcanlar. 2017. Sustainable development of smart cities: A systematic review of the literature. *Journal of Open Innovation: Technology, Market, and Complexity* 3, 3 (2017), 11.

[73] Will R Turner, Toshihiko Nakamura, and Marco Dinetti. 2004. Global urbanization and the separation of humans from nature. *Bioscience* 54, 6 (2004), 585–590.

[74] Ming Wang, Qingquan Li, Qingwu Hu, and Meng Zhou. 2013. Quality analysis of open street map data. *International archives of the photogrammetry, remote sensing and spatial information sciences* 2 (2013), W1.

[75] Robert C Weih and Norman D Riggan. 2010. Object-based classification vs. pixel-based classification: comparative importance of multi-resolution imagery. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38, 4 (2010), C7.

[76] Johnny Wood. 2018. The 10 fastest-growing cities in the world are all in India. https://www.weforum.org/agenda/2018/12/all-of-the-world-s-top-10-cities-with-the-fastest-growing-economies-will-be-in-india/

[77] Jianguo Wu, G Darrel Jenerette, Alexander Buyantuyev, and Charles L Redman. 2011. Quantifying spatiotemporal patterns of urbanization: The case of the two fastest growing metropolitan regions in the United States. *Ecological Complexity* 8, 1 (2011), 1–8.