We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Chapter

A Review of the Machine Learning in GIS for Megacities Application

Nasim Tohidi and Rustam B. Rustamov

Abstract

Machine learning (ML) is very useful for analyzing data in many domains, including the satellite images processing. In the remote sensing data processing, ML tools are mainly founded out a place for filtering, interpretation and prediction information. Filtering aims at removing noise and performing transformations, which is vital segment of data processing as useful performance of data validation. An interpretation is significant part of it as the stage of objects classification depends of existing task for solution. Prediction is performed to estimate precise values of underlying parameters or future events in the data. It can be used successfully above achievements in a variety of areas. An urbanization is one of the spheres of advance technology application where highly need to collect appropriate data for understanding of challenges facing society. The process of urbanization becomes very important problem, thanks to city expansion. Each city is a complicated system. It consists of various interactive sub-systems and is affected by multiple factors, including population growth, transportation and management policies. To understand the driving forces of the urban structure change, the satellite-based estimates are considered to monitor these changes, in long term. GIS (geographic information system) is equivalent to methods related to the use of geospatial information. Besides, the increasing application of ML techniques in various fields, including GIS, is undeniable. Thus, the chapter attempts to review the application of ML techniques in GIS with a focus on megacities and theirs features fixing/identification and solution.

Keywords: geographic information system, machine learning, urbanization, data processing, modeling

1. Introduction

Today there is a growing need for the collection, processing, management and efficiently use of reliable spatial information. Therefore, it is very significant to be aware of relevant approaches and to share experiences and develop best practices. This growing demand is due to the most important developments in society, which in turn are magnified by rapid urbanization and the conditions of the megacities.

Location, in the form of spatial data, is a key point for visualizing the current location, predicting events and enhancing service delivery. Information about location can integrate and strengthen the complex analysis of the distribution of locations, events, and services. This provides many opportunities for improving government services in terms of best governmental segments, interacting with

customers and optimizing processes. As cities get larger, spatial information becomes like a key tool in efficient urban service delivery, public safety, and overall resource management.

On the other hand, today, artificial intelligence methods, especially ML techniques, have come to the attention of scientists and officials in various fields, to analyze and manage the enormous data that is produced at any given moment, and one of the most exciting tools that have entered the material science toolbox in recent years is ML. Undoubtedly, one of these fields is GIS.

In practice, a GIS allows users to understand the spatial dimensions of their work and relate it to information such as population information as well. The data collected and stored by the GIS can be used for different purposes ranging from transport, draught analysis, agriculture, disease-outbreak analysis, land occupancy, etc. At the same time GIS makes possible to storage a big volume of data in safely stage and access to them at any needed time and rapid base. So, the goal of this chapter is to review past works and research in this area, because it can be supposed that can help greatly in understanding the current situation and capabilities; besides, it will be attempt of step in planning for future developments in the field of GIS.

The remainder of this chapter is organized and structures as follows. In Section 2, main definitions are mentioned. Section 3 presents an overview of ML application in GISs and related works in this area. In Section 4, it has been introduced the evaluation metrics and datasets. The last Section 5 provides conclusions.

2. Fundamental principles

In order to review the ML application in GIS, the first is needed to familiarize with the basic concepts in this regard. The followings are some fundamental principles and definitions.

2.1 Machine learning

ML is an application of artificial intelligence that provides systems the ability to automatically learn and improve their performance from experience without being explicitly programmed. ML focuses on the development of computer programs that can access data and use it in the process of learning [1].

The process of learning begins with observations or data, such as examples, direct experience, or instruction. The data will be used in order to look for patterns in it and make better decisions based on the provided examples. The primary goal is to allow the machines learn without human intervention or assistance and adjust actions accordingly.

ML algorithms are often categorized as supervised or unsupervised, however this categorization is very general and it cannot cover all of the available methods [2]:

- Supervised ML algorithms can use what has been learned in the past by using labeled examples to predict future events from unseen data. Starting from the analysis of a training dataset (labeled examples), the learning algorithm predicts the output values. The system is capable of providing targets for each new input after sufficient training. Besides, the algorithm can compare its output with the correct output and find errors to modify the model accordingly. Examples of these algorithms: Support Vector Machine (SVM), Decision Tree, Random Forest, KNN, Regression, etc.

- Unsupervised ML algorithms are used when the training data is not labeled or classified. The purpose of these algorithms is to examine how systems can derive a function to describe the hidden pattern of unlabeled data. They may not specify the appropriate output, but it explores data and can infer to describe hidden structures from unlabeled data. Examples of these Learning: Apriori algorithm, K-means, EM.

- Semi-supervised ML algorithms fall in between the two types of previously mentioned algorithms, because they use both labeled and unlabeled data for training. Usually, a small portion of data is labeled and a large amount of it, is unlabeled. The systems that use these algorithms can achieve high level of accuracy. Typically, semi-supervised learning is selected when the acquired labeled data requires skilled and relevant resources in order to learn from it (producing labeled data costs money and takes time.). Otherwise, accessing to unlabeled data generally does not require additional resources.

- Reinforcement learning algorithms are learning methods that interacts with its environment by generating actions and receiving punishments or rewards. Trial and error search and delayed reward are the most important features of these algorithms. They allow systems agents to automatically determine the ideal behavior in a particular context in order to maximize its performance quality. Simple reward feedback is known as the reinforcement signal. Examples of these Learning: Q-learning, Markov Decision Process.

ML enables analysis of massive amount of data. Besides, it generally provides faster, more accurate results in order to identify profitable opportunities or dangerous risks, it may also require additional time and resources to train it properly. ML requires formatted data that is analyzed to build a ML model. In other words, it requires an appropriate set of data that can be applied to a learning process.

ML can be used in cases where using human resources is not time/cost effective or when many variables are being considered simultaneously. ML uses the prepared data to train a ML algorithm. An algorithm is a computerized procedure or recipe. When the algorithm is trained on the data, a ML model will be generated. Once the data is prepared and the algorithm trained, the ML model can make predictions about the unseen data, on its own.

Selecting the right algorithm for the issue is necessary for applying ML successfully. Selection is largely influenced by the application and the data available.

2.1.1 Choosing the most appropriate ML algorithm

There are a large number of ML algorithms available. Choosing the optimal algorithm for a specific problem is dependent on its features such as speed, accuracy, training and predicting time, amount of data required to train, data type, how easy is it to implement, etc. Most of the time, for GIS applications, time is very important.

To avoid dependence on the specific conditions, it is common to analyze the runtime of algorithms in an asymptotic sense. So, considering n the number of training sample, p the number of features, n_t the number of trees, n_{sv}, the number of support vectors and k, the number of clusters, following are time complexity factors of some ML algorithms, which help to choose the correct algorithm for the issue (**Table 1**):

Algorithm	Learning	Predicting
Regression	$O(p^2n + p^3)$	$O(p)$
Decision Tree	$O(n^2p)$	$O(p)$
Random Forest	$O(n^2pn_t)$	$O(pn_t)$
Naïve Bayes	$O(np)$	$O(p)$
SVM	$O(n^2p + n^3)$	$O(pn_{sv})$
KNN	—	$O(np)$
K-means	$O(n^{pk+1})$	$O(k)$

Table 1.
Time complexity of some ML algorithms.

Where:

- Time for Learning is time associate with training of dataset. It varies with size of data and algorithm we are using in that.

- Time for Predicting is time associate with testing of dataset or predicting unseen data. It varies with size of data and algorithm we are using in that.

Most of the time, about 80 percent of the dataset will be used for training and the remaining part will be used for tuning and testing. In addition, it should be noted that, as the training phase most of the time can be performed offline, the predicting time is more important for developers.

Generally, it can be used the points above to shortlist a few algorithms, but it is hard to know right at the start which algorithm will work in the best way. It is usually desirable to work iteratively. Among the ML algorithms can be identified as potential good approaches, throw the data into them, run them all in either parallel or serial, and at the end evaluate the performance of the algorithms to select the best one(s).

2.2 Megacity

A megacity is defined by the United Nations (UN) as a city which has a population of 10 million or more people. Currently, there are 38 megacities in the World (**Figure 1**). The UN statistics indicate that the city with the largest populations worldwide is Tokyo with 38.8 million people. Recently, the UN has predicted that the number of megacities will rise to 41 by the year 2030.

The urbanization process poses enormous challenges for governments, social and environmental planners, engineers, architects and the residents of the megacities. No wonder, the growing population of cities creates demand, in areas such as housing and services. The environmental destruction and poverty are two other concerns, which city administrations have to take care of, as especially poor people do not have the necessary financial resources to tackle these problems.

Megacities affect a variety of living conditions for citizens. Although stress level, traffic jams, poor air quality and increasing health risks, make life more difficult in megacities, most people still choose to live there. Therefore, more accurate governmental programs are needed to help improve living conditions for the metropolitan inhabitants.

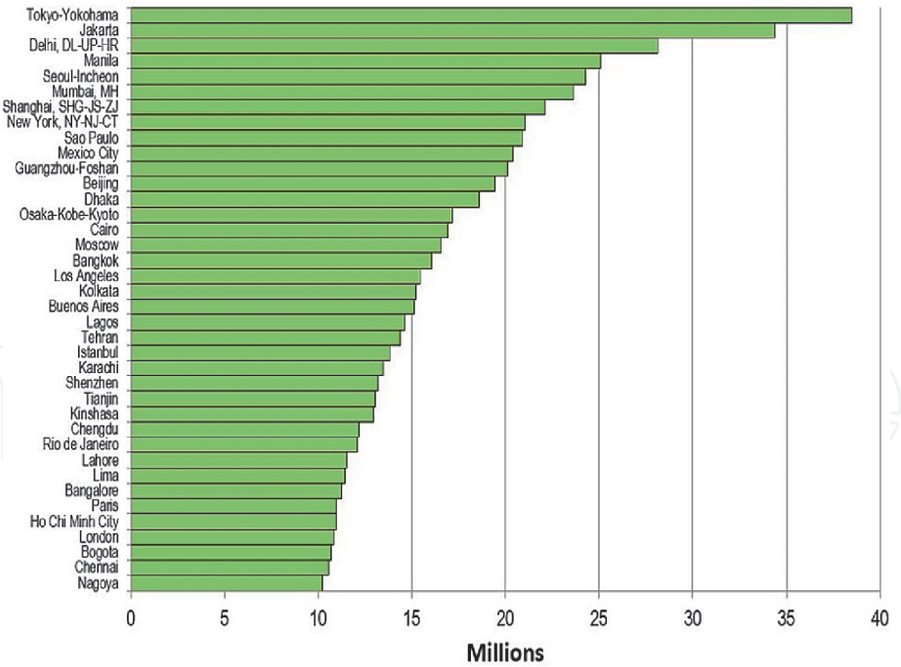


Figure 1.
The 38 megacities in 2019 [3].

As more cities are becoming megacities and existing megacities are growing, policymakers and urban planners are grappling with the questions of how to make growth at this scale sustainable, and how to tackle the escalating social, economic and environmental problems evident in the world's megacities. One of the most popular solutions is ML.

3. Application of machine learning in GIS

Urban dispersal and expansion has become an important issue for municipalities, environmental scientists and urban planners. Especially, in megacities, this issue becomes more vital. Currently, more than %50 of the world's population lives in urban areas and then it is predicted to grow over the %65 by 2050, according to the United Nation report. For example, all population in the 500,000+ urban areas of Australia and New Zealand combines to equal that of Moscow or Bangkok, and only slightly larger than Los Angeles (16.4 million). It is known that developing countries have already begun a rapid urbanization [4]. The fact that the global population has increased rapidly since the industrial revolution of the 18th century, highlights the problems of urban planning and urbanization, because of the population gathering in certain centers [5]. This unnatural pace of urbanization has created significant social and environmental challenges for decision-makers [6]. In addition, modeling and simulation are effective tools for discovering the urban development mechanisms and for providing planning in growth management. Therefore, monitoring and modeling the urban sprawl of cities is a necessary key parameter to prevent precautions [7, 8].

As it has been illustrated in **Figure 2**, Asia remains the dominance in terms of megacities, with nearly 58 percent of the population in larger metropolitan areas. This is approximately five times as many greater urban area residents as in North America or Africa. Besides, Asia has more than five times as many larger urban area residents as Europe and eight times that of South America [3].

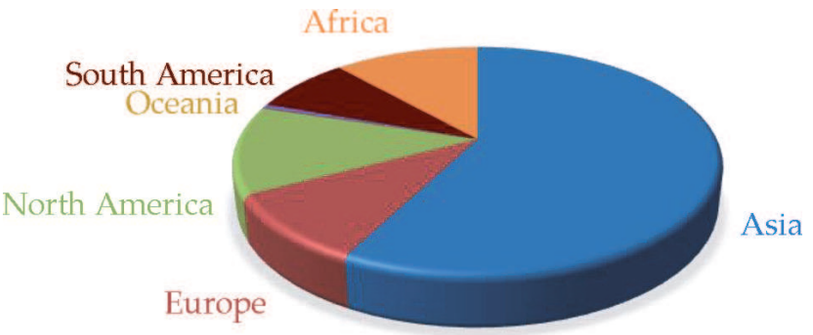


Figure 2.
Built-up urban area population in 2019.

Urban expansion modeling became widespread in the 1960s [9] and have accelerated with the developments of the technologies like Remote Sensing (RS) and GIS. Today, RS has been widely recognized as an essential tool for urban planning, management and design due to allows to get spatiotemporal data that are necessary for modeling environmental impacts, urban expansion and population growth. Particularly, the benefits of satellite-based image data have attracted attention of the scientific studies on urban expansion and environment [10, 11]. RS enables the collection of spatial details data for large areas at different time intervals; therefore, it provides a unique perspective on revealing the spatial and temporal dynamics of the change process in the land use and urban expansion [12]. GIS technology is described as an effective tool for identifying and monitoring the land cover change at different scales [13–16].

The dynamic modeling via GIS as a tool for urban simulation has rapidly gained popularity [7, 17]. The application of ML models has increased noticeably in RS filed due to the increased availability of powerful and flexible ML software and improvements in computing performance [18, 19].

The useful application of artificial neural networks (ANNs) in interpreting spatial resource information have been proven i.e. one of the most common are back-propagation neural networks, which are widely used by spatial planners. However, to improve the usability of ANNs for map-based applications, a more efficient method for communicating between the GIS and a trained ANN, is critical. When ANNs and GIS are used together for many applications to improve decision-making quality. The ANN design will consist of numerous layers, all of which can have different weights. The training process of an ANN involves changing the weightings over time until as it is desired the network reaches the static or optimum firing state. After training, an ANN can be used for applications effectively and consistently. Through the application of an ANN, GIS professionals can add another dimension to their spatial capabilities. In some research, a combination of neural networks with remote sensing image data for mapping the urbanization dynamics, has been proposed [20–22].

As another helpful technic, recently evolutionary algorithms have been used to tackle a variety of complex computation and optimization problems, such as natural language processing [23], route finding [24] and image processing [25]. One of the most important applications of evolutionary algorithms is in the field of GIS [26–28]. It should be noted that generally these algorithms are not fast in comparison with other ML algorithms and their main usage is for optimization.

All in all, there are two main ways for satellite-image processing that each of them has its own advantages and disadvantages. Sometimes there is high-resolution data (satellite images) so processing this amount of data would take a lot of time, with high accuracy. Some other cases, depending on the issue the resolution is not good, but the measurement methods are highly qualified.

3.1 Process description: impact and influence determination

Over the years, the fast development of map services [29] and volunteered geographic information (VGI) [30] has provided a huge number of geo-tagged images. This data source has given information on every corner of a city and has been enabling broader and more in-depth quantitative research in related fields. These data improve the understanding of the dynamic and physical features of the city by identifying landmark [31], detecting urban identities [32, 33], assessing the living environment inequality [34], and modeling human activities [35] and popular places [36]. Also, they provide information on the social and physical structures of dynamic metropolitan environments [36, 37].

The MIT Media Lab launched the "Place Pulse" program in 2013, which is a data collection platform that enables volunteers to take part in the urban perception rating experiment. By the end of 2016, the MIT Place Pulse dataset had collected 1,170,000 pairwise comparisons from 81,630 online participants for 110,988 cityscape images. Given this dataset and advances in ML techniques, many studies have tried to analyze human perceptions of urban appearance [34, 38–41].

Since previous approaches use low or mid-level image attributes, they have problem in extracting high-level information about the natural image. Some examples of these attributes are: Gist, SIFT- Fisher Vectors, DeCAF features [41], geometric classification map, color Histograms, HOG2x2, and Dense SIFT [40].

For instance, according to building models, SVM and Linear Regression were used in [41] to predict image labels. Support Vector Regression was used in [40], Ranking SVM it has been used in [42], and several convolutional neural network (CNN) based approaches were used in [39, 41, 42].

Among the various image representations and models, approaches that uses deep convolutional neural network (DCNN), have outperformed conventional methods to a large extent. In [39], authors introduce a DCNN model that is based on the Deep Residual Network (ResNet) [43], which won the first place in the ImageNet Large Scale Visual Recognition Competition [44].

Recently, a shift-invariant and hierarchical model has emerged in the form of DCNN, because of the availability of large-scale annotated datasets and the rapid development of high-performance computing systems. In some research, DCNN was employed to conduct human perception modeling and prediction. Due to its powerful capability to learn and represent automatic image feature, this model has attracted a lot of attention and achieved great success in multiple fields, including speech recognition [45], natural language processing [46], and visual object detection [43, 47].

A very deep convolutional neural network is hard to train and optimize because of disappearance gradients and the curse of dimensionality [43, 48]. ResNet is known as an acceptable attempt to address this problem. It was designed to learn the residual functions with regard to the layer inputs rather than learning the unreferenced functions [43].

In [49], a data-driven ML approach that measures how people perceive a place in a large-scale urban region, was proposed. In particular, a deep learning model, that had been trained with millions of human ratings of street-level imagery, was used to predict human perceptions of a street view image. The model achieved a high accuracy rate in predicting six human perceptual indicators. These indicators are: beautiful, boring, depressing, lively, safe and wealthy. It could help to map the distribution of the city-wide human perception for a new urban area. Besides, to determine the visual points that may cause a place to be perceived as various perceptions, a series of statistical analyses was performed. From the 150 object categories that had been segmented from the images of the street view, many

objects were identified as being positively or negatively correlated with each of the indicators. The mentioned results helped urban planners and researchers to take a step toward getting the interactions of the place sentiments and semantics.

Big data are voluminous and complex data of different qualities, that have the potential to generate new hypotheses and new methods for understanding interactions between social, biophysical and infrastructure domains of complex urban systems that face the challenges of climate change [50, 51].

The movement of people tracked through cell phones is an example of crowd-sourced and big data, which offers manifold new possibilities for assessing the city's inner workings, and the availability, quality and quantity of data is evolving, rapidly (**Figure 3**). Crowd-sourced information can be used as a reliable proxy, with much better resolution and replication, for more traditional methods of empirical social survey [53].

Similar analyses of social media provide the opportunity to complement the existing traditional ways of collecting information about human behavior in cities, which can be brought together with other sources of biophysical and infrastructural data, especially in spatial formats, collected through GIS [54, 55].

Big data can also come from urban hotlines, city planning offices, tax assessor databases, records about utility use and repair, and the rapid emergence of sensors and instrumented buildings, ecological spaces and even roads [56].

In **Figure 3**, The direction of change is shown by color, where I equals warmer and wetter; II colder and wetter; III colder and drier; and IV warmer and drier conditions. The direction of change is measured with the Euclidean distance in the 2D space including temperature and precipitation change. The classification of the magnitude of change corresponds to quartiles.

The usefulness of big data for understanding urban systems such as efficacy of climate solutions and climate impacts will only increase with time [56].

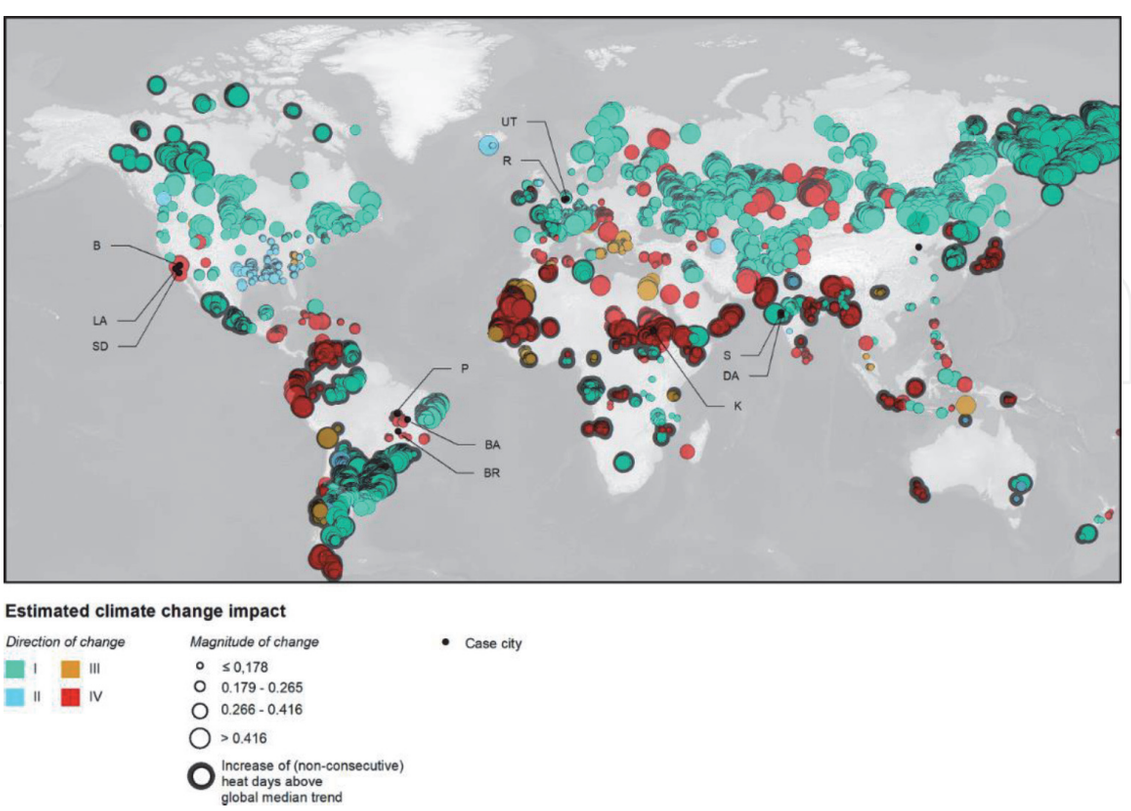


Figure 3.
Mapping direction of temperature change from 1901 until 2014 and rainfall from 1901 until 2013 in cities [52].

Decision-makers of cities need improved information, that regularly updated, about human behavior and perceptions and how they relate to climate change, both globally and locally. Considering human behavior in cities and linking it to down-scaled climate projections and remotely sensed observations of urban form, land cover, land-use patterns and social-demographic information from national and international databases, has the potential for improved decision-making to drive a much more nuanced and highly spatially resolved platform. Over the past decade, with the advancement of the digital social sciences and big data, as well as the increasing use of social media data (SMD) in geographical studies, new opportunities have emerged for augmenting urban systems and climate impacts research and expanding them [57].

Geocoded SMD, which comes from social media users (e.g., Twitter, Facebook, Instagram) opens up a significant new opportunity to fill data gaps and address many of the obstacles that prevent researchers and practitioners from understanding the human behavior component of urban system dynamics and climate change.

SMD and other big data let researchers to ask a wide range of spatially explicit questions at an unprecedented scale. Most of the time, social media allows users to manually select the location from where they post a message, or automatically adds it via geolocation tracking services. However, at present, geo-located tweets and Flickr photographs represent a small portion of the overall volume of SMD (e.g., only %1 of all tweets are tweets that geocoded via GPS constitute) [58] the sheer quantity of these data makes them worth investigating. Geotagged tweets can reinforce traditional control data (e.g. remotely sensed images, roads, parcels). For instance, for modeling population distribution, geo-located Twitter messages can serve as control data [59].

Research using geo-located SMD is also starting to take shape to study socioeconomic disparities and their relationship to climate impacts in cities. For example, crowd-sourced data from Foursquare users in London, had been used to be a reliable proxy for the localization of income variability and highlighting places which are more at risk across the city [60].

Yet, mapping based on data that are demographically unrepresentative, can also regenerate spatial segregation and give an unfair picture of the places which matter citywide [61]. The same is true for global-scale analyses. The amount of geocoded tweets widely varies among nations. The United States and Brazil are some of the countries whit the highest proportion of geocoded and non-geocoded tweets, while countries such as Norway and Denmark record considerably lower values [62].

The emergence of various types of big data provides interesting options for evaluating how people use and respond to urban events, policies, programs and designs to adapt and mitigate climate change. New types of data can be an important source for examining the use, value and social equity of specific spaces in the city, that provide refuge during climate driven extreme events, such as parks, vacant areas and open spaces that can provide, like cooling during heat waves. Working with big data can provide opportunities for multi-year to decadal datasets to understand the interactions between human and nature in the city and could be crucial to evaluate progress on examining influences of climate change and of mitigation options in cities [57].

Various sources of big data have already been helpful for awareness of disaster risk management and climate adaptation planning. In [63], authors for assessing the desired location and capacity of flood evacuation shelters, used volunteered geographic information through SMD as a source; while, in [64], researchers used SMD sourced observations of flooding to develop a method for estimating flood extent in Jakarta. In addition, following the devastating impact of Hurricane Sandy in New

York City, scientists used SMD in Twitter to reveal the geographies of a range of social processes and actions that happened shortly after the event [65].

In another research, Twitter data collected have been used during the devastating Sendai earthquake in Japan to assess social networks and build a database to study the human landscape of post-disaster effects [66]. Understanding interactions between climate change and fire prone landscapes is another major concern for adapting with climate change and for reducing the disaster risk. In [67], authors were able to use SMD to evaluate spatial patterns of situational awareness during the Horsethief Canyon Fire in Wyoming, besides they demonstrated the usefulness of SMD for actionable content during a crisis.

Another promising route which has been used in previous research, is the combination of satellite data with other datasets and analyzing it by ML. For example, in [68], authors proposed an accurate, scalable and cheap method for estimating consumption expenditure and asset wealth from high-resolution satellite imagery. In this research, using satellite data from five African countries: Nigeria, Tanzania, Uganda, Malawi, and Rwanda, they showed how a CNNN can be trained to detect image features that can explain up to 75% of the variation in local-level economic outcomes, which result in estimation of poverty levels via satellite data.

Another point is that, big data can become a central tool for online monitoring of urban risks and climate policies, made possible by sensor-based cities and the large amounts of data typically generated by their residents through social media. Applications include [69]:

- Use real-time data extracted from local weather stations, rainfall and sewer gauges to collect real-time data in hydrodynamic models for improved flood prediction;

- Combine local high frequency observations, with regional monitoring and forecasts, along with tracking of geospatial social messaging (e.g., posts about occurring events) to provide improved early warning about potential effects;

- Use image processed CCTV feeds to understand the risks, for instance, water surface locations and social media feeds to validate in real-time the emergent the flooding patterns;

- Integrate spatially heterogeneous sensor data from flows and movements with geospatial social messaging, CCTV and other data to reach a better understanding of the temporal dynamics of impacts;

- Combine CCTV monitoring with social media data feeds to improve understandings of citizen reaction and response to emerging impacts for optimized hazard mitigation and planning in future;

- Apply knowledge from previous events, like modeling result sets of both risks and impacts, to improve pre-response event from the site to the city-scale for future events.

ML techniques, especially neural networks, are powerful tools for multi-dimensional and complex big data analysis, where complexity needs to be reduced to understand its main drivers [70]. CNNs work well to classify images [71], and are widely used to evaluate land-use patterns [72]. Some researchers have gone even further and combined this approach with the socio-economic data analysis [68, 73].

As mentioned before, authors' work in [68] is very instructive. Using a combination of CNN, daytime satellite imagery and nightlight data, they predict poverty in five African countries at a village scale. For this purpose, they did their analysis in three steps. In the first step the CNN is trained on ImageNet [74] to learn how to recognize visual attributes like edges and corners. Second, it was well tuned so as to be able to predict intensities of the night-time in daytime images. Nightlights are a universally consistent poverty predictor. Therefore, the model was trained to focus on the aspects in daytime imagery, which are relevant to poverty estimation. In the third and final step, socio-economic survey data was used to train ridge-regression models on both household surveys and the image features from the previous two steps. Their approach used night-time data as a globally consistent, but very noisy proxy for poverty in an intermediate step and eventually explains %37–55 of average household consumption, and %55–75 of the variation in average household asset wealth. While it used publicly available data, it delivered better results than cellphone-based studies and outperforms products that rely solely on nightlights.

Recently another study used data extracted from Google Street View images and ML methods, such as v-support regression and feature extraction, to estimate high income areas in US cities [38]. From another perspective, phone records were used to reveal detailed mobility patterns for improving the understanding of travel behavior and traffic management [75].

3.2 High-precision measurement

An example of the activity recognition task is transportation mode detection can be in which data from smartphone sensors carried by users are employed to deduce what transportation mode the individuals have used. Microelectromechanical systems (MEMS), such as gyroscopes and accelerometers are embedded in most smartphone devices [76] from which the data can be obtained at high frequencies.

Nowadays, smartphones have powerful sensors like Global Positioning System (GPS), accelerometer, light sensors, gyroscope, etc. Having such sensors that embedded in a small device carried in all life activities has allowed researchers to investigate new research areas. Some of the benefits of these smart devices include ability to send and receive data through various ways (e.g. Wi-Fi/cellular network/Bluetooth), ubiquity and processing data [77]. The knowledge of individuals' transport mode can be adopted in many applications and also can facilitate several tasks, as follows:

1. Knowing transportation mode is a necessary part of urban planning for transportation, which is usually examined through questionnaires or telephone interviews or travel diaries [77, 78]. Most of the time, this traditional method of polling is inaccurate, expensive, limited to a specific area, and not up-to-date [79].

2. As environmental applications, by obtaining the transport mode, the carbon footprint and the amount of calories burnt of individuals can be determined. Besides, health situation and physical activities can be monitored, the risk exposure can be tracked, and the environmental influences of one's activities can be examined [80].

3. Other applications such as giving real-time information to users with the knowledge of speed and transport mode from the them as probes [77, 81], offering individuals with personalized messages and advertisements based on their transportation mode [77].

Several studies have used GPS data for classification purposes. However, using GPS sensors have some limitations, such as: in shielded areas like tunnels, GPS information is not available and the GPS signals may be lost especially in high dense locations, which results in erroneous position information. In addition, the GPS sensor consumes power a lot, so sometimes users turn it off to save the battery [79, 81]. Some research focus on developing detection models using ML techniques and data obtained from smartphone sensors like gyroscope, accelerometer and rotation vector, without GPS data [82]. In this way, it has the advantage of considering multiple sensors even without using GPS, the transportation modes can be identified.

4. Evaluation methods

Given the area of research, there are always some standard methods to evaluate a system that uses a ML algorithm. In addition, there should be some standard datasets the are prepared for the learning process (training, tuning and testing). Therefore, in the following the evaluation metrics and datasets in GIS for urbanization are introduced.

4.1 Metrics

There is no single connotation for the word "quality", because it is difficult to define quality with an absolute concept. Obviously, the data quality within software systems relates to the benefits that can be achieved by an organization. Furthermore, it is dependent on various aspects. Thus, to measure data quality accurately, one unique feature has to be chosen for considering the contribution of other attributes of the data quality as a whole. Following aspects can be used to describe the data quality (**Table 2**) [83].

It should be noted that by getting a high score in any of the mentioned dimensions, does not simply mean that a high quality data has been achieved. For example, the timelines may only matter in terms of correctness (correct user information is available, but if it is not updated, then it is useless). Sometimes, these features complement to each other [83].

Dimensions	Definition
Relevance	The importance of each piece of information stored in the database.
Reliability	The sources of data are reliable.
Correctness	The real world situation is represented by each set of stored data.
Timeliness	The data has been updated on time and with adequate frequency
Precision	The accuracy of the stored data is enough to characterize it.
Unambiguous	Each piece of data carries a unique meaning.
Accuracy	The level of data that can be accurately represented.
Objectivity	Data is objective: do not need people to judge, interpret, or evaluate.
Security	Access is secure and limited.
Completeness	The absence of the essential data: how much available data is missing.

Table 2.
Patterns of data quality dimensions.

The goals of data quality metrics are multi-dimensional. Indeed, they can set information quality objectives for data creators and managers to achieve, set standards for data to be produced, acquired and curated, and introduce measurement methods for quality judgment.

These metrics include rules that determine the thresholds of meeting appropriate professional expectations and govern the measurement of data quality aspects and levels. In order to configure and organize the rules, a basic structure is needed to distinguish the transformation process from data quality expectations to a set of applicable claims and to prevent unprofessional conduct [84].

Defining dimensions of data quality metrics can meet some purposes. Most of the time, the dimensions are classified according to accepted standard of scholarly activities within an academic discipline as well as other related disciplines that use the data. Scientists have developed several sets of data quality dimensions [85].

The dimension categories differ from each other, according to the academic field(s) in which data are regulated or by the different researchers' understanding and preference. Not only their dimensions are categorized differently among scholars, but also their definitions vary according mostly to different types of data. In practice, variations exist, such as integrity may be described in a different way to measurement adjusted strategies, and accuracy may be calculated at different levels of explanation [85].

Landslide susceptibility mapping (LSM) is a prime step in implementing the immediate disaster management planning and risk mitigation measures. All susceptibility models must be verified for their predictions accuracy. An unverified prediction model and susceptibility maps are nonetheless meaningless and hence do not have any scientific significance. The issue of LSM validation have tackled by many studies [86].

Several LSM approaches have been developed and described in numerous papers. These approaches are mainly divided into three groups: heuristic, deterministic and statistical techniques.

The heuristic techniques are based on the expert's knowledge to group landslide-prone areas into several ranks from high to low classes. It is often used for susceptibility mapping in large areas. While, deterministic techniques rely on numerical modeling of the physical mechanism that controls slope failure. However, they are not suitable for a large-scale mapping, due to their problematic and impractical need for a huge array of data, namely rock mechanical properties, the wetness and soil saturation and soil depth. Statistical and probabilistic techniques including bivariate, multivariate statistical methods, certainty factor, as well as knowledge-based techniques such as ANNs and fuzzy logic approaches are promising methods for predicting the landslides [87].

In most cases, the models are tested with an independent set of data, which was not used for training the model. In [88], authors reported a three following approach to obtain an independent sample of the landslide for validation purpose [87].

1. From the entire landslide inventory map of the study area, two sets of randomly divided landslide polygons should be created, one for the susceptibility analysis and one for validation the models;

2. In a part of the whole study area, the susceptibility analysis should be performed; the obtained result should be tested in another part, distinctly with different landslides;

3. The analysis should be performed using landslides happened in specific period and validation should be carried using landslides occurred in a different

period. This is the most sufficient to test the validity of the prediction mode, however, the toughest to apply as it requires knowledge of the temporal distribution of landslides over an adequately long-time spans.

As an example, image classification is not valid without evaluating its accuracy. The source of errors could stem from the classification itself, image recording, inappropriate training data and so on, however, in accuracy evaluation it is assumed that all differences between classification results and reference data comes from the classification errors.

Confusion matrix is one of the most common methods that evaluates classification accuracy. This matrix contains a category comparison of relationship between known, ground-truth data and classification results for the same category.

The overall accuracy of the classification process, is measured in percent and indicates the number of pixels which correctly classified divided by the total number of pixels. Kappa coefficient is a measure of overall statistical agreement. It measures the overall agreement of classification results, excluding agreement acquired, not on purpose, but by chance [89].

4.2 Data

From the very first satellite launched in 1972 till the Landsat 8, launched in 2013, Landsat satellite data have been recognized as a source of objective and reliable information. These missions provide high quality worldwide multispectral data and have been successfully used in countless applications in science [90].

The Landsat archive has provided multispectral data over the Earth for about 40 years. This fact makes Landsat data an attractive information source for studies related to change detection, especially for identifying land use and land cover changes indications.

World population was more than 7 billion at the time of the latest Landsat, Landsat 8. Considering the valuable information about changes to Earth's land surface for more than 40 years, the Landsat program has given decision makers a reliable source for managing Earth's resources for the planet's burgeoning population with integral information about the World's food, water, forests and how land resources are being used [90].

Imagery from these satellites is distributed for free and was obtained from the USGS website: http://earthexplorer.usgs.gov/.

Landsat 5 had Multi-Spectral Scanner (MSS) and Thematic Mapper (TM) sensors. TM sensor has 6 spectral bands with the resolution – 30 m and 1 thermal infrared band with resolution of 120 m (**Table 3**) [91].

Landsat 7 has Enhanced Thematic Mapper Plus (ETM+) sensor with 6 multispectral bands with 30 m resolution, 1 thermal band with the resolution of 60 m and 1 panchromatic band with 15 m resolution (**Table 4**). Bands 1–5&7 were used for LULC classification, while band 6 for LST extraction in both cases of Landsat 5 and 7 [91].

High-resolution maps of settlements and urban footprints form the basis for an integrated evaluation of global settlement patterns. In the past decade, there has been rapid progress in preparation of such maps. New satellite technology and improved data processing using ML have facilitated rapid improvement in their accuracy and resolution. The MODIS 500 urban land cover [92], until recently represented the state of the art in urban land cover datasets [93]. It is now outperformed by both the Global Urban Footprint (GUF) dataset which have higher resolution and accuracy than any other urban land cover dataset [94], even if comparing it to the high quality Global Human Settlement Layer (GHSL) [95, 96] or GlobeLand 30 [97]. The GUF attributes a binary urban footprint at a resolution as

Band	Spectral band	Resolution
1	0.45–0.52 μm	30 m × 30 m
2	0.52–0.60 μm	30 m × 30 m
3	0.63–0.69 μm	30 m × 30 m
4	0.76–0.90 μm	30 m × 30 m
5	1.55–1.75 μm	30 m × 30 m
6	10.4–12.5 μm	120 × 120 m
7	2.08–2.35 μm	30 m × 30 m

Table 3.
Landsat 5 TM bands.

Band	Spectral band	Resolution
1	0.45–0.515 μm	30 m × 30 m
2	0.525–0.605	30 m × 30 m
3	0.63–0.69 μm	30 m × 30 m
4	0.75–0.90 μm	30 m × 30 m
5	1.55–1.75 μm	30 m × 30 m
6	10.4–12.5 μm	60 m × 60 m
7	2.09–2.35 μm	30 m × 30 m

Table 4.
Landsat 7 ETM+ bands.

high as 0.4″ (approximately 12 m) at the equator and 0.6″ in the mid-latitudes on a global coverage. Also it is freely available for scientific use. This high resolution constitutes a paradigm shift in studying urban extent for cities around the world.

The importance of the satellite imageries for evaluating urbanization by measuring land use and land cover change for cities and their surroundings, is undeniable. Remote sensing (RS) is a reliable data source, which provides spatially consistent coverage of large areas with temporal frequency and high spatial detail. Besides, it is useful for analyzing phenomenon that is time dependent, such as urban expansion [98]. Therefore, RS is an accurate and effective data source for monitoring expansion of metropolitans, especially in cases that information related to the land use management is inconsistent and inappropriate.

This is a list of some other datasets that provide information related to GIS for urbanization:

1. GLOBAL Map[1]: It is a set of digital maps that cover the entire world to express the status of global environment, accurately. It is developed through the cooperation of National Geospatial Information Authorities (NGIAs) in the world. The Global Map provides eight main map themes at a nominal ground resolution of 1 km for raster data and at a scale of 1:1,000,000 for vector data. These themes are:

    - Transportation
    - Boundary

[1] https://nationalmap.gov/small_scale/atlas-ftp-global-map.html?openChapters=chptrans#

- Drainage
- Population Centers
- Elevation
- Vegetation
- Land Cover
- Land Use

2. Gridded Population of the World (GPW)[2]: It is the dataset of NASA's socioeconomic data and applications center, which includes raw population, and population density of the past, current and future prediction. The purpose of GPW is to provide a spatially disaggregated population layer that is compatible with datasets from social, economic, and Earth science disciplines, and RS. This data is globally consistent and spatially explicit for research, decision-making and communication.

3. World Bank Geodata[3]: In this data, a wide range of World Bank datasets converted to KML format, including GNP, schooling and financial data.

4. Global ADMINISTRATIVE Areas[4]: Administrative areas in this database are countries and lower level subdivisions such as provinces and departments. The latest version is 3.6 and it was released in 2018. It restricts 386,735 administrative areas, and scientists can download the spatial data by country.

5. Armed CONFLICT Location and Event Dataset[5]: This data includes all reported conflict events in 50 countries in developing world, from 1997 to present.

6. Global Rural-Urban Mapping Project (GRUMP)[6]: It is the dataset from NASA'S socioeconomic data and applications center, which includes information on rural and urban population balances.

7. Open Street Map (OSM)[7]: Crowdsourced data for the whole world, which contains many important things like points of interest, buildings, roads and road names, ferry routes, etc.

8. Geohive[8]: the initiative is made available by Ordnance Survey Ireland for easy access to public spatial data, and includes population and county statistics. it is not provided in GIS data formats, but it is easily convertible from CSV.

---

[2] https://sedac.ciesin.columbia.edu/data/collection/gpw-v4

[3] https://sourceforge.net/projects/googleworldbank/

[4] https://gadm.org/

[5] https://www.acleddata.com/

[6] https://sedac.ciesin.columbia.edu/data/collection/grump-v1

[7] https://learnosm.org/en/osm-data/osm-in-qgis/

[8] https://geohive.ie/

4.2.1 Urban landscapes

The World Bank, in south and east Asia, has explored the patterns, consequences and policy implications related to spatial development of cities by outlining the increasing availability of spatial data and developments in analytics. Data from Earth observation (EO) satellite can give valuable results which are useful for measuring urban growth over a wide range of spatial and temporal scales, especially when combined with data from other sources. The resulting digital urban maps give an accurate, up-to-date and cost effective resource to assist governments in understanding the nature of urban development and making informed decisions. EO datasets allow for harmonized and standardized measurements. Also, they enable planners to make spatially and temporally consistent comparisons and global assessment. In addition, they are particularly significant for monitoring and understanding the evolution of cities. For instance, allowing authorities to know when built-up areas spill across formal administrative boundaries. This shows the need to cooperate with adjoining administrative areas on issues like collecting garbage or connective infrastructure [99].

The World Bank has created a database to analyze the speed, magnitude and spatial form of urbanization in EO data. These data help researchers examine the drivers and influences of the urbanization nature and how the urban landscape has evolved into its current state. It offers a basis for understanding the effects of policy change and identifying priorities for new initiatives. In particular, the focus is on exploring the institutional frameworks for urban management, like mechanisms to coordinate service delivery across administrative jurisdictions, investment for example in transport and other network infrastructure and regulation such as zoning and pricing of services.

About twelve years ago, the World Bank launched the "Earth Observation for Development" initiative. So, data in areas where data are commonly scarce and unreliable, are provided. Such information is useful for building project fundamentals against which progress can be gauged, high priority issues identified and mitigation measures determined. Focus of this project is on areas like metropolitan development and related fields including disaster risk management, the environment, water and energy. The bank has also developed the Urban Management and Analysis (PUMA) platform to facilitate more collaboration between policymakers and other development stakeholders, toward these purposes. By using this tool, users with no GIS experience would be able to access, analyze and share urban spatial data in an interactive and customizable way [100].

These activities have resulted in more than 30 technical helping projects that done for urban planners and partners, in the period 2008–2018. As a result, highly specialized big data mapping products and monitoring systems that leverage EO data for South Asian cities have been launched.

4.2.2 Megacities

In the South Asia Megacities Improvement Program, EO big data was used to analyze 20 years of urban expansion in the metropolitan areas of Delhi, Mumbai and Dhaka. These data make it possible to measure the qualitative and quantitative aspects of transformation, like the distribution and density of urban sprawl, the growth rate of built-up areas and urban land use change. This information helps analysts to trace how informal settlements grow outside the cities' boundaries, and to understand the drivers of land use [99]. Therefore, some important insights into land cover and use in the three cities revealed (**Figure 4**). Furthermore, it showed

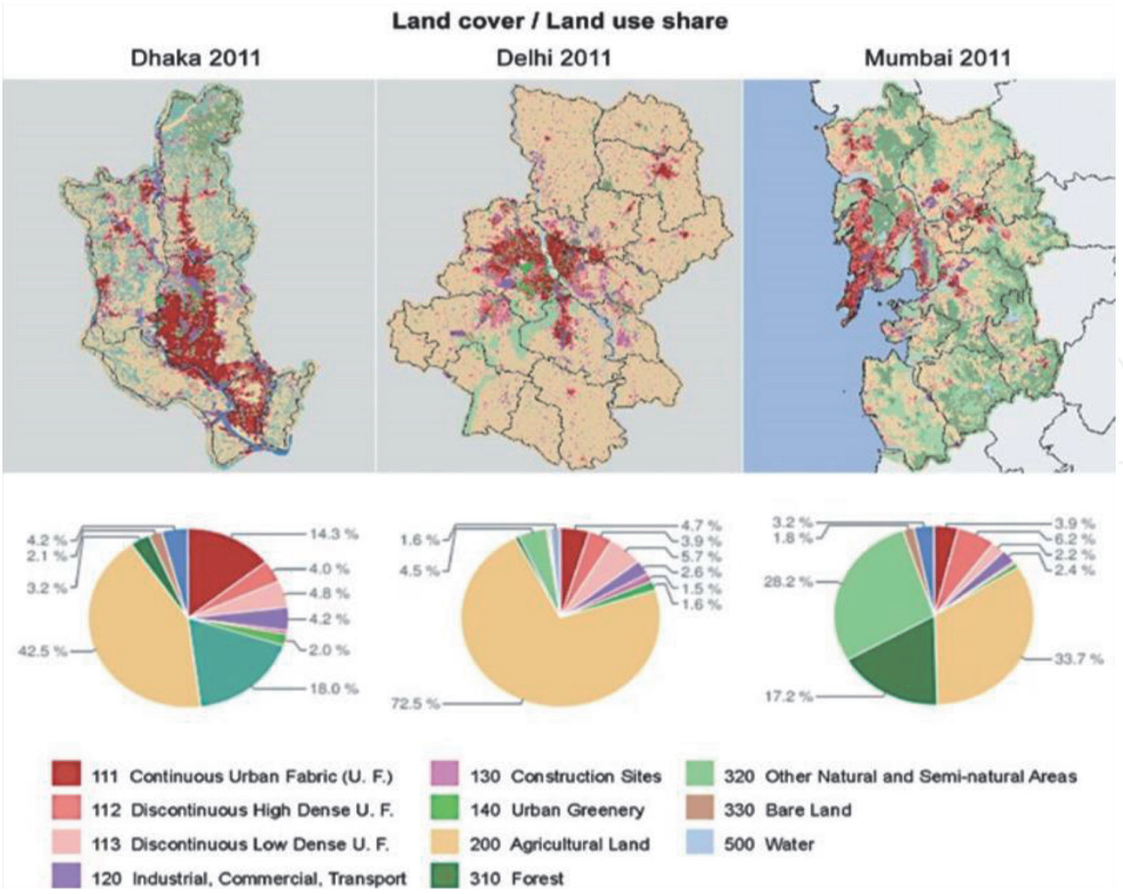


Figure 4.
Sample visualizations from the South Asia geospatial analysis [101].

the percentage of land taken by settlements and industrial build-up, agriculture, natural or semi-natural vegetation and forest [101].

By using it, urban planners and development stakeholders could understand existing demands and plan for future needs. For instance, in Delhi, the maps illustrate that with industrial development, the urban expansion is accelerated. This mostly happened between 2003 to 2010; however, a considerable increase in construction sites shows that it will continue in the future, so it must be planned [101].

By using digitized spatial data, analysts would be able to study the target at different administrative levels: metropolitan, city, district or sub-district, and also other non-administrative units. These datasets make it possible to aggregate flexibly. One example is showing the proportion of sprawl by district, its density, the drivers of urban change and class evolution within urban areas. Together with environmental or socio-economic data, the data can prepare information on the proportion of population to urban growth, and can measure indicators like compactness, the ratio of green space to citizens, and the accessibility of these areas.

The results of applying EO big data can be crucial for coordination between public, private and household investment in infrastructure, productive capital and housing, respectively. Thus, policymakers would be able to promote optimal spatial and transportation links between businesses, affordable housing and commercial units, health and education services and recreational areas. In addition, these views can be applied to support rural-to-urban migrants and ensuring that rapid urbanization is inclusive. Since EO big data methods spread across the world's megacities, and are refined and adapted, they will provide valuable tools to policymakers, and greater benefits for the citizens of the future.

4.2.3 Residential cities

EO big data approaches are also contribute to drive sustainable urban development. The mentioned research on the use of high-resolution satellite data for poverty mapping, draws emerging techniques that can show fast changing urban areas in near real-time. These methods can determine built-up area, density of cars and buildings, and types of roofing and road. Via ML techniques and image processing algorithms, also they can calculate whether buildings are more rectangular or have more chaotic angles, that indicates higher poverty level, and construct poverty indicators like the ratio of paved roads in an area. So, stakeholders can target their interventions exactly where they are mostly useful [101].

All in all, analysis of EO big data can be an important tool for managing city development in low-income countries. It can measure and track the urban expansion and highlight the drivers of economic growth. This result in better understanding the factors contributing to inefficiencies and inequality in urban areas, and providing optimized policies. Besides, they can create flexibility in urban environments, so that residents, businesses and systems can adapt to persistent stresses or shocks. Also they can provide residential cities that meet their residents' needs.

5. Conclusions

This book chapter briefly introduced ML and past research about the application of ML algorithms for processing of daily satellite imagery. It has been demonstrated several aspects of detecting and classification of Earth features merging into local geographical and geodetical system with further GIS development. The main purpose of the chapter is to provide existing resources for researchers to be aware of the up-to-date status of development of ML application in GIS in particular in studies of megacities.

The real potential of ML in GIS is not sufficiently developed yet. On one hand, both fields intersect in analytical discussions. At the same time, most GIS applications which are desirable for ML implementation, are driven by conventional approach and standard tools of commercial GIS packages.

Merging GIS and ML offers a potential mechanism to reduce the cost of analysis of spatial information by decreasing the amount of time spent on data interpretation. This integration allows the interpretive outcome from a small area to be transferred to a larger, geographically similar area, without the extra time and expense of putting geographers in the field for a time sufficient to cover geographical area.

ML can be considered both as a science and as engineering, depending on the goal. This technology is often seen as part of computing; however, it has links with various other areas including philosophy, psychology and linguistics. Its techniques can provide benefits within GIS over traditional methods, like statistical analysis, especially if data show some form of non-linearity. Thanks to such an opportunity of ML/GIS technology makes most successfully to apply for monitoring and observation consequences of megacity development.

Most people are unaware that they use artificial intelligence in their daily life. Finding solutions to decision-making issues by using models that allow decision makers to express their limitations and imprecise concepts that are used with large volume of geographic data, costs a lot. This chapter is expected to open opportunity to understand clearly fundamental aspects of ML/GIS development with basically related to the megacity studies.

IntechOpen

IntechOpen

Author details

Nasim Tohidi and Rustam B. Rustamov*
Alzahra University, Tehran, Iran

*Address all correspondence to: r_rustamov@hotmail.com

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

References

[1] Ray, S. A Comparative Analysis and Testing of Supervised Machine Learning Algorithms. 2018.

[2] Ayodele, T. Types of Machine Learning Algorithms. In: New Advances in Machine Learning. InTech; 2010. DOI: 10.5772/9385

[3] 15th Annual Demographia World Urban Areas. Demographia. Parliament of Hungary, Budapest. 2019.

[4] Fenta, A.A. Yasuda, H. Haregeweyn, N. Sewale Belay, A. Hadush, Z. Gebremedhin, M.A. Mekonnen, G. The dynamics of urban expansion and land use/land cover changes using remote sensing and spatial metrics: the case of Mekelle City of northern Ethiopia. International Journal of Remote Sensing. 2017:38:4107–4129. DOI: 10.1080/01431161.2017.1317936

[5] Akintunde, J.A. Adzandeh, E.A. Fabiyi, O.O. Spatio-temporal pattern of urban growth in Jos Metropolis, Nigeria. Remote Sensing Applications: Society and Environment. 2016:4:44–54. DOI: 10.1016/j.rsase.2016.04.003

[6] Farooq, M. Muslim, M. Dynamics and forecasting of population growth and urban expansion in Srinagar City – A Geospatial Approach. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Hyderabad; 2014. DOI: 10.5194/isprsarchives-XL-8-709-2014

[7] Dereli, M.A. Uğur, M.A. Polat, N. Yalçın, M. Spatio-Temporal Analysis of Urban Expansion Using Remote Sensing Data. In: 17th International Multidisciplinary Scientific GeoConference SGEM2017; Albena, Bulgaria; 2017. DOI: 10.13140/RG.2.2.35243.34084/1

[8] Kivilcim, C.Ö. Duran, Z. Impact of rapid urbanization on the morphology of historical Istanbul: Uskudar district case study. Fresenius Environmental Bulletin. 2015:24:14–18.

[9] Wilson, A.G. Urban and regional models in geography and planning. John Wiley and Sons. 1974.

[10] Karakus, C.B. Kavak, K.S. Cerit, O. Determination of variations in land cover and land use by remote sensing and geographic information systems around the city of Sivas (Turkey). Fresenius Environmental Bulletin. 2014: 23:667–667.

[11] Son, N.T. Chen, C.F. Chen, C.R. Thanh, B.X. Vuong, T.H. Assessment of urbanization and urban heat islands in Ho Chi Minh City, Vietnam using Landsat data. Sustainable Cities and Society. 2017:30:150–161. DOI: 10.1016/j.scs.2017.01.009

[12] Herold, M. Goldstein, N.C. Clarke, K.C. The spatiotemporal form of urban growth: measurement, analysis and modeling. Remote Sensing of Environment. 2003:86:286–302. DOI: 10.1016/S0034-4257(03)00075-0

[13] LStefanov, W. Ramsey, M.S. Christensen, P.R. Monitoring urban land cover change: An expert system approach to land cover classification of semiarid to arid urban centers. Remote Sensing of Environment. 2001:77:173–185. DOI: 10.1016/S0034-4257(01)00204-8

[14] Hoffhine Wilson, E. Hurd, J.D. Civco, D.L. Prisloe, M.P. Arnold, C. Development of a geospatial model to quantify, describe and map urban growth. Remote Sensing of Environment. 2003:86:275–285. DOI: 10.1016/S0034-4257(03)00074-9

[15] Weng, Q. Land use change analysis in the Zhujiang Delta of China using satellite remote sensing, GIS and

stochastic modelling. Journal of Environmental Management. 2002:64: 273–284. DOI: 10.1006/jema.2001.0509

[16] Chrysoulakis, N. Mitraka, Z. Stathopoulou, M.I. Cartalis, C. A comparative analysis of the urban web of the greater athens agglomeration for the last 20-years period on the basis of landsat imagery. Fresenius Environmental Bulletin. 2013:22:2139–2143.

[17] Sariyilmaz, F.B. Musaoglu, N. Tanik, A. Investigation of land use/cover changes of Sazlidere basin by using band ratio for builtup area (BRBA). Fresenius Environmental Bulletin. 2017:26:39–45.

[18] Skapura, D.M. Building Neural Networks, New York: Addison-Wesley Professional; 1996.

[19] Tayyebi, A. Delavar, M.R. Saeedi, S. Amini, J. Alinia, H.S. Monitoring land use change by multi-temporal Landsat remote sensing imagery. In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Beijing. 2008. DOI: 10.13140/2.1.2736.3204

[20] Tayyebi, A. Pijanowski, B.C. Tayyebi, A.H. An urban growth boundary model using neural networks, GIS and radial parameterization: An application to Tehran, Iran. Landscape and Urban Planning. 2011:100:35–44. DOI: 10.1016/j.landurbplan.2010.10.007

[21] Li, J. Huang, X. Gong, J. Deep neural network for remote-sensing image interpretation: status. National Science Review, pp. 1–4, 2019. DOI: 10.1093/nsr/nwz058

[22] Liu, X. Han, F. Ghazali, K.H. Mohamed, I.I. Zhao, Y. A review of Convolutional Neural Networks in Remote Sensing Image. In: Proceedings of the 2019 8th International Conference on Software and Computer Applications. 2019. DOI: 10.1145/3316615.3316712

[23] Tohidi, N. Hasheminejad, S.M.H. MOQAS: Multi-objective question answering system. Journal of Intelligent and Fuzzy Systems. 2019:36:3495–3512. DOI: 10.3233/JIFS-181364

[24] Davoodi, M. Mesgari, S.M. GIS-based Route Finding using Ant Colony Optimization and Urban Traffic Data from Different Sources. In: International Conference on Sensors & Models in Remote Sensing & Photogrammetry; Kish Island, Iran; 2015.

[25] Zhang, Y. Zhang, G. Zhao, H. Cao, Y. Liu, Q. Shen, Z. Li, A. A Convenient Tool for District Heating Route Optimization Based on Parallel Ant Colony System Algorithm and 3D WebGIS. International Journal of Geo-Information. 2019:8:1–18. DOI: 10.3390/ijgi8050225

[26] Eldrandaly, K.A. Hassan, M.M. AbdelAziz, N.M. A Modified Artificial Bee Colony Algorithm for Solving Least-Cost Path Problem in Raster GIS. International Journal of Applied Mathematics. 2015:9:147–154.

[27] Davoodi, M. Malekpour Golsefidi, M. Mesgari, S.M. A hybrid optimization method for vehicle routing problem using artificial bee colony and genetic algorithm. In: Joint Conferences of SMPR and GI Research; Karaj, Iran; 2019.

[28] García-Alonso, C.R. Pérez-Naranjo, L.M. Fernández-Caballero, J.C. Multiobjective evolutionary algorithms to identify highly autocorrelated areas: the case of spatial distribution in financially compromised farms. Annals of Operations Research. 2014:219:187–202. DOI: 10.1007/s10479-011-0841-3

[29] Anguelov, D. Dulong, C. Filip, D. Frueh, C. Lafon, S. Lyon, R. Ogale, A. Vincent, L. Weaver, J. Google Street View: Capturing the World at Street Level. Computer. 2010:43:32–38. DOI: 10.1109/MC.2010.170

[30] Goodchild, M. Citizens as Sensors: The World of Volunteered Geography. GeoJournal. 2007:69:211–221. DOI: 10.1007/s10708-007-9111-y

[31] Hays, J. Efros, A.A. Large-Scale Image Geolocalization. Multimodal Location Estimation of Videos and Images. 2015:41–62. DOI: 10.1007/978-3-319-09861-6_3

[32] Liu, L. Zhou, B. Zhao, J. Ryan, B.D. C-IMAGE: city cognitive mapping through geo-tagged photos. GeoJournal. 2016:81:817–861. DOI: 10.1007/s10708-016-9739-6

[33] Zhang, F. Zhang, D. Liua, Y. Lin, H. Representing place locales using scene elements. Computers. Environment and Urban Systems. 2018:71:153–164. DOI: 10.1016/j.compenvurbsys.2018.05.005

[34] Salesses, P. Schechtner, K. Hidalgo, C. The Collaborative Image of The City: Mapping the Inequality of Urban Perception. PLoS ONE. 2013:8. DOI: 10.1371/journal.pone.0068400

[35] Arase, Y. Xie, X. Hara, T. Nishio, S. Mining people's trips from large scale geo-tagged photos. In: Proceedings of the 18th ACM international conference on Multimedia; Firenze, Italy; 2010.

[36] Crandall, D. Backstrom, L. Huttenlocher, D.P. Kleinberg, J.M. Mapping the world's photos. In: Proceedings of the 18th International Conference on World Wide Web; Madrid, Spain; 2009.

[37] Less, E.L. McKee, P. Toomey, T. Nelson, T. Erickson, D. Xiong, S. Jones-Webb, R. Matching study areas using Google Street View: A new application for an emerging technology. Evaluation and Program Planning. 2015:53:72–79. DOI: 10.1016/j.evalprogplan.2015.08.002

[38] Glaeser, E.L. Kominers, S.D. Luca, M. Naik, N. Big Data and Big Cities: The Promises and Limitations of Improved Measures of Urban Life. Economic Inquiry. 2018:56:114–137.

[39] Dubey, A. Naik, N. Parikh, D. Raskar, R. Hidalgo, C.A. Deep Learning the City: Quantifying Urban Perception At A Global Scale. in European Conference on Computer Vision (ECCV), 2016. DOI: 10.1007/978-3-319-46448-0_12

[40] Naik, N. Philipoom, J. Raskar, R. Hidalgo, C. Streetscore – Predicting the Perceived Safety of One Million Streetscapes. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2014.

[41] Ordonez, V. Berg, T.L. Learning High-Level Judgments of Urban Perception. In: European Conference on Computer Vision, 2014. DOI: 10.1007/978-3-319-10599-4_32

[42] Porzi, L. Rota Bulò, S. Lepri, B. Ricci, E. Predicting and Understanding Urban Perception with Convolutional Neural Networks. In: Proceedings of the 23rd ACM international conference on Multimedia, 2015. DOI: 10.1145/2733373.2806273

[43] He, K. Zhang, X. Ren, S. Sun, J. Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas; 2016. DOI: 10.1109/CVPR.2016.90

[44] Russakovsky, O. Deng, J. Su, H. Krause, J. Satheesh, S. Ma, S. Huang, Z. Karpathy, A. Khosla, A. Bernstein, M. Berg, A.C. Fei-Fei, L. ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision. 2015:115:211–252. DOI: 10.1007/s11263-015-0816-y

[45] Zhang, S. Zhang, S. Huang, T. Gao, W. Speech Emotion Recognition Using Deep Convolutional Neural Network and Discriminant Temporal Pyramid

Matching. IEEE Transactions on Multimedia. 2018:20:1576–1590. DOI: 10.1109/TMM.2017.2766843

[46] Conneau, A. Schwenk, H. Barrault, L. Lecun, Y. Very Deep Convolutional Networks for Natural Language Processing. ArXiv. vol. abs/1606.01781, 2017.

[47] Ren, S. He, K. Girshick, R. Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017: 39:1137–1149. DOI: 10.1109/TPAMI.2016.2577031

[48] Glorot, X. Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS); Sardinia, Italy; 2010.

[49] Zhang, F. Zhou, B. Liu, L. Liu, Y. Fung, H.H. Lin, H. Ratti, C. Measuring human perceptions of a large-scale urban region using machine learning. Landscape and Urban Planning. 2018: 180:148–160. DOI: 10.1016/j.landurbplan.2018.08.020

[50] Herold, M. Scepan, J. Clarke, K. The Use of Remote Sensing and Landscape Metrics to Describe Structures and Changes in Urban Land Uses. Environment and Planning. 2002:34: 1443–1458. DOI: 10.1068/a3496

[51] McPhearson, T. Pickett, S.T.A. Grimm, N.B. Niemelä, J. Alberti, M. Elmqvist, T. Weber, C. Haase, D. Breuste, J. Qureshi, S. Advancing Urban Ecology toward a Science of Cities. BioScience. 2016:66:1–15. DOI: 10.1093/biosci/biw002

[52] Scheuer, S. Haase, D. Volk, M. Integrative assessment of climate change for fast-growing urban areas: Measurement and recommendations for future research. PLoS ONE. 2017:12. DOI: 10.1371/journal.pone.0189451

[53] Wood, S.A. Guerry, A. Silver, J.M. Lacayo-Emery, M. Using social media to quantify nature-based tourism and recreation. Scientific Reports. 2013:3. DOI: 10.1038/srep02976

[54] Mora, H.M. Pérez-del H.R. Paredes-Pérez, J.F.; Sirvent, R.M. Analysis of Social Networking Service Data for Smart Urban Planning. Sustainability. 2018:10. DOI: 10.3390/su10124732

[55] Martí, P. Serrano-Estrada, L. Nolasco-Cirugeda, A. Social Media data: Challenges, opportunities and limitations in urban studies. Computers, Environment and Urban Systems. 2019: 74:161–174.

[56] Knox, E.G. The Slippery Slope of Material Support Prosecutions: Social Media Support to Terrorists. The Hastings law journal. 2014:66:295–329.

[57] Ilieva, R.T. McPhearson, T. Social-media data for urban sustainability. Nature Sustainability. 2018:1:553–565. DOI: 10.1038/s41893-018-0153-6

[58] Crampton, J.W. Graham, M. Poorthuis, A. Shelton, T. Stephens, M. Wilson, Matthew W. Zook, M.A. Beyond the Geotag: Situating 'Big Data' and Leveraging the Potential of the Geoweb. Cartography and Geographic Information Science. 2013:40:130–139. DOI: 10.1080/15230406.2013.777137

[59] Lin, J. Cromley, R. Evaluating geo-located Twitter data as a control layer for areal interpolation of population. Applied Geography. 2015:58:41–47. DOI: 10.1016/j.apgeog.2015.01.006

[60] Quercia, D. Saez-Trumper, D. Mining Urban Deprivation from Foursquare: Implicit Crowdsourcing of City Land Use. IEEE Pervasive Computing. 2014:13:30–36. DOI: 10.1109/MPRV.2014.31

[61] Cranshaw, J. Schwartz, R. Hong, J. Sadeh, N.M. The livehoods project: Utilizing social media to understand the dynamics of a city. In: Proceedings of the Sixth International Conference on Weblogs and Social Media; Dublin, Ireland; 2012.

[62] Graham, M. Hale, S.A. Gaffney, D. F. Where in the World Are You? Geolocation and Language Identification in Twitter. The Professional Geographer. 2013:66:568–578. DOI: 10.1080/00330124.2014.907699

[63] Kusumo, A.N.L. Reckien, D. Verplanke, J.J. Utilising volunteered geographic information to assess resident's flood evacuation shelters. Case study: Jakarta. Applied Geography. 2017:88:174–185. DOI: 10.1016/j.apgeog.2017.07.002

[64] Brouwer, R. Hadzhiyska, D. Ioakeimidis, C. Ouderdorp, H. The social costs for marine litter along the European coasts. Ocean & Coastal Management. 2017:138:38–49. DOI: 10.1016/j.ocecoaman.2017.01.011

[65] Shelton, A.O. Thorson, J.T. Ward, E. J. Feist, B.E. Spatial semiparametric models improve estimates of species abundance and distribution. Canadian Journal of Fisheries and Aquatic Sciences. 2014:71:1–12. DOI: 10.1139/cjfas-2013-0508

[66] Stefanidis, A. Crooks, A. Radzikowski, J. Harvesting ambient geospatial information from social media feeds. GeoJournal. 2013:78:319–338. DOI: 10.1007/s10708-011-9438-2

[67] Kent, J.D. Capello Jr, H.T. Spatial patterns and demographic indicators of effective social media content during the Horsethief Canyon fire of 2012. Cartography and Geographic Information Science. 2013:40:78–89. DOI: 10.1080/15230406.2013.776727

[68] Jean, N. Burke, M. Xie, M. Davis, W.M. Lobell, D.B. Ermon, S. Combining satellite imagery and machine learning to predict poverty. Science. 2016:353:790–794. DOI: 10.1126/science.aaf7894

[69] Creutzig, F. Lohrey, S. Bai, X. Baklanov, A. Richard, D. Shobhakar, D. Lamb, W.F. McPhearson, T. Minx, J. Munoz, E. Walsh, B. Upscaling urban data science for global climate solutions. Global Sustainability. 2019:2:1–25. DOI: 10.1017/sus.2018.16

[70] Hinton, G.E. Salakhutdinov, R.R. Reducing the Dimensionality of Data with Neural Networks. Science. 2006:313:504–507. DOI: 10.1126/science.1127647

[71] Pansombut, T. Wikaisuksakul, S. Khongkraphan, K. Phon-on, A. Convolutional Neural Networks for Recognition of Lymphoblast Cell Images. Computational Intelligence and Neuroscience. 2019:2019:1–12. DOI: 10.1155/2019/7519603

[72] Castelluccio, M. Poggi, G. Sansone, C. Verdoliva, L. Land Use Classification in Remote Sensing Images by Convolutional Neural Networks. arXiv. 2015. DOI: arXiv:1508.00092

[73] Tapiador, F.J. Avelar, S. Corrêa, C.T. Zah, R. Deriving fine-scale socioeconomic information of urban areas using very high-resolution satellite imagery. International Journal of Remote Sensing. 2011:32:6437–6456. DOI: 10.1080/01431161.2010.512928

[74] Deng, J. Dong, W. Socher, R. Li, L.J. Li, K. Li, F.F. ImageNet: a Large-Scale Hierarchical Image Database. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009); Miami, Florida, USA; 2009. DOI: 10.1109/CVPR.2009.5206848

[75] Toole, J.L. Colak, S. Sturt, B. Alexander, L.P. Evsukoff, A. González,

M.C. The path most traveled: Travel demand estimation using big data resources. Transportation Research Part C: Emerging Technologies. 2015:58:162–177. DOI: 10.1016/j.trc.2015.04.022

[76] Susi, M. Renaudin, V. Lachapelle, G. Motion Mode Recognition and Step Detection Algorithms for Mobile Phone Users. Sensors. 2013:13:1539–1562. DOI: 10.3390/s130201539

[77] Stenneth, L. Wolfson, O. Yu, P.S. Xu, B. Transportation Mode Detection using Mobile Phones and GIS Information. In: 19th ACM SIGSPATIAL International Symposium on Advances in Geographic Information Systems; Chicago, IL, USA; 2011.

[78] Xiao, Y. et al. Transportation activity analysis using smartphones. In: IEEE Consumer Communications and Networking Conference (CCNC); Las Vegas; 2012.

[79] Widhalm, P. Nitsche P. Brändie, N. Transport mode detection with realistic Smartphone sensor data. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012); Tsukuba; 2012.

[80] Reddy, S. Mun, M. Burke, J. Estrin, D. Hansen, M. Srivastava, M. Using Mobile Phones to Determine Transportation Modes. ACM Transactions on Sensor Networks. 2010: 6. DOI: 10.1145/1689239.1689243

[81] Manzoni, V. Maniloff, D. Kloeckl, K. Ratti, C. Transportation mode identification and real-time CO2 emission estimation using smartphones. Massachusetts Institute of Technology; Cambridge; 2010.

[82] Yazdizadeh, A. Patterson, Z. Farooq, B. An automated approach from GPS traces to complete tripinformation. International Journal of TransportationScience and Technology. 2019:8:82–100. DOI: 10.1016/j.ijtst.2018.08.003

[83] Gaba, S. Gaba, D. Tomar, A. Metric based Approach for Measuring Data Quality. International Journal for Innovative Research in Science & Technology. 2016:2:46–51.

[84] Xia, J. Metrics to Measure Open Geospatial Data Quality. Issues in Science and Technology Librarianship. 2012:68. DOI: 10.5062/F4B85627

[85] Loshin, D. Monitoring Data Quality Performance Using Data Quality Metrics. Informatica Corporation; Redwood City, USA; 2006.

[86] Nohani, E. Moharrami, M. Sharafi, S. Khosravi, K. Pradhan, B. Pham, B.T. Lee, S. Melesse, A.M. Landslide Susceptibility Mapping Using Different GIS-Based Bivariate Models. Water. 2019:11. DOI: 10.3390/w11071402

[87] Dou, J. Yunus, A.P. Bui, D.T. Sahana, M. Chen, C. Zhu, Z. Wang, W. Pham, B.T. Evaluating GIS-Based Multiple Statistical Models and Data Mining for Earthquake and Rainfall-Induced Landslide Susceptibility Using the LiDAR DEM. Remote Sensing. 2019: 11. DOI: 10.3390/rs11060638

[88] Irigaray, C. Del Castillo, T.F. El Hamdouni, R. Chacón, J. Evaluation and Validation of Landslide-Susceptibility Maps Obtained by a GIS Matrix Method: Examples from the Betic Cordillera (Southern Spain). Natural Hazards. 2007:41:61–79. DOI: 10.1007/s11069-006-9027-8

[89] Bobrinskaya. M. Remote Sensing for Analysis of Relationships between Land Cover and Land Surface Temperature in Ten Megacities [Thesis]. School of Architecture and the Built Environment Royal Institute of Technology (KTH); Stockholm, Sweden; 2012.

[90] Wulder, M.A. Loveland, T.R. Roy, D.P. Crawford, C.J. Masek, J.G. Woodcock, C.E. Allen, R.G. Anderson,

M.C. Belward, A.S. Cohen, W.B. Dwyer, J. Erb, A. Gao, F. et al. Current status of Landsat program, science, and applications. Remote Sensing of Environment. 2019:225:127–147. DOI: 10.1016/j.rse.2019.02.015

[91] Bobrinskaya, M. Remote Sensing for Analysis of Relationships between Land Cover and Land Surface Temperature in Ten Megacities. Geodesy and Geographic Information Technology; 2012.

[92] Schneider, A. Friedl, M.A. Potere, D.T. Mapping global urban areas using MODIS 500-m data: New methods and datasets based on 'urban ecoregions'. Remote Sensing of Environment. 2010: 114:1733–1746. DOI: 10.1016/j.rse.2010.03.003

[93] Potere, D.T. Schneider, A. Angel, S. Civco, D. Mapping urban areas on a global scale: Which of the eight maps now available is more accurate? International Journal of Remote Sensing. 2009:30:6531–6558. DOI: 10.1080/01431160903121134

[94] Esch, T. Heldens, W. Hirner, A. Keil, M. Marconcini, M. Roth, A. Zeidler, J. Dech, S. Strano, E. Breaking new ground in mapping human settlements from space -The Global Urban Footprint. ISPRS Journal of Photogrammetry and Remote Sensing. 2017:134. DOI: 10.1016/j.isprsjprs.2017.10.012

[95] Pesaresi, M. Ehrlich, D. Freire, S. The Global Human Settlement Layer (GHSL) – New Tools and Geodatasets for Improving Disaster Risk Assessment and Crisis Management. In: 11th International ISCRAM Conference; Pennsylvania, USA; 2014.

[96] Corbane, C. Politis, P. Pesaresi, M. Kemper, T. Siragusa, A. Estimation of Land Use Efficiency from the Global Human Settlement Layer (GHSL). In: QGIS and Applications in Territorial Planning. vol. 3; John Wiley & Sons; 2018. DOI: 10.1002/9781119457121.ch2

[97] Chen, J. Chen, J. GlobeLand30: Operational global land cover mapping and big-data analysis. Science China Earth Science. 2018:61:1533–1534. DOI: 10.1007/s11430-018-9255-3

[98] Jensen, J.R. Cowen, D.C. Remote Sensing of Urban/Suburban Infrastructure and Socio-Economic Attributes. Photogrammetric Engineering & Remote Sensing. 1999:65: 611–622.

[99] Esch, T. Taubenböck, H. Heldens, W. Thiel, M. Wurm, M. Geiss, C. Dech, S. Urban remote sensing - How can earth observation support the sustainable development of urban environments? In: Proceedings, Real CORP; Vienna; 2010.

[100] García, L.E. Rodríguez, D.J. Wijnen, M. Pakulski, I. The World Bank and Remote Sensing. In: Earth Observation for Water Resources Management: Current Use and Future Opportunities for the Water Sector; Washington; World Bank Group; 2016. p. 39–48

[101] Using Geospatial Data to Track Changes in Urbanization [Internet]. Available from: https://olc.worldbank.org/system/files/Using%20Geospatial%20Data%20to%20Track%20Changes%20in%20Urbanization.pdf World Bank Group.