ISPRS Journal of Photogrammetry and Remote Sensing 178 (2021) 203–218

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Mapping essential urban land use categories with open big data: Results for five metropolitan areas in the United States of America

Bin Chen [a,*], Ying Tu [b], Yimeng Song [c,d], David M. Theobald [e,f], Tao Zhang [b], Zhehao Ren [b], Xuecao Li [g], Jun Yang [b,h,i], Jie Wang [j], Xi Wang [k], Peng Gong [l], Yuqi Bai [b,h,i,*], Bing Xu [b,h,i,*]

[a] Division of Landscape Architecture, Faculty of Architecture, The University of Hong Kong, Hong Kong Special Administrative Region
[b] Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China
[c] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region
[d] Smart Cities Research Institute, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region
[e] Conservation Planning Technologies, Fort Collins, CO 80521, USA
[f] Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80523, USA
[g] College of Land Science and Technology, China Agricultural University, Beijing 100083, China
[h] Tsinghua Urban Institute, Tsinghua University, Beijing 100084, China
[i] Center for Healthy Cities, Institute for China Sustainable Urbanization, Tsinghua University, Beijing 100084, China
[j] State Key Laboratory of Remote Sensing Science, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China
[k] AI for Earth Laboratory, Cross-Strait Institute, Tsinghua University, Beijing 100084, China
[l] Departments of Geography and Earth Sciences, The University of Hong Kong, Hong Kong Special Administrative Region

ARTICLE INFO

Keywords:
Land use classification
Block-level mapping
Geospatial big data
Ensemble learning
NAIP
Sentinel-1/2

ABSTRACT

Urban land-use maps outlining the distribution, pattern, and composition of various land use types are critically important for urban planning, environmental management, disaster control, health protection, and biodiversity conservation. Recent advances in remote sensing and social sensing data and methods have shown great potentials in mapping urban land use categories, but they are still constrained by mixed land uses, limited predictors, non-localized models, and often relatively low accuracies. To inform these issues, we proposed a robust and cost-effective framework for mapping urban land use categories using openly available multi-source geospatial "big data". With street blocks generated from OpenStreetMap (OSM) data as the minimum classification unit, we integrated an expansive set of multi-scale spatially explicit information on land surface, vertical height, socio-economic attributes, social media, demography, and topography. We further proposed to apply the automatic ensemble learning that leverages a bunch of machine learning algorithms in deriving optimal urban land use classification maps. Results of block-level urban land use classification in five metropolitan areas of the United States found the overall accuracies of major-class (Level-I) and minor-class (Level-II) classification could be high as 91% and 86%, respectively. A multi-model comparison revealed that for urban land use classification with high-dimensional features, the multi-layer stacking ensemble models achieved better performance than base models such as random forest, extremely randomized trees, LightGBM, CatBoost, and neural networks. We found without very-high-resolution National Agriculture Imagery Program imagery, the classification results derived from Sentinel-1, Sentinel-2, and other open big data based features could achieve plausible overall accuracies of Level-I and Level-II classification at 88% and 81%, respectively. We also found that model transferability depended highly on the heterogeneity in characteristics of different regions. The methods and findings in this study systematically elucidate the role of data sources, classification methods, and feature transferability in block-level land use classifications, which have important implications for mapping multi-scale essential urban land use categories.

* Corresponding authors at: Division of Landscape Architecture, Faculty of Architecture, The University of Hong Kong, Hong Kong Special Administrative Region (B. Chen); Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing 100084, China (Y. Bai, B. Xu).
E-mail addresses: binley.chen@hku.hk (B. Chen), yuqibai@tsinghua.edu.cn (Y. Bai), bingxu@tsinghua.edu.cn (B. Xu).

https://doi.org/10.1016/j.isprsjprs.2021.06.010
Received 2 April 2021; Received in revised form 22 May 2021; Accepted 14 June 2021
Available online 25 June 2021

1. Introduction

Urban land use information that reflects socio-economic functions and human activities is fundamentally important for urban planning and management (Gong et al., 2020a; Zhang et al., 2018). As the highest level human modification on land surface, urban land uses have widespread effects on modulating local hydrology, climate, biodiversity, agriculture, living environment, and public health (Chen et al., 2017b; Clinton et al., 2018; Grimm et al., 2008; Seto and Shepherd, 2009; Watts et al., 2015; Wheater and Evans, 2009). With increasing urbanization and population growth, around 66% of global population will live in urban areas by 2050 (UNDESA, 2014). The accelerated conversion of land from rural to urban uses is expected to continue and lead to rapidly changing urban land uses from regional to global scales (Gao and O'Neill, 2020). However, detailed urban land use classification that outlines the distribution and composition of different land use types remains limited, in particular towards large-scale practices, due to extreme difficulties in (*i*) differentiating complex urban built-up areas to derive high-level semantic labels (Zhang et al., 2018; Zhang et al., 2017); (*ii*) coordinating financial input and skills of mapping personnel (Gong et al., 2020a); and (iii) securing the support of spatially and temporally explicit datasets with high to very high resolutions. All these challenges indicate the importance of developing a robust and cost-effective data-model framework and approach to derive accurate and timely urban land use classification maps.

A number of previous efforts have been made in the field of urban land use classification (Gong et al., 2020a; Hu et al., 2016; Liu et al., 2017; Liu and Long, 2016; Theobald, 2014; Tu et al., 2020; Yao et al., 2017; Zhang et al., 2018, 2019; Zhong et al., 2020), which can be categorized into three classes by the minimum size of their spatial representation as pixels, objects, or blocks. The pixel-based approaches mainly use spectral and textural signatures to classify land cover/use types. For example, there are several available gridded urban-extent maps at global scales such as Global Human Settlement Layer (Gong and Howarth, 1990, 1992; Pesaresi et al., 2013), Human Built-up and Settlement Extent (Wang et al., 2017), Global Man-made Impervious Surface (Brown de Colstoun et al., 2017), Global Artificial Impervious Area (Gong et al., 2020b), and others (Liu et al., 2020; Schneider et al., 2009; Zhou et al., 2018). However, the majority of these datasets with a moderate spatial resolution above 30 m generally regard urban as the single class of impervious area, rather than distinguishing sub-category land use types, or they are just experimental research with a small tract of land in a particular city. Lu and Weng (2006) have applied spectral mixture analysis to classify residential areas with different intensities and the integrated class of commercial/industrial/transportation uses in the city of Indianapolis, Indiana, United States (US). Theobald (2014) combined information of census housing, employment, and infrastructure, and land cover from satellite imagery to present a map of land uses for the conterminous US at a resolution of 30 m. These efforts greatly advance our understanding of land-use patterns from a macro perspective, but the pixel-based classification practices have not fully utilized the spatial information to uncover the composition of multiple land use types in urban settings.

With the rapid development of very-high-resolution (VHR, $\leq$1 m) satellite observations, it is now possible to identify the geometry, texture, size, location, and adjacent information of ground objects at a finer scale (Zhang et al., 2018; Zhong et al., 2020). Object-based approaches thus are popular for classifying urban land use types, based on segmented objects from VHR remote sensing imagery. Two types of information are always included for consolidating the classification models, i.e., intra-object features (e.g., spectral, texture) and inter-object features (e.g., connectivity, contiguity, adjacent alignment) (Zhang et al., 2018). In addition to machine learning algorithms such as random forest (RF) and support vector machines (SVMs) that bring together low-level features for urban land use classification through the train-and-predict protocol (Petropoulos et al., 2012; Tu et al., 2020), recent advances in deep learning convolutional neural networks (CNNs) make it promising to transform these features into classes at a higher, slightly abstract level (LeCun et al., 2015; Schmidhuber, 2015), thus facilitating image classification (Huang et al., 2018; Liu et al., 2019; Zhang et al., 2018, 2019) and object recognition (Cheng et al., 2016; Guo et al., 2018; Long et al., 2017) at different spatial scales. For example, Du et al. (2021) proposed an object-based urban land use classification method by combining a multi-scale semantic segmentation network and a conditional random field framework using VHR remote sensing images. However, these approaches are mostly applied to case studies and have not yet been scaled-up to large challenging datasets from regional to global scales because of the data availability and computational costs. Moreover, the segmented units from object-based classifications are largely influenced by the spatial scale effect (Myint et al., 2011), which cannot be easily applied in practical urban land-use planning and management, on account of the "application gap" (Zhong et al., 2020).

Given the fact that a street block representing a relatively homogeneous urban function (Erol and Akdeniz, 2005; Liu and Long, 2016) is more compatible with the base unit for urban planning and management, block-based approaches have been increasingly developed and applied in classifying urban land use (Gong et al., 2020a; Hu et al., 2016; Liu and Long, 2016; Zhong et al., 2020). Based on the statistics of Point of Interests (POIs) allocated within street blocks, land uses for 297 cities of China have been estimated (Liu and Long, 2016). Without considering other features into the differentiation of different land use classes, the accuracy of this derived map is highly determined by the quality and quantity of POIs. Several studies further explored to include more predictors from medium-resolution satellite imagery, POI data, and other auxiliary geospatial information, and experimental tests of block-based land use classification practices at the city level yielded more robust classification maps (Hu et al., 2016; Li et al., 2021; Liu and Long, 2016; Su et al., 2020; Yao et al., 2017). For example, Huang et al. (2021) integrated high-resolution multispectral and multi-view Ziyuan3-01 satellite images and Jilin1-07 nighttime light images to derive urban land use function for two megacities of Wuhan and Beijing in China. A data-driven point, line, and polygon semantic object mapping framework is recently proposed to integrate POIs, OpenStreetMap (OSM) data, and VHR Google Earth imagery for block-based urban land-use mapping in four cities of China (Zhong et al., 2020). However, the Google Earth imagery at regional scales is always spatially mosaicked using temporally irregular VHR satellite observations, which prevents spatially and temporally consistent information for categorizing land use types. Additionally, this work is also at the experimental stage focusing on specific testing zones without having a full picture of urban land use patterns in China. The implementation of this method will be challenging to scale up to large-scale regions due to the limitation of computational costs and model transferability. Gong et al. (2020a) report a new map of essential urban land use categories for entire China (EULUC-China) that uses 10-m satellite images, OSM, nighttime lights, POIs, and Tencent location-based service data in 2018 as input features, marking the beginning of a new approach of collaborative urban land use mapping over large areas. However, this product has several shortcomings: due to the incomplete coverage of OSM roads in China, the segmented land blocks with certain big sizes are typically mixed with different land uses, thus resulting in a relatively low overall accuracy; the predictors are extracted from low-level features such as spectral and derived remote sensing indices, without considering higher-level information that is highly correlated with urban land uses such as vertical height, texture, and gradients; a unified model is used to classify nationwide urban land use categories, which may lead to a biased performance for localized experiments.

The purpose of this study is to explore detailed urban land use classification in five metropolitan areas of the United States with numerous blocks (>194,000 blocks), using an expansive set of multi-source geospatial data layers, including multi-scale spatially explicit information on land surface, vertical height, socio-economic attributes,

social media, demography, and topography. The ultimate goal is to present a robust and cost-effective framework for mapping urban land use categories. Specifically, we seek to answer the following scientific questions: (1) How is the performance of block-based urban land use classifications in metropolitan areas of the United States? (2) How to leverage multiple machine learning models to achieve relatively robust and accurate performance? (3) What is the relative importance of different datasets and predictor features? (4) Regarding feature transferability, are locally trained models suitable for predicting land use classification in non-local areas? An improved understanding of these issues is needed to guide and move forward the campaign of block-based urban land use classification from local to regional and continental scales.

2. Materials and methods

2.1. Study area

We selected five representative metropolitan areas spanning over the continental US as our study area (Fig. 1), specifically, the San Francisco Bay area, Denver, New Orleans, Chicago, and New York City, which included different geographic distributions of urban land uses and landscape settings for testing our proposed framework in mapping urban land use categories.

2.2. Data

We included an expansive set of geospatial data layers (Table 1) in the mapping of essential urban land use categories: OSM, Global Urban Boundary (GUB), National agriculture imagery program (NAIP) imagery, Sentinel-2 multi-spectral imagery, Sentinel-1 ground range detected (GRD) data, Suomi NPP VIIRS Day-Night Band imagery, WorldPop population dataset, Twitter check-in records, and topography data. The details regarding each category of the dataset are provided as follows.

Table 1
Overview of categories, datasets, spatial resolution, and years data used in the mapping of essential urban land use categories.

Category	Dataset	Resolution (m)	Year (s)
Road network	OpenStreetMap (OSM)	±20	2018
Urban boundary	Global Urban Boundary (GUB)	30	2018
VHR multi-spectral	National Agriculture Imagery Program (NAIP)	0.6–1	2017–2018
HR multi-spectral	Sentinel-2	10–20	2018
SAR	Sentinel-1	10	2018
Human activities	Twitter	–	2014–2017
	Visible Infrared Imaging Radiometer Suite (VIIRS) nighttime light	500	2018
	WorldPop population	100	2018
Topography	National Elevation Data (NED)	10	2012

Initiated in 2004 as a volunteer effort, OSM (https://www.openstreetmap.org/) is now a substantial global spatial database that maps a variety of point, line, and polygon features. The positional accuracy of mapped features (±20 m) is mainly determined by the positioning technologies (e.g., GPS) employed and references used while digitizing these features. It has been reported that road features from OSM largely surpassed the accuracy of other publicly available global datasets such as Global Roads Open Access Data Set (±500 m) (Haklay, 2010), and the high precision and wide coverage make OSM the best available seamless

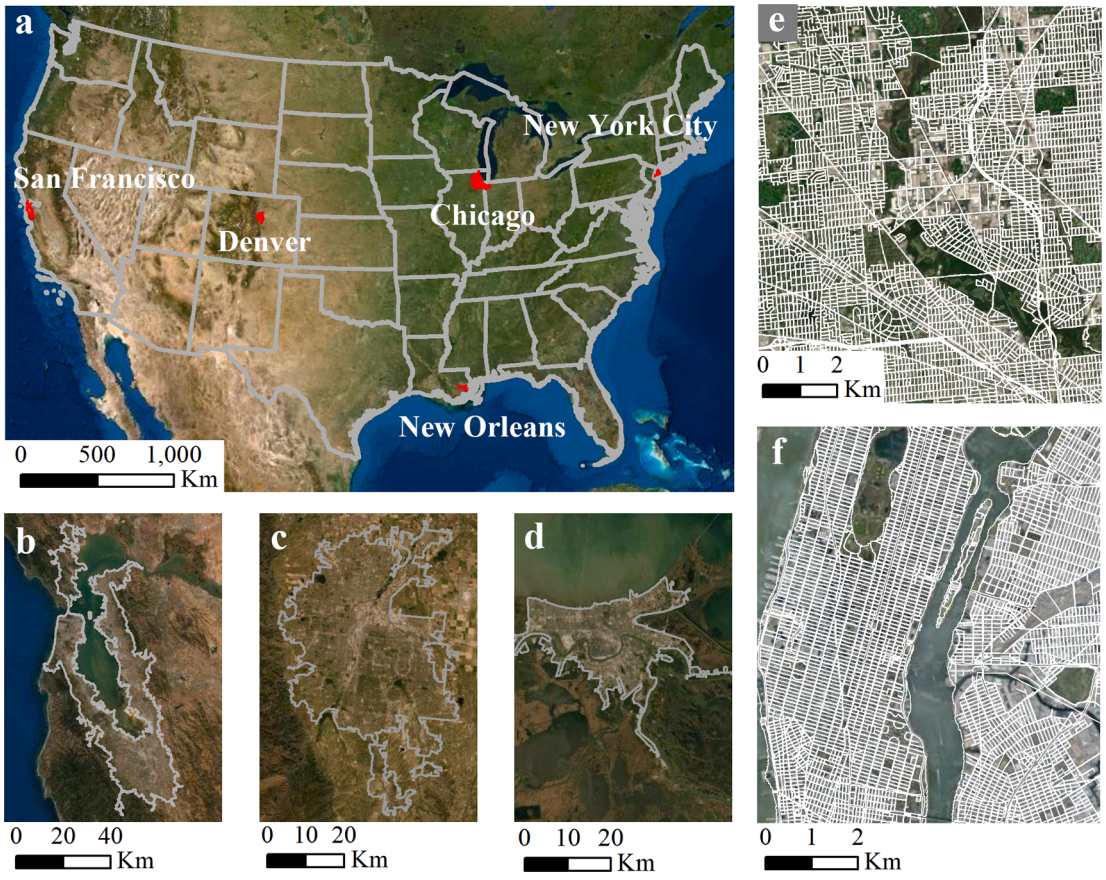

Fig. 1. Geographic distribution of five metropolitan areas in the United States (a). Enlarged metropolitan boundaries in white lines of (b) San Francisco Bay area, (c) Denver, and (d) New Orleans. Examples of street blocks generated in this study are shown in (e) Chicago and (f) New York City.

dataset (Barrington-Leigh and Millard-Ball, 2017; Meijer et al., 2018). Similar to practices in mapping global human modification (Kennedy et al., 2019), we grouped road features as either "highway", "motorway", "trunk", "primary", or "secondary" into a single layer of major roads, and grouped road features coded as "tertiary", "unclassified", and "residential" into another layer of minor roads.

We collected the GUB dataset in 2018, which is derived from 30-m Landsat imagery to represent the urban extent (Li et al., 2020a). Different from the commonly used administrative boundaries, the urban boundaries we used mark a physical region that consists of not only the built-up areas but also the associated natural lands in the urban center such as greenspace and water bodies. Two main steps were used in the process of generating GUB, firstly, the 30-m impervious surface pixels were aggregated to a coarse resolution (1 km) to derive a kernel density map; and secondly, urban boundaries in the urban fringe area were extracted using thresholding methods and refined using morphological operations.

The US Department of Agriculture's NAIP acquires aerial imagery during the agricultural growing seasons in the continental US. NAIP imagery is acquired at a one-meter ground sample distance (GSD) with a horizontal accuracy that matches within six meters of photo-identifiable ground control points, which are used during image inspection (USDA-FarmServiceAgency, 2020). Given the data availability, here we collected NAIP imagery in 2018 for the San Francisco Bay area, in 2017 for Denver, Chicago, New Orleans, and New York City.

The Sentinel-2 mission initiated by the European Commission and the European Space Agency constellation aims to provide systematic global acquisition of high-resolution multi-spectral imagery with a high revisit frequency (Drusch et al., 2012). With the full operation of two identical satellites, Sentinel-2A/B has now been able to provide an unprecedented observation of global land surface with a spatial resolution of 10–60 m and a high revisit of 5 days. Here we collected the Sentinel-2 Level-2A imagery in 2018 for the five metropolitan areas.

The Sentinel-1 mission provides Ground Range Detected (GRD) data from a dual-polarization C-band Synthetic Aperture Radar (SAR) instrument (Torres et al., 2012). Given the availability of multiple combinations of instrument mode and polarization in the Sentinel-1 data, we chose a homogeneous GRD subset by selecting GRD scenes with a dual polarization (i.e., VV and VH) at the spatial resolution of 10 m from the instrument mode of an interferometric wide swath. Specifically, we acquired the full coverage maximum composite of Sentinel-1 GRD data covering the five metropolitan areas in 2018.

The Suomi National Polar-orbiting Partnership Visible Infrared Imaging Radiometer Suite (VIIRS) supports a Day-Night Band (DNB) sensor that provides global daily measurements of nocturnal visible and near-infrared (NIR) light that are suitable for Earth system science and applications (Elvidge et al., 2017). The VIIRS DNB's ultra-sensitivity in lowlight conditions enables us to generate a new set of science-quality nighttime products that manifest substantial improvements in sensor resolution and calibration when compared to the previous era of Defense Meteorological Satellite Program/Operational Linescan System's (DMSP/OLS) nighttime lights image products (Elvidge et al., 2017; Shi et al., 2014). We collected monthly average radiance composites from the VIIRS stray light corrected DNB datasets at the spatial resolution of ~500 m in 2018 (Mills et al., 2013).

WorldPop (www.worldpop.org) provides different types of gridded population count datasets, depending on the methods and end-user application. Specifically, it provides the estimated number of people residing in each grid cell. Given the superiority of its fine spatial resolution and yearly updated frequency over other population datasets such as the Gridded Population of the World (GPW) (CIESIN, 2018) and LandScan (Dobson et al., 2000), we used the WorldPop 100-m resolution population count dataset in 2018, derived from the random forest model (Stevens et al., 2015; Tatem, 2017).

Twitter is one of the most popular social media platforms in the world, which allows users to post messages and record their real-time locations. As for 2016, the platform had more than 270 million active users, and 80% of them were cellphone-based (Lansley and Longley, 2016). The spatial-temporal features of geotagged Tweets are considered a good representative of dynamic population distributions (Frias-Martinez et al., 2012). In this study, we collected geotagged Tweets generated from April 2014 to December 2017 for the five selected metropolitan areas via the Twitter Streaming Application Programming Interface (API) (https://developer.twitter.com). Although the collected social media dataset, i.e., Tweets, had a ~2-year difference from other datasets used in 2018, it did not affect much on revealing the geographic distribution and diurnal changes of human activities over different land use blocks (Chen et al., 2020). To avoid overestimation, we further deleted all the duplicated Tweets with the same messages (e.g., advertising information produced by robots).

We collected the National Elevation Dataset (NED) with a spatial resolution of 1/3 arc-second (Evans, 2010). The NED is a seamless dataset with the best available raster elevation data of the conterminous United States.

All the datasets used in this study are open source. Except for the Twitter data collected using API, we collected the other datasets and extracted multi-source block-level features in the Google Earth Engine platform (Gorelick et al., 2017), and exported the derived results to be coupled with samples for offline classifications using different machine learning algorithms.

2.3. Mapping urban land use categories

Five main procedures are involved in mapping urban land use categories in this study (Fig. 2): preparation of multi-source open big data; generation of street blocks in metropolitan areas; feature extraction from multi-source geospatial big data; collection of training and validation samples; and mapping urban land use categories using automatic ensemble learning framework and accuracy assessment.

2.3.1. Generation of street blocks in metropolitan areas

The street block that represents a relatively homogeneous function was used as the basic unit for urban land use classification in this study (Hu et al., 2016; Liu and Long, 2016; Watts et al., 2007). We used the road centerlines of major roads and minor roads from OSM to generate street blocks in the five selected metropolitan areas. Street blocks are polygons bounded by road networks, thus we can use buffered road centerlines to divide the metropolitan areas into polygon-based blocks. Given the fact that roads at different types and locations have different widths (Gong et al., 2020a), we randomly selected 150 samples for major roads and 150 samples for minor roads across the states where the five metropolitan areas were located. For each sample, we measured its road width using high-spatial-resolution (HR) imagery in the Google Earth Pro software. Road widths varied among different states in terms of both major and minor roads (Fig. 3). Based on these findings, we used state-level thresholds (i.e., mean values) of the buffered road width for major roads and minor roads (Table 2). With the threshold of buffered road width and OSM road network, we derived the initial street blocks and further overlapped them with the GUB data in 2018 to achieve the final street blocks for these five metropolitan areas (Fig. 4).

2.3.2. Feature extraction from geospatial big data

Block-level features were extracted for urban land use classification from NAIP imagery, Sentinel-2 imagery, Sentinel-1 SAR imagery, VIIRS nighttime light imagery, WorldPop gridded population, Twitter check-in records, and topography datasets. Detailed descriptions of each categorized feature were provided as below and their specific items were summarized in Table 3.

2.3.2.1. Multi-spectral features and texture features from NAIP imagery. We mosaicked NAIP imagery together to generate a seamless composite

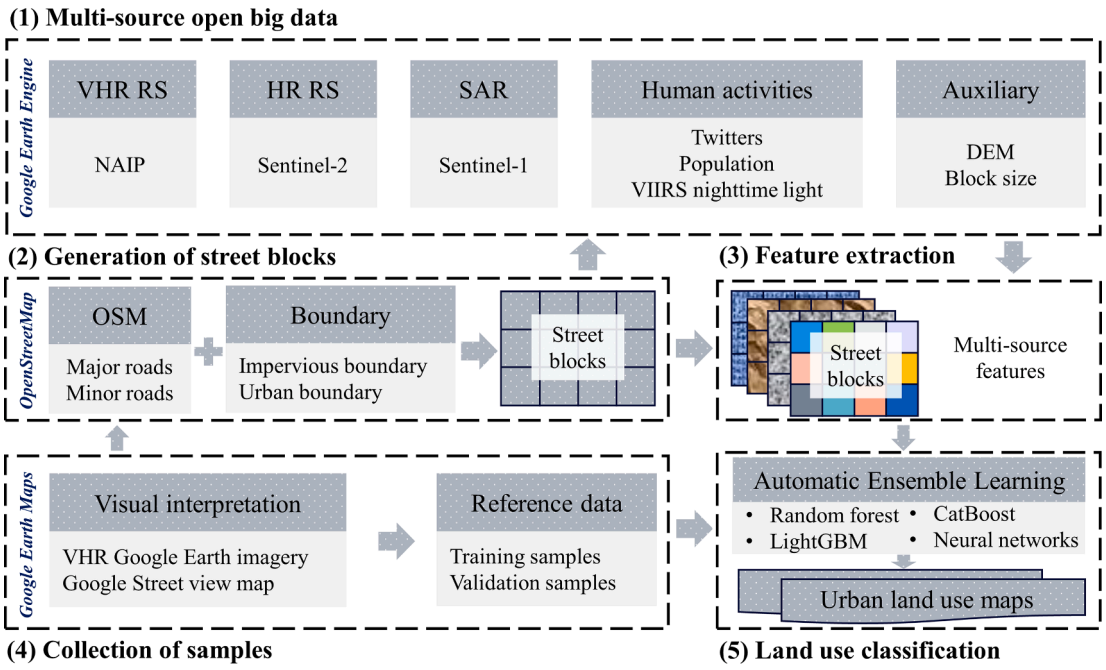

Fig. 2. Flowchart of the research data and methods. Noted that the abbreviations are: very-high-resolution (VHR), remote sensing (RS), high-resolution (HR), synthetic aperture radar (SAR), Visible Infrared Imaging Radiometer Suite (VIIRS), digital elevation model (DEM).

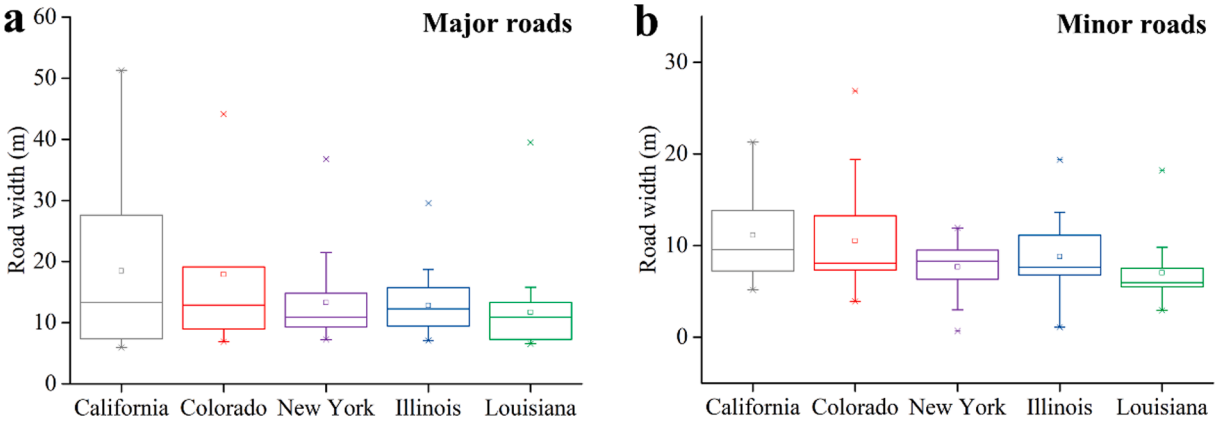

Fig. 3. Boxplots of road widths for major roads (a) and minor roads (b) in five states.

Table 2
Road widths (meters) used for generating street blocks.

Metropolitan (state)	San Francisco (California)	Denver (Colorado)	New York City (New York)	Chicago (Illinois)	New Orleans (Louisiana)
Major road	18.5	17.9	13.3	13.8	11.7
Minor road	11.1	10.5	7.7	8.4	7.0

for each metropolitan area. Given the fact that NAIP aerial imagery was collected without calibrating into radiance or reflectance, we directly used its digital number (DN) values of blue, green, red, and near-infrared bands. Similarly, we also used the band combination to indicate the concept of normalized difference vegetation index (NDVI = (NIR-Red)/(NIR + Red)) and normalized difference water index (NDWI = (Green-NIR)/(Green + NIR)). Due to the high spatial resolution of NAIP imagery, we extracted texture information including entropy and gradient from each band. Specifically, we calculated the entropy of blue, green, red, and near-infrared bands, with a square kernel of 4-pixel radius (approximately 2.4 or 4 m in radius) to quantify the adjacent texture information. Regarding the gradient, we first computed the image gradient in both horizontal and vertical directions, and then calculated the magnitude of the gradient through Eq. (1),

$$grad = \sqrt{grad_x^2 + grad_y^2} \quad (1)$$

where $grad_x$ and $grad_y$ represent the gradient in x-axis and y-axis directions, respectively.

Finally, all these available and derived bands were aggregated over each street block to obtain their mean and standard deviations (Table 3).

2.3.2.2. Multi-spectral features and texture features from Sentinel-2 imagery. We used the co-constellation Sentinel-2A/B imagery from January 1 to December 31, 2018 to extract multi-spectral features. We first did a pixel-based quality check to screen and filter out the poor-quality surface reflectance values using cloud mask and quality assessment (QA) information in the Sentinel-2A/B metadata. This eliminated the observations contaminated by clouds and shadows from the entire

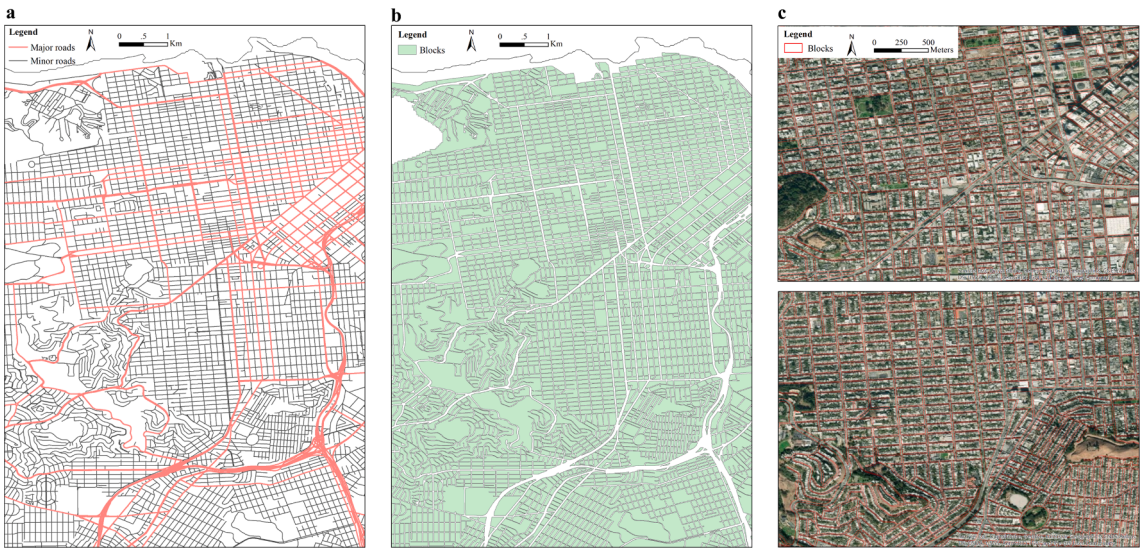

Fig. 4. Illustration of the generation of land use blocks. (a) The OSM major (red lines) and minor (black lines) roads, (b) the derived blocks as the basic unit for land use classification, and (c) the zoomed-in subsets of derived blocks overlaid on the top of high-resolution Google Earth imagery. Noted that it is an example of the City of San Francisco, California. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Sentinel-2 archive. NDVI values were then calculated from the retained reflectance in the Red and NIR bands for each pixel. We further used the pixel-based maximum NDVI values as the quality index to merge the whole-year Sentinel-2A/B images and derive the cloud-free greenest Sentinel-2 composite in 2018 (i.e., blue, red, green, and near-infrared bands at 10-m spatial resolution, and four red edge bands and two shortwave-infrared bands at 20-m spatial resolution). Using this base image with 10-m bands, we calculated mean and standard deviations of blue, green, red, and near-infrared bands, NDVI, NDWI for each street block. Similarly, using the base image with remaining 20-m bands, we calculated mean and standard deviations of four red-edge bands and two shortwave-infrared bands for each street block. Additionally, we calculated the entropies of inclusive bands and remote sensing indices using a square kernel of 4-pixel radius, and derived their mean and standard deviations over each street block (Table 3).

2.3.2.3. Height features from Sentinel-1 SAR imagery. The height of building structure is one of explanatory variables accounting for the difference in land use type. For example, compared with residential and industrial regions, commercial service and business offices are generally with high-rise buildings. Although OSM data include certain building height information, they are not spatially complete because of the crowdsourcing collection and estimation. However, backscatter coefficients from the Sentinel-1 GRD VV and VH have been verified to show a high correlation with building height (Frantz et al., 2020; Koppel et al., 2017; Li et al., 2020b). Therefore, we directly used the backscatter coefficients of VV and VH from Sentinel-1 GRD data as measures of building height. Additionally, we also included the angle band representing the approximate viewing incidence angle, to characterize the viewing geometry of quantitative VV and VH observations accounting for potential differences across spatial and temporal scales. All these three bands were further aggregated over each street block to derive their mean and standard deviation values (Table 3).

2.3.2.4. Auxiliary features. We averaged the 12-month VIIRS stray light corrected DNB images to obtain the monthly nighttime light imagery, and calculated the mean value of digital number for each street block (Table 3). The 100-m gridded WorldPop data in 2018 were used as the reference to calculate the mean of population for each street block. In addition to the total number of Tweets, we calculated the number of geotagged Tweets during three time periods (0:00–8:00, 8:00–16:00, and 16:00–24:00) within the generated street blocks (Table 3). In addition, the size of street block may help to account for the difference in land use type. For example, compared with industrial and entertainment/recreational regions, residential land uses typically occurring smaller blocks. Therefore, we also included the block size as one predictor in the classification (Table 3).

2.3.3. Collection of training and validation samples

We applied the two-level classification system proposed in our previous paper that comprises Essential Urban Land Use Categories (EULUC): residential, entertainment, transportation, industrial, and office (Table 4) (Gong et al., 2020a).

According to the defined classification scheme, we collected ground truth samples using the combination of the following two approaches: (i) visual inspections using HR Google Earth imagery, Google Street Views, Google Map point of interests (POIs), and 3-D modelled imagery; and (ii) reference from regional survey-based land use maps. The crowdsourcing information provided us relatively high confidential references to interpret the land use type of sampled blocks. Practically, we first generated randomly distributed blocks using stratified sampling strategies for each metropolitan area and then overlapped them with Google Earth imagery for interpretations (Fig. 5). We also provided examples of three representative features in different land use categories to illustrate the difference in multi-source data features (Figs. S1–3). The sampling rate was 13.66%, 4.69%, 6.86%, 3.97%, and 7.94% for San Francisco, Denver, New York City, Chicago, and New Orleans, respectively (Table S1) and the spatial distribution of samples was provided in Fig. S4. All samples in terms of Level-II classes collected in five metropolitan areas were summarized in Table 5.

2.3.4. Mapping essential urban land use categories using automatic ensemble learning

We trained the classification models using automatic ensemble learning through multi-layer stacking. As the basic architecture of multi-layer stacking shown in Fig. 6, L_n represents the nth stack layer consisting of several individual models (Base Learner-BL) and a meta-learning model (ML). For each stack L_n, the derived results from BL_n and original input features are stacked together for training ML_n. Iteratively, each model in BL_n will be individually trained using the output from ML_{n-1} in the previous stack layer. The input features from original data are also concatenated into input vectors of ML_{n-1}, which enables

Table 3
Summary of features used in block-level mapping of essential urban land use categories.

Data source	Features	Variables
NAIP	Mean of blue, green, red, near-infrared, NDVI, NDWI, entropy, gradient bands	B_mean, Bent_mean, Bgrad_mean, G_mean, Gent_mean, Ggrad_mean, R_mean, Rent_mean, Rgrad_mean, N_mean, Nent_mean, Ngrad_mean, NDVI_mean, NDWI_mean, NDVIgrad_mean, NDWIgrad_mean
	Standard deviation of blue, green, red, near-infrared, NDVI, NDWI, entropy, gradient bands	B_std, Bent_ std, Bgrad_ std, G_ std, Gent_ std, Ggrad_ std, R_ std, Rent_ std, Rgrad_ std, N_ std, Nent_ std, Ngrad_ std, NDVI_std, NDWI_std, NDVIgrad_std, NDWIgrad_std
Sentinel-2A/B	Mean of B2, B3, B4, B5, B6, B7, B8A, B11, B12, NDVI, NDWI, and entropy bands	S2_B2_mean, S2_B2ent_mean, S2_B3_mean, S2_B3ent_mean, S2_B4_mean, S2_B4ent_mean, S2_B5_mean, S2_B5ent_mean, S2_B6_mean, S2_B6ent_mean, S2_B7_mean, S2_B7ent_mean, S2_B8A_mean, S2_B8Aent_mean, S2_B11_mean, S2_B11ent_mean, S2_B12_mean, S2_B12ent_mean, S2_NDVI_mean, S2_NDVIent_mean, S2_NDWI_mean, S2_NDWIent_mean, S2_NDWI_mean, S2_NDWIent_mean
	Standard deviation of B2, B3, B4, B5, B6, B7, B8A, B11, B12, NDVI, NDWI, and entropy bands	S2_B2_std, S2_B2ent_std, S2_B3_std, S2_B3ent_std, S2_B4_std, S2_B4ent_std, S2_B5_std, S2_B5ent_std, S2_B6_std, S2_B6ent_std, S2_B7_std, S2_B7ent_std, S2_B8A_std, S2_B8Aent_std, S2_B11_std, S2_B11ent_std, S2_B12_std, S2_B12ent_std, S2_NDVI_std, S2_NDVIent_std, S2_NDWI_std, S2_NDWIent_std, S2_NDWI_std, S2_NDWIent_std
Sentinel-1 GRD	Mean of VV, VH, angle	VV_mean, VH_mean, slope_mean
	Standard deviation of VV, VH, angle	VV_std, VH_std, slope_std
VIIRS DNB	Mean of nighttime light	Nighttime_light
Twitter	Counts of Tweets during 0:00–8:00, 8:00–16:00, and 16:00–24:00	TW_P1, TW_P2, TW_P3, TW_ALL
WorldPop	Mean of population	Pop_mean
NED-DEM	Mean of elevation and slope	Elevation, slope
Block	Area of street block	Block_size

Table 4
The two-level Residential-Entertainment-Transportation-Industrial-Office classification schemes.

Level-I	Level-II	Descriptions
01 Residential	0101 Residential	Houses and apartment buildings—places where people live.
02 Entertainment/ Recreational	0201 Sport and cultural	Lands used for public sports and training, cultural services, including gym centers, libraries, museums, exhibition centers, etc.
	0202 Park and greenspace	Parks and greenspace lands used for entertainment and environmental conservation.
03 Transportation	0301 Road	Paved roads including freeways, major and minor city-roads.
	0302 Transportation station	Transportation facilities including motor, bus, train stains and ancillary facilities.
	0303 Airport	Airports for civil, military, and mixed uses.
04 Industrial	0401 Industrial	Land and buildings used for manufacturing, warehouse, mining, etc.
05 Office	0501 Business office	Buildings where people work, including office buildings, and commercial office places for finance, internet technology, e-commerce, media, etc.
	0502 Commercial service	Houses and buildings for commercial retails, restaurants, lodging, and entertainment.
	0503 Administrative	Lands used for government, military, and public service agencies.
	0504 Educational	Lands used for education and research, including schools, universities, institutes and their ancillary facilities.
	0505 Medical	Lands used for hospitals, disease prevention, and emergency services.

higher-layer stackers to revisit the original data in training process for more robust and accurate model performance. In addition to multi-layer stacking, we employed 5-fold cross validation to reduce model variability and mitigate over-fitting problems in the automatic ensemble learning. Specifically, for any model at any stack layer, we randomly split the input data into 5 folds with equal size (stratified sampling based on labels). Among the 5-fold subsamples, one-fold subsample was retained for model validation, and the remaining 4 folds were used as training data. The cross-validation process in 5 replicated runs were then averaged to produce an average estimation.

We here applied the AutoGluon package (Erickson et al., 2020) to implement automatic ensemble learning. AutoGluon is an open-source Python library that automates the process of model selection, hyperparameter tuning, and model ensembling during machine learning (Erickson et al., 2020). By setting parameters such as bagging strategy, stack level, and model parameters, AutoGluon will automatically train and ensemble multiple models to obtain the best classification result within a given time. In this study, the parameter 'num_bag_folds' was set to 5 for 5-fold cross-validation, 'auto_stack' was set to True for automatic multi-layer stacking, and 'time_limit' was set to 3600 for a maximum learning time of 3600 s in total. Base models here included random forest, extremely randomized trees, light gradient boosting machine (LightGBM) (Machado et al., 2019), CatBoost boosted trees (Dorogush et al., 2018), and neural networks. For each base model, we tested classification performance under 20 sets of parameter combinations (Table S2). For neural networks, the choice of hyperparameters did have certain impact on the classification results, with a validation accuracy ranging from 67.70% to 81.47% and a training accuracy ranging from 67.97% to 81.29% for Level-II classification. In contrast, as for random forest, extremely randomized trees, and CatBoost boosted trees, the classification accuracy among models did not vary significantly with standard deviations less than 1%. We therefore selected parameters with the highest validation accuracy as the optimal parameters for each base model and used them for automatic ensemble learning accordingly. Specifically, For random forest and extremely randomized trees, the number of trees was 500 and 450, respectively, and the criterion was set to Gini. For LightGBM, the learning rate was 0.08, the number of leaves was 138, and the boosting type was set to traditional gradient boosting decision tree. In CatBoost boosted trees, the number of iterations was 10,000, and the learning rate was 0.1. In neural networks, the epoch was set to 10, the learning rate was 0.005, the activation function was ReLU, the batch size was 512, and the dropout probability was set to 0..

2.3.5. Accuracy assessment and comparison

We evaluated the accuracy of model performance in terms of two aspects. First, we grouped all samples from five metropolitan areas and split them into 75% for training and 25% for validation, which was treated as accuracy assessment for the global model. Second, we conducted the accuracy assessment individually across each of the metropolitan areas, which was termed as localized models. Similarly, 75% of samples were used for training and the remaining 25% of samples were

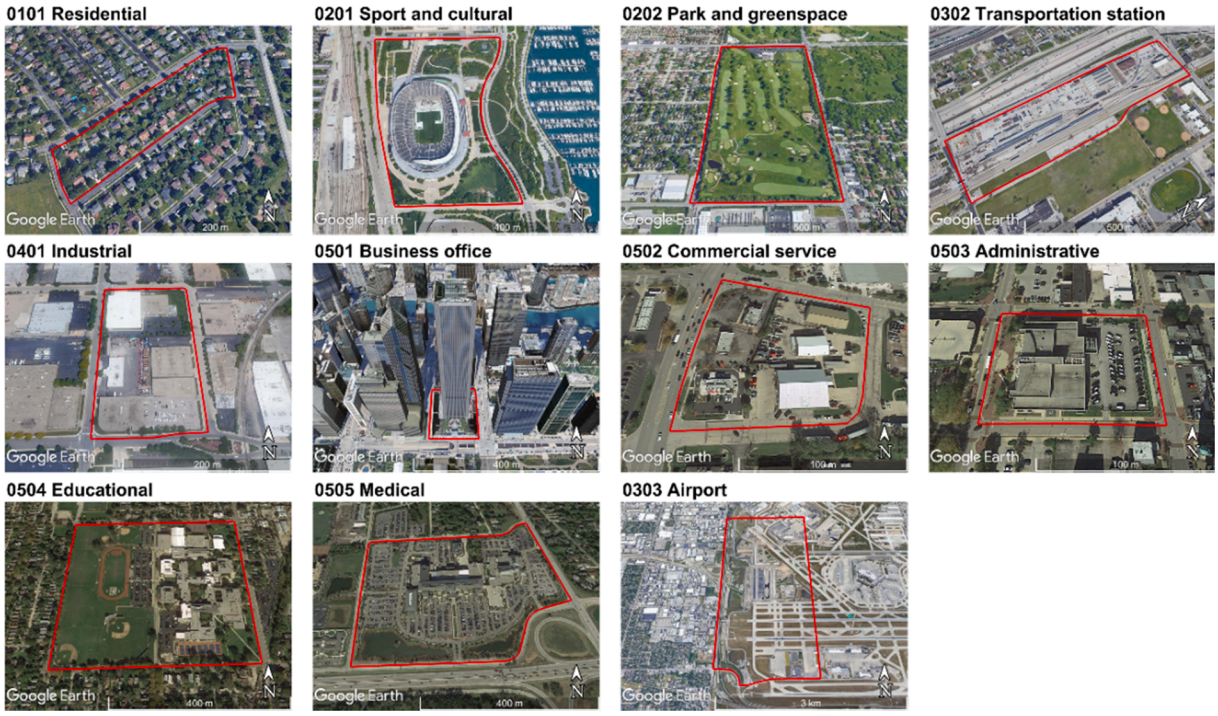

Fig. 5. Examples of sampled blocks of different urban land use categories overlaid with Google Earth imagery in Chicago.

Table 5
The number of collected samples of different Level II categories in different metropolitan areas.

Level-II	Number of samples					
	San Francisco	Denver	New York	Chicago	New Orleans	In sum
0101	2744	946	1148	1823	1062	7723
0201	34	19	74	84	38	249
0202	106	80	160	172	68	586
0302	23	2	62	89	6	182
0303	3	–	6	2	–	11
0401	502	100	143	233	32	1010
0501	572	24	170	97	15	878
0502	753	102	166	326	85	1432
0503	11	13	34	43	18	119
0504	215	76	217	187	56	751
0505	31	21	41	43	21	157

used for validation. During the training process, we used 5-fold cross-validation to achieve the averaged overall accuracy as an indicator for model training performance. To better inform the accuracy assessment, we included Kappa Coefficient for validation accuracy assessments. For model comparison, we further included the weighted F1 score to better justify the accuracy assessment. Specifically, the F1 score can be interpreted as an average of the precision (user's accuracy) and recall (producer's accuracy), where a F1 score reaches its best value at 1 and worst value at 0. The relative contribution of precision and recall to the final F1 score is equal, and the calculation of F1 score is as below.

$$F1 = 2*(precision*recall)/(precision + recall) \quad (2)$$

According to Eq. (2), we can calculate the F1 score for each class, respectively. By accounting for the weight of each class, defined by the number of true instances for each class, we can calculate the weighted F1 score as another indicator for a more justified accuracy assessment.

To differentiate the relative contribution of inclusive variables in the full model, we calculated the mean decrease of prediction accuracy to quantify variable importance (Erickson et al., 2020). Given the fact that

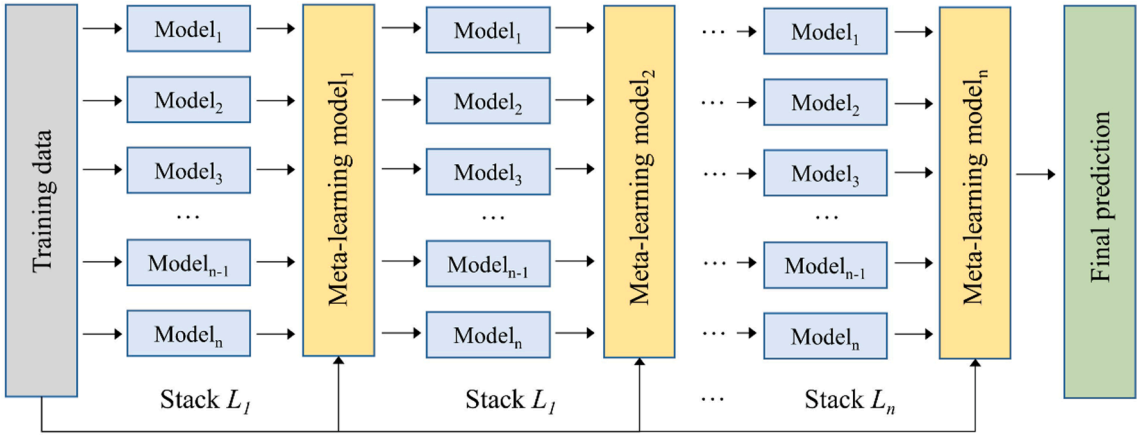

Fig. 6. The architecture of automatic ensemble learning.

we have included multi-source datasets into this block-level urban land use classification, it will be useful to identify specific types of data sources and features with a higher contribution to the classification performance, thus gaining potential insights of data selections and method transferability for multi-scale land use classification practices. Therefore, we conducted another set of classification comparisons using different combinations of inclusive features: (1) NAIP imagery only; (2) Sentinel-1 imagery (S1) only; (3) Sentinel-2 imagery (S2) only; (4) auxiliary data including block size, population, Twitter locations, topography and nighttime light (i.e., others) (5) NAIP and S1; (6) NAIP and S2; (7) NAIP and others; (8) S1 and S2; (9) S1 and others; (10) S2 and others; (11) NAIP, S1, and S2; (12) NAIP, S1, and others; (13) NAIP, S2, and others; (14) S2, S1, and others. All 14 scenarios were trained and validated using the 5-fold cross-validation scheme as described in Section 2.3.4.

2.3.6. Feature transferability for urban land use classification

In addition to the models trained using samples from all five metropolitan areas (i.e., global models), we also developed independent models of urban land use classification metropolitan by metropolitan (i.e., localized models). Using the collected samples from each metropolitan, we were able to train metropolitan-wise classification models. We proposed a paired-metropolitan training and validation scheme to investigate whether localized features and models could be transferred to non-local urban land use classification. For example, we first built the machine learning algorithm using training samples collected in the San Francisco Bay area, and then applied it in urban land use classification in New York City. The performance of classification practice was independently validated using samples collected in New York City. Following this protocol, we validated the classification performance through all paired-metropolitan training practices. The overall accuracy was used as the major indicator to account for its transferability.

3. Results

3.1. Comparison of different models

Results of the multi-model comparison in Tables 6 and 7 showed that multi-stacking ensemble models achieved the best performance in training accuracy for Level-I classes (overall accuracy: 90.93%, and Kappa coefficient: 0.84) and Level-II classes (overall accuracy: 86.14%, and Kappa coefficient: 0.77), followed by random forest, LightGBM, ExtraTrees, NeuralNetwork, and CatBoost models with slightly lower overall. Meanwhile, we found that the multi-stacking strategy did help improve model performance in land use classification. For example, the overall accuracy could be improved from 89.72% to 90.93% when we increased the multi-stacking ensemble model from 2 layers to 3 layers. This general pattern was also identified in other base models such as Extremely Randomized Trees and Random Forest models. Due to the relatively superior performance of multi-stacking ensemble models (Tables 6 and 7), we chose them as the main models for subsequent analysis.

Although multi-stacking ensemble models achieved relatively robust and promising classification performance with overall accuracies of 90.93% and 86.14% for Level-I and Level-II, the confusion between classes (at both Level-I and Level-II) was still a challenge. As the confusion matrixes shown in Tables S3-4, the producer's accuracy (PA) was quite high for the majority of classes in general, whereas the user's accuracy (UA) for certain classes was not plausible, for example, commercial and transportation lands at Level-I scheme, and sport and cultural, transportation station, and medical lands at Level-II scheme achieved UA less than 67%. The relatively lower accuracy may result from two major factors. First, the samples of these land use types are much less than those of residential and industrial lands. Second, the similarity in spectral, textural, and anthropogenic characteristics of different land uses will lead to confusion between classes using machine learning algorithms. For example, sport and cultural lands (e.g., gyms, libraries, museums) will be confused with residential lands (Table S4), and medical lands will be confused with administrative and educational lands (Table S4).

3.2. Land use classification maps for five metropolitan areas

We derived the urban land use classification maps for these five metropolitan areas using the multi-stacking ensemble models, which achieved the best training accuracy for both Level-I and Level-II classification schemes (Tables 6 and 7). Given the contrasting characteristics in urban morphologies and components, the metropolitan-wise models were used. As shown in Fig. 7, the derived block-level maps could accurately depict the geographic distribution, pattern, and composition of various land use types for each metropolitan area. For example, in the city of San Francisco, the residential lands were mainly distributed in the western side, and the business office and commercial areas were distributed in the upper eastern side (Fig. 7a). In Denver, the majority of industrial land use was distributed in the northeast side while the residential and commercial lands in the middle to southwest side (Fig. 7b). In New Orleans, the majority of residential, business office and commercial areas were distributed in the north of the Mississippi River, and we could clearly identify several industrial land uses along with the Mississippi River (Fig. 7c). In Chicago, residential, business office and commercial areas were densely distributed along the Lake Michigan, while the spatial coverage of parks and greenspaces were much larger in the suburban areas (Fig. 7d). New York City was quite different, and business offices are densely distributed in Manhattan (Fig. 7e).

Quantitative results revealed that the overall accuracy of land use classification maps was consistently higher than 87% and 81% for Level-I and Level-II schemes in all five metropolitan areas (Table 8). Specifically, the highest Level-I overall accuracy could be up to 95.38% and 93.71% for Denver and New Orleans, while their corresponding Level-II overall accuracy could be up to 91.62% and 89.71%.

Table 6
Level-I accuracy comparison of different models in terms of overall accuracy (OA), Kappa Coefficient (Kappa), weighted F1 score, training time and the number of stack layers.

Model	OA	Kappa	Weighted F1 score	Training time (s)	Number of stack layers
WeightedEnsemble_L3	**90.93%**	**0.8405**	**90.60%**	3284.56	3
RandomForest_BAG_L2	90.77%	0.8376	90.42%	1691.14	2
LightGBM_BAG_L2	90.77%	0.8374	90.43%	2018.04	2
ExtraTrees_BAG_L2	90.64%	0.8354	90.29%	1616.73	2
NeuralNetMXNet_BAG_L2	90.58%	0.8346	90.29%	1856.80	2
CatBoost_BAG_L2	90.55%	0.8337	90.23%	2479.29	2
CatBoost_BAG_L1	89.88%	0.8214	89.46%	1132.00	1
WeightedEnsemble_L2	89.72%	0.8180	89.23%	1572.69	2
LightGBM_BAG_L1	89.56%	0.8148	89.01%	229.13	1
NeuralNetMXNet_BAG_L1	88.36%	0.7936	87.81%	143.49	1
RandomForest_BAG_L1	87.37%	0.7737	86.52%	65.16	1
ExtraTrees_BAG_L1	86.83%	0.7639	85.90%	18.22	1

Table 7
Level-II accuracy comparison of different models in terms of overall accuracy (OA), Kappa coefficient (Kappa), weighted F1 socre, training time and the number of stack layers.

Model	OA	Kappa	Weighted F1 score	Training time (s)	Number of stack layers
WeightedEnsemble_L3	**86.14%**	**0.7733**	**84.82%**	3247.34	3
LightGBM_BAG_L2	86.07%	0.7720	84.77%	1558.68	2
CatBoost_BAG_L2	85.88%	0.7704	84.79%	2267.11	2
ExtraTrees_BAG_L2	85.82%	0.7678	84.35%	1185.04	2
RandomForest_BAG_L2	85.72%	0.7663	84.19%	1243.50	2
WeightedEnsemble_L2	85.44%	0.7600	83.83%	1152.12	2
CatBoost_BAG_L1	85.25%	0.7577	83.76%	844.22	1
NeuralNetMXNet_BAG_L2	84.87%	0.7533	83.75%	1582.80	2
LightGBM_BAG_L1	84.42%	0.7419	82.53%	98.20	1
NeuralNetMXNet_BAG_L1	83.41%	0.7286	82.25%	153.41	1
RandomForest_BAG_L1	81.88%	0.6964	79.62%	38.21	1
ExtraTrees_BAG_L1	81.25%	0.6852	78.97%	14.37	1

3.3. Contribution of inclusive features

According to the variable importance quantified by mean decrease in the prediction accuracy (Fig. 8, Table S5), we could identify that, for both Level-I and Level-II classifications, mean population (pop_mean), elevation, mean of angle (angle_mean), and block size (F_AREA) were leading important predictors. It makes sense that population density accounts most for differentiating different urban land use types, because the function of urban land use is characterized by human activities, for example, the population in residential areas is significantly higher than that in industrial areas. Similarly, street block size is another factor representing the extent of different urban land uses. For example, residential land uses are always characterized with much denser street networks, thus leading to a smaller block size. In contrast, the industrial and transportation land uses always require much larger land availability, which are characterized by the larger block size. Meanwhile, as revealed in Fig. 8, elevation is another important factor since the location of different urban land uses will be influenced by topographic effects such as elevation. We also identified significant variations of variable importance across these five metropolitan areas (Fig. S5-9). The differences of physical environment and socio-economic status among the five metropolitan areas will be the main driver accounting for the changes in variable importance.

The inclusion of all features achieved the best performance of urban land use classification. Scenarios with different combinations of input features yielded different classification outcomes. Given the fine spatial details from NAIP imagery, we found that classification results from NAIP derived features (Level-I: 87.56% and Level-II: 81.95%) achieved much higher accuracies than that from Sentinel-1 imagery (Level-I: 80.62% and Level-II: 74.24%) or Sentinel-2 imagery (Level-I: 77.63% and Level-II: 72.46%). However, expect for scenarios with the addition of NAIP, the integration of Sentinel-1, Sentinel-2, and other auxiliary data could achieve a relatively plausible classification performance with overall accuracies of 88.42% for Level-I classes and 81.44% for Level-II classes (Table 9), which was higher than the combination of S1 and S2 (Level-I: 85.37% and Level-II: 78.08%), the combination of S1 and others (Level-I: 87.82% and Level-II: 81.12%), the combination of S2 and others (Level-I: 84.93% and Level-II: 78.52%), S1 only (Level-I: 80.62% and Level-II: 74.24%), S2 only (Level-I: 77.63% and Level-II: 72.46%), and others only (Level-I: 79.82% and Level-II: 74.90%). The combination of moderate-resolution spectral, textural, height, and topographic information from satellite-based observations and human activities from geospatial big data is able to provide multi-dimensional lens to uncover the composite, pattern, and distribution of urban land use types.

3.4. Transferability of features and models across metropolitan areas

As shown in Table 10, the overall accuracies of training models in each metropolitan area were above 87% and 81% for Level-I and Level-II schemes. By applying the model trained using samples collected in one metropolitan to other metropolitan areas, the derived overall accuracies differed a lot across different metropolitan areas (Table 10). For example, the models trained in Denver achieved quite plausible performance in classifying Level-I urban land uses in San Francisco with an overall accuracy of 76.60%, but yielded much lower accuracies in Chicago and New York. The models trained in New York even achieved better performance in classifying Level-I urban land uses in San Francisco with an overall accuracy of 81.89%. However, most localized models did not perform well in non-local regions, especially for the Level-II classification practices with overall accuracies of less than 50% in several metropolitan areas.

We selected four important features to show their difference among five metropolitan areas in terms of different urban land use categories (Fig. 9). We found there was a significant difference in the mean population within residential (Fig. 9a) and entertainment/recreational blocks (Fig. 9b) between Chicago and other metropolitan areas; and the elevation was much higher in Denver than other metropolitan areas (Fig. 9a-d). Block size was more comparable among different metropolitan areas for residential blocks (Fig. 9a) but showed more variations for other land use blocks such as entertainment/recreational (Fig. 9b), industrial (Fig. 9c), and office lands (Fig. 9d). The difference in the averaged angle among different metropolitan areas was similar across four selected land use types (Fig. 9a-d). This dived-in analysis helps gain insights that the suitability of model transferability should be dependent on the similarity in characteristics of different metropolitan areas.

4. Discussion

The integration of features extracted from VHR satellite observations is able to provide important information for urban land use classification, especially for unique textures of buildings and infrastructures. However, the majority of VHR satellite imagery is publicly inaccessible and expensive (Chen et al., 2017a), thus hindering VHR feature extraction and semantic classification at large scales. A cost-effective data source is the freely available moderate-spatial-resolution images such as Sentinel-2 and Landsat series. The 30-m resolution of Landsat imagery is still challenging in spatial details to identify building objects at the block level, but the 10-m pixels from Sentinel-2 observations are promising to differentiate spectral and texture characteristics among different land use types at the block level (Gong et al., 2020a; Su et al., 2020; Tu et al., 2020). Compared with the classification performance using VHR NAIP based features, the derived classification using Sentinel-1/2 based features with other auxiliary features achieved comparable overall accuracies (Table 9). This finding elucidates the possibility of accurate block-level urban land use classification at large scales without VHR satellite imagery, and we could rely on global free-accessible Sentinel-1/2 for substitute. Additionally, the VHR remotely sensed data is limited by its spectral bands from visible to near infrared wavelength. However, the red edge bands, shortwave infrared bands,

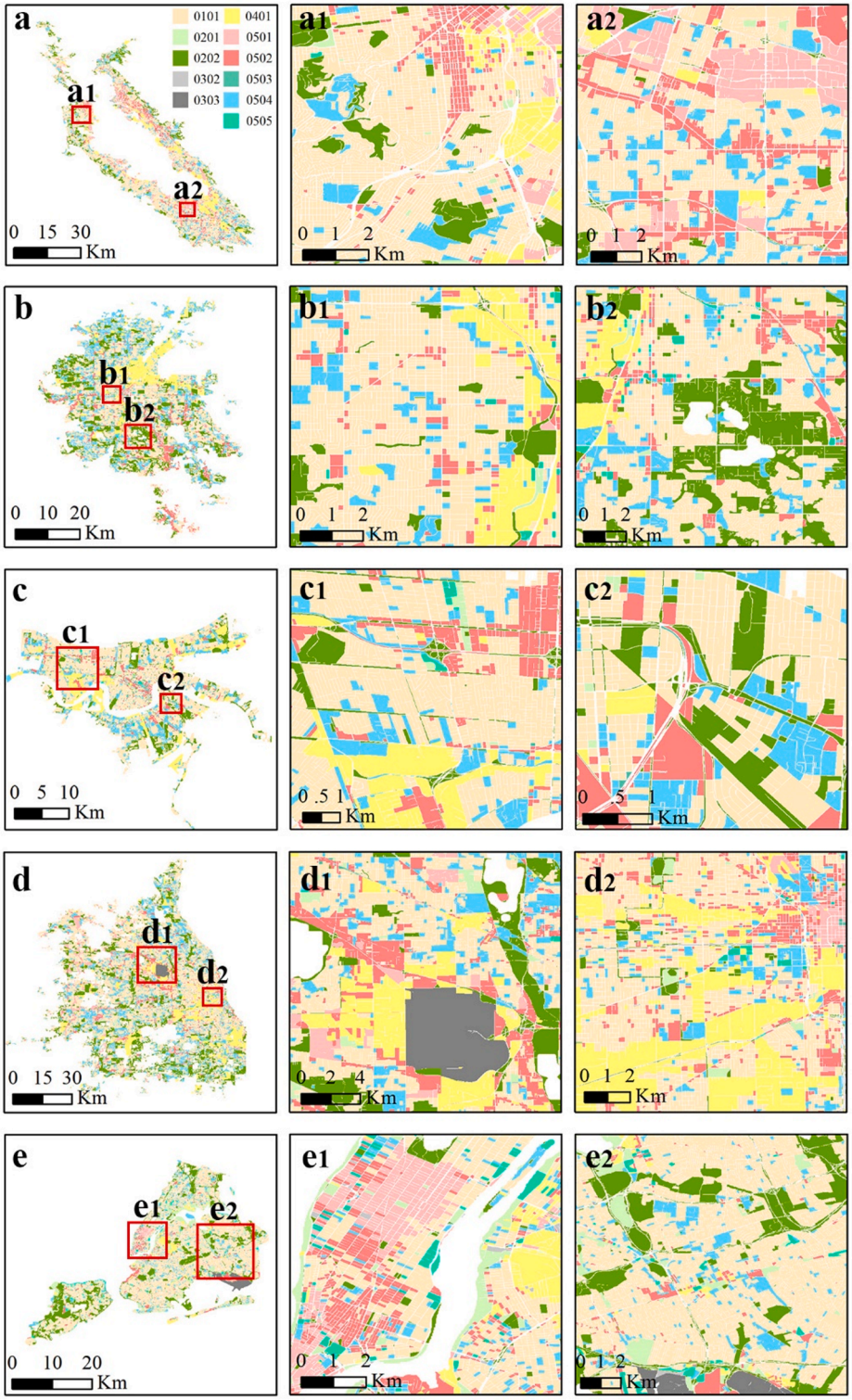

Fig. 7. Urban land use classification map of (a) the San Francisco Bay area, (b) Denver, (c) New Orleans, (d) Chicago, and (e) New York City. The middle and right panels are zoomed-in subsets from corresponding metropolitan areas.

and derived remote sensing indices such as NDWI and NDBI from Sentinel-2 provide more features to feed machine learning algorithms in urban land use classification. Backscatter coefficients from Sentinel-1 GRD observations have proven to be useful in urban land use classification, and in this study, the overall accuracies of classification practices using only Sentinel-1 based features are up to 80.62% and 74.24% for Level-I and Level-II classes, respectively (Table 9), which is even higher to the results using only Sentinel-2 based features. However, the integration of both Sentinel-1 and Sentinel-2 observations can lead to an increase in the overall accuracy of 5%-8% for Level-I classes and 4%-6% for Level-II classes (Table 9). Human activities and auxiliary features such as population and nighttime light are also playing important roles in increasing classification accuracies (Fig. 8). Although topography features do not change dramatically in the urban environment, we do

Table 8
Quantitative comparison of Level-I and Level-II classification in five metropolitan areas using overall accuracy (OA) and Kappa Coefficient (Kappa).

Metropolitan areas	Level-I		Level-II	
	OA	Kappa	OA	Kappa
San Francisco	92.63%	0.87	88.54%	0.82
Denver	95.38%	0.90	91.62%	0.83
New Orleans	93.71%	0.84	89.71%	0.75
Chicago	91.61%	0.86	86.32%	0.78
New York	87.21%	0.80	81.62%	0.75

find that for the global model, elevation plays an important role in both Level-I and Level-II classifications (Fig. 8). However, the role of topography features in urban land use classification varies across different regions. For example, elevation is the first or second leading variable in Level-I/II classification in San Francisco Bay area (Fig. S5) and Denver (Fig. S6), but it is not ranked as the leading variable of relative contribution in other three metropolitan areas (Fig. S7-9). With the investigation of variable contribution using an expansive set of multi-source open big data, our results revealed that the integration of free-accessible datasets including Sentinel-1, Sentinel-2, nighttime lights, and population could be robust and cost-effective input features for large-scale mapping of urban land use categories. On the other hand, feature combination should depend on the urban landscape of specific study areas, since the contribution of datasets and predictor features will be different across different urban landscapes (Fig. S5-9), especially for different countries and continents.

Mixed land uses have been challenging to urban land use categories

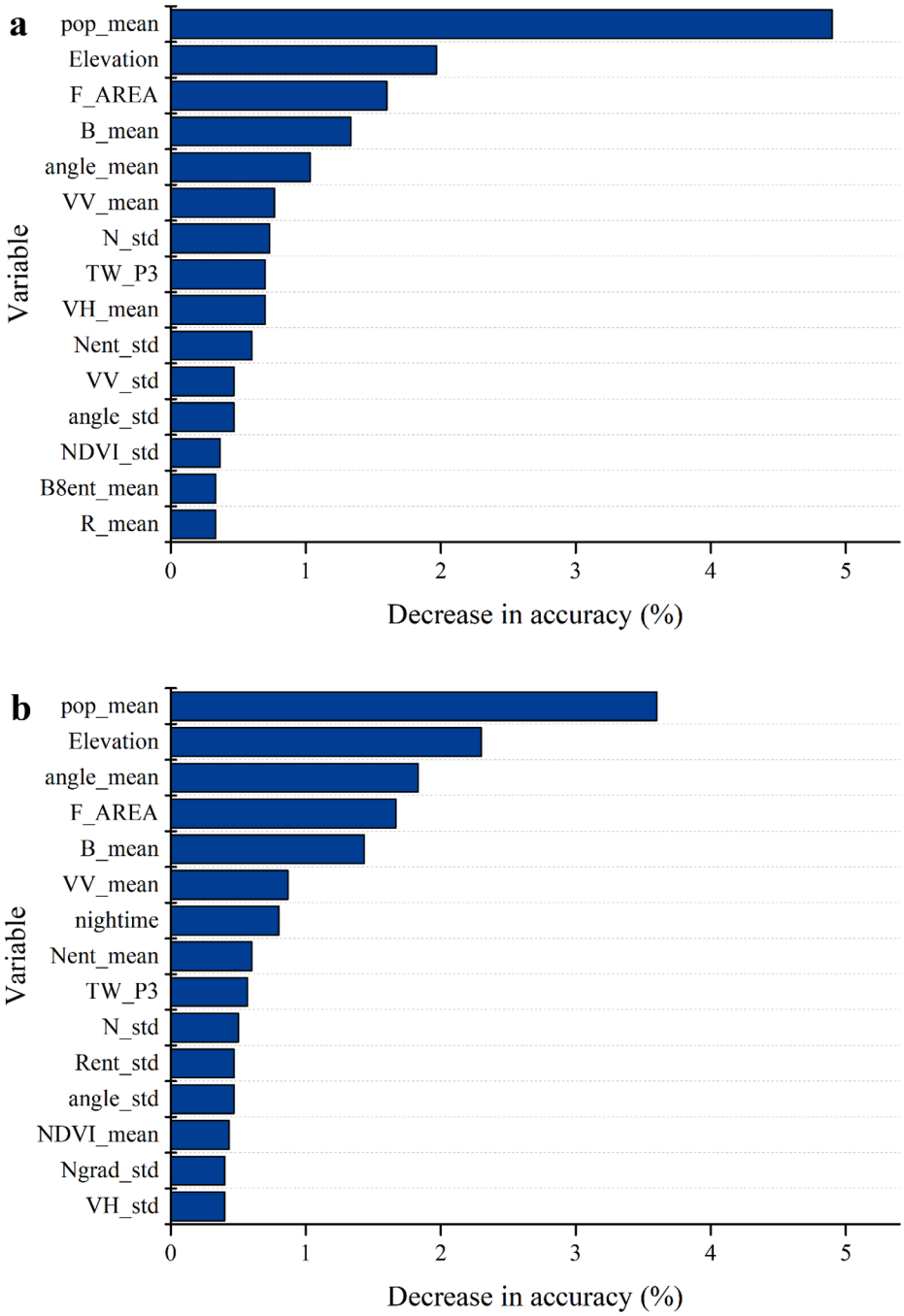

Fig. 8. Variable importance in terms of mean decrease in the prediction accuracy for (a) Level-I and (b) Level-II classification schemes.

Table 9
Accuracy comparison of Level-I and Level-II classification using different combinations of features. Noted that OA represents overall accuracy and Kappa represents Kappa Coefficient.

Categories		NAIP	S1	S2	Others	NAIP + S1	NAIP + S2	NAIP + others
Level-I	OA	87.56%	80.62%	77.63%	79.82%	88.48%	86.87%	89.40%
	Kappa	0.78	0.65	0.58	0.63	0.80	0.77	0.81
Level-II	OA	81.95%	74.24%	72.46%	74.90%	82.87%	81.25%	84.52%
	Kappa	0.70	0.56	0.50	0.57	0.72	0.69	0.75
Categories		S1 + S2	S1 + others	S2 + others	NAIP + S1 + S2	NAIP + S1 + others	NAIP + S2 + others	S1 + S2 + others
Level-I	OA	85.31%	87.82%	84.93%	89.05%	90.13%	88.67%	88.42%
	Kappa	0.74	0.78	0.73	0.81	0.83	0.80	0.80
Level-II	OA	78.08%	81.12%	78.52%	83.06%	85.22%	84.30%	81.44%
	Kappa	0.63	0.69	0.64	0.72	0.76	0.74	0.69

Table 10
Comparison of the Level-I and Level-II overall accuracies using paired-metropolitan training and validation experiments.

Level-I overall accuracy (%)					
	Validate				
Train	**San Francisco**	**Denver**	**New Orleans**	**Chicago**	**New York**
San Francisco	92.63	21.68	39.71	52.00	76.40
Denver	76.60	95.38	58.00	69.16	60.00
New Orleans	74.60	31.21	93.71	38.32	54.96
Chicago	29.89	51.73	24.57	91.61	22.34
New York	81.89	70.23	68.00	75.23	87.21
Level-II overall accuracy (%)					
	Validate				
Train	**San Francisco**	**Denver**	**New Orleans**	**Chicago**	**New York**
San Francisco	88.54	11.85	26.29	50.97	62.34
Denver	10.50	91.62	1.71	8.00	7.75
New Orleans	59.62	11.27	89.71	33.03	37.30
Chicago	19.07	69.94	33.43	86.32	10.27
New York	65.95	74.43	74.57	61.29	81.62

mapping. The low accuracy of EULUC-China practice mainly results from the mixed land use within land blocks. Generally, the overall accuracy decreases rapidly with the increase in land use mixture (Gong et al., 2020a). Although we adopted the same protocol of using the OSM road network to generate polygon-based blocks, the derived street blocks in five metropolitan areas of the US are in much finer scales with less mixed land uses, due to the following two reasons: (1) the data coverage of OSM road networks is much higher in the US with almost complete coverage spatially, leading to pure land uses for generated street blocks; and (2) land use types are more unique in urban planning practices in the US (Sarzynski et al., 2014). Therefore, the road network based generation of street blocks is cost-effective and practically applicable for urban land use classification. However, for areas with highly mixed land uses, an optimal approach will be to refine the division of land blocks using more detailed road networks and multi-scale image segmentation (Tu et al., 2020). In addition to mixed land use and block size, the impact of the number of size (Su et al., 2020) and similarity of land uses should also be considered to better refine urban land use classification.

Machine learning algorithms have been widely used in urban land use/land cover classification. The concept of ensemble learning has also enjoyed growing attention within the artificial intelligence and machine learning community (Zhang and Ma, 2012). The contribution of this study to method development and application could be concluded in two aspects. First, we proposed to apply the automatic ensemble learning framework in leveraging a group of machine learning algorithms for urban land use classification. As urban land use classification is a complex and challenging task with multi-source and high-dimensional features and different landscape settings, compared with single machine learning algorithm, the ensemble learning framework will be more useful to derive robust classification outcomes. Second, using the same set of training and validation block samples, our study provided a comparison of multi-model performance (i.e., random forest, extremely randomized trees, LightGBM boosted trees, CatBoost boosted trees, and neural networks) in urban land use classification, which will provide potential guidance for selecting models and strategies in urban land use classification practices in different locations. Experimental results revealed that multi-stacking ensemble models achieved relatively robust and better performance in classification accuracy. This classification strategy and framework is especially suitable for processing high dimensional features. Meanwhile, the tradeoff between classification accuracy and computational cost should be acknowledged. For each inclusive base model, we tested its classification accuracy and computational cost under different scenarios of no-stack, one-layer stack, and two-layer stack, respectively. We also leveraged all based models for multi-layer stacking (i.e., the weighted ensemble model) for baseline comparison. As for Level-I classification scheme, results showed that ensemble learning did achieve better performance than single models, with higher weighted F1 scores (Fig. S10). In the meantime, ensemble learning required more training time and memory size, which increased with the larger number of stack layers (Fig. S10). For scenarios with no-stack single modes, the average training time was 317.6 s with an average weighted F1 score of 87.74%. In contrast, as for the three-layer stack ensemble learning, the training time was approximately ten times longer (i.e., 3284.56 s on average), but the average weighted F1 score reached up to 90.60% (Table S6). Given the added computational cost due to ensemble learning, on the one hand, we felt it was still worthwhile considering the increase in classification accuracy. On the other hand, the balance between classification accuracy and computational cost can be adjusted according to different purposes of practical applications.

Although recent advances in deep learning that transform multi-spectral remote sensing imagery to high-level abstract features have proven great utilities in land cover and land use classification, the model interpretability of deep learning based approaches continues to be a major challenge. Moreover, the essential urban land use classification in this study is in the form of block-level mapping practices determined by road networks, which makes it difficult to create paired training samples in unified sizes and to link with multi-sensor and multi-format data sources. Additionally, deep learning based approaches are more appropriate to high or very high-resolution remote sensing observations for thematic information extraction, which may be hindered by computational costs at large-scale urban land use classification practices. In contrast, ensemble machine learning in this study that brings together multi-source features representing spectral, textural, height, topographic, and anthropogenic characteristics has demonstrated its robust and cost-effective capability of block-level urban land use classification with quite plausible accuracies. More importantly, the model interpretability of variable importance could help gain insights about the optimal selection of data sources and features in urban land use

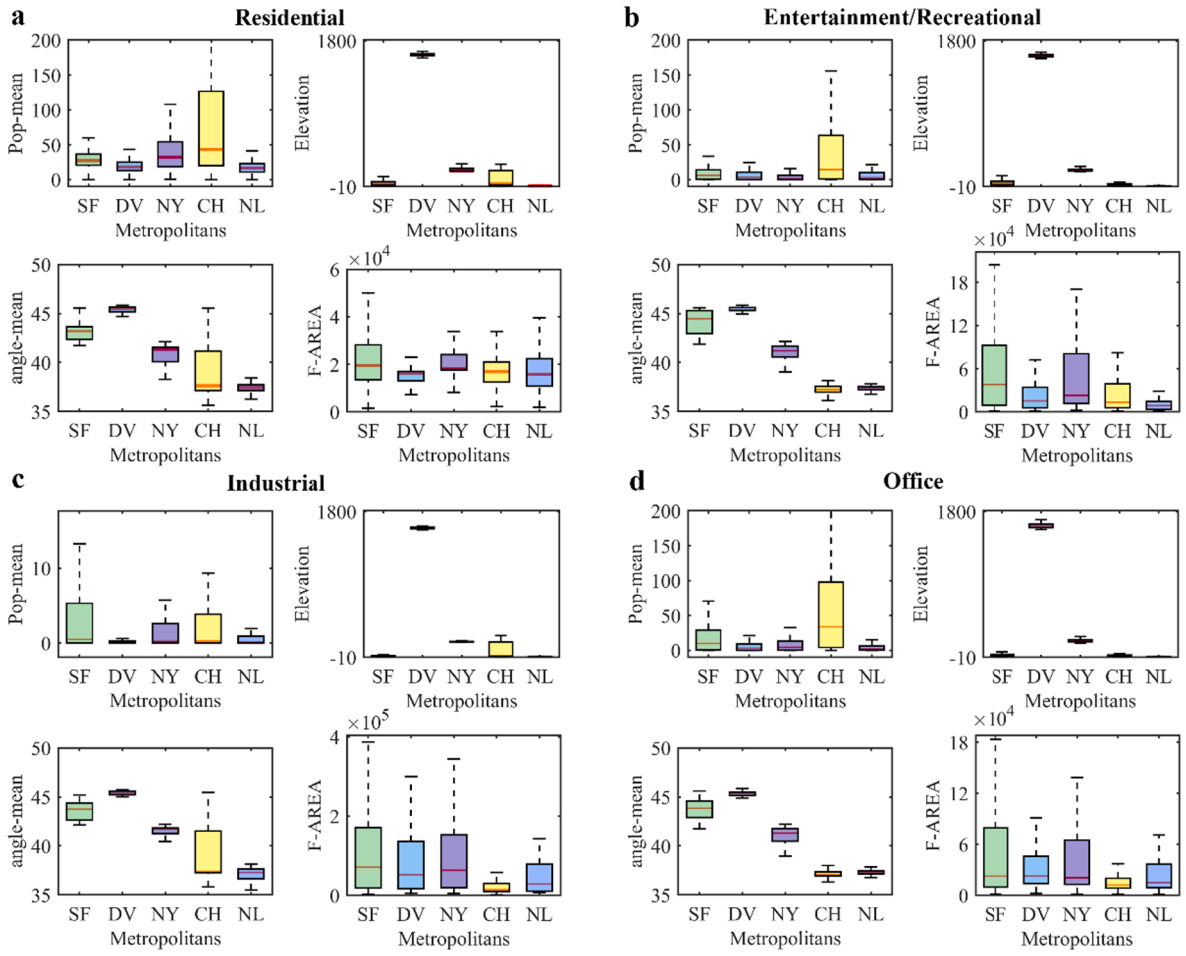

Fig. 9. Comparison of selected important features among five metropolitan areas in terms of different urban land use categories: (a) residential, (b) entertainment/recreational, (c) industrial, and (d) office lands. Noted that the San Francisco Bay area, Denver, New York City, Chicago, and New Orleans are abbreviated by SF, DV, NY, CH, and NL, respectively.

classification across different spatial and temporal scales. In addition to the model selection, the model transferability is another important issue. Given the difference in characteristics of urban land uses across states and countries, the training samples collected in some regions may not be appropriate for the land use classification in other regions. Our experimental results of five metropolitan areas also observed the varying overall accuracies when we applied the model trained using samples collected in one metropolitan to other metropolitan areas (Table 10). For large-scale mapping of urban land use categories, in particular, it will be unrealistic to apply a global model in universal land use classification for different regions. How to better define non-local similarities among different regions and develop spatially adjusted models will be an open topic in the direction of model and feature transferability from local to regional, and continental scales. For example, transfer learning which makes use of all parameters of neural network pre-trained over training datasets has been proven to be quite helpful and efficient to solve a different but related problem (Weiss et al., 2016).

This study presents a cost-effective framework for mapping urban land use categories using openly available multi-source geospatial "big data", using examples of five U.S. metropolitan areas. The potential of extending this framework to broader scale urban land use mapping can be summarized in terms of the following points. First, global open big data. Apart from the NAIP VHR imagery, the other datasets used including Sentinel-1, Sentinel-2, WorldPop, OpenStreetMap, nighttime light, Twitter, etc. are all globally free available, which makes it possible for regional to global urban land use classification practices. Second, cost-effective models. Automatic ensemble learning models have been verified to be effective and robust to achieve plausible accuracies in mapping urban land use categories. Third, crowdsourcing sample collections. Given the time-consuming cost of sample collections, we adopted the crowdsourcing scheme to leverage available geospatial big data to better facilitate sample collection from local to broader scales, which will be much more cost-effective than the on-site survey practices in EULUC-China (Gong et al., 2020a). Fourth, transferability of training and prediction across regions. Although localized models and samples will be more suitable for mapping urban land use categories at local scales, our experimental tests also demonstrate the potential possibility of transferring samples and predictions across different regions, which will be very useful to conduct broader-scale urban land use classification.

5. Conclusions

Leveraging multi-source geospatial big data, this study sought to present a robust and cost-effective framework for mapping urban land use categories, including five major procedures: (1) multi-source open big data collection; (2) generation of street blocks in metropolitan areas; (3) feature extraction from multi-source geospatial big data; (4) collection of training and validation samples; and (5) mapping urban land use categories using automatic ensemble learning strategy. Following this framework, we conducted block-level urban land use classification in five metropolitan areas of the United States, using a complete set of open-source geospatial data layers from VHR NAIP, Sentinel-1 GRD, Sentinel-2A/B, nighttime light, topography, population, and Twitter

data. Results showed that the overall accuracies of Level-I and Level-II classification among five metropolitan areas could be up to 91% and 86%, respectively. Multi-model comparisons revealed that compared with base machine learning models, the multi-stacking ensemble models achieved relatively robust and better performance in urban land use classification with high dimensional features. We found the classification result derived from Sentinel-1, Sentinel-2, and other open big data based features could also achieve comparable accuracies to models that included NAIP imagery, which supported the possibility of accurate block-level urban land use classification at large scales without VHR satellite imagery. We further found the model transferability was highly dependent on non-local heterogeneity in characteristics of different regions, which enlightened that cross-city model training and transferring should be cautious in practical applications. This study systematically elucidates the role of data sources, classification methods, and feature transferability in block-level land use classifications, and the methods and findings may carry implications for mapping multi-scale essential urban land use categories.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Major Program of the National Natural Science Foundation of China (20201321441 and 20201320003), The University of Hong Kong HKU-100 Scholars Fund, and donations made by the Cyrus Tang Foundation to Tsinghua University.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.isprsjprs.2021.06.010.

References

Barrington-Leigh, C., Millard-Ball, A., 2017. The world's user-generated road map is more than 80% complete. PLoS ONE 12, e0180698.
Brown de Colstoun, E.C., Huang, C., Wang, P., Tilton, J.C., Tan, B., Phillips, J., Niemczura, S., Ling, P.Y., Wolfe, R.E., 2017. Global Man-made Impervious Surface (GMIS) Dataset From Landsat. In. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC).
Chen, B., Huang, B., Xu, B., 2017a. Multi-source remotely sensed data fusion for improving land cover classification. ISPRS J. Photogramm. Remote Sens. 124, 27–39.
Chen, B., Nie, Z., Chen, Z., Xu, B., 2017b. Quantitative estimation of 21st-century urban greenspace changes in Chinese populous cities. Sci. Total Environ. 609, 956–965.
Chen, B., Song, Y., Huang, B., Xu, B., 2020. A novel method to extract urban human settlements by integrating remote sensing and mobile phone locations. Sci. Remote Sens. 100003.
Cheng, G., Zhou, P., Han, J., 2016. Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images. IEEE Trans. Geosci. Remote Sens. 54, 7405–7415.
CIESIN, 2018. Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11. NASA Socioeconomic Data and Applications Center (SEDAC). Palisades, NY.
Clinton, N., Stuhlmacher, M., Miles, A., Uludere Aragon, N., Wagner, M., Georgescu, M., Herwig, C., Gong, P., 2018. A global geospatial ecosystem services estimate of urban agriculture. Earth's Future 6, 40–60.
Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C., Worley, B.A., 2000. LandScan: a global population database for estimating populations at risk. Photogramm. Eng. Remote Sens. 66, 849–857.
Dorogush, A. V., Ershov, V., Gulin, A., 2018. CatBoost: gradient boosting with categorical features support. arXiv preprint arXiv:1810.11363.
Drusch, M., Del Bello, U., Carlier, S., Colin, O., Fernandez, V., Gascon, F., Hoersch, B., Isola, C., Laberinti, P., Martimort, P., 2012. Sentinel-2: ESA's optical high-resolution mission for GMES operational services. Remote Sens. Environ. 120, 25–36.
Du, S., Du, S., Liu, B., Zhang, X., 2021. Mapping large-scale and fine-grained urban functional zones from VHR images using a multi-scale semantic segmentation network and object based approach. Remote Sens. Environ. 261, 112480.
Elvidge, C.D., Baugh, K., Zhizhin, M., Hsu, F.C., Ghosh, T., 2017. VIIRS night-time lights. Int. J. Remote Sens. 38, 5860–5879.
Erickson, N., Mueller, J., Shirkov, A., Zhang, H., Larroy, P., Li, M., Smola, A., 2020. Autogluon-tabular: Robust and accurate automl for structured data. arXiv preprint arXiv:2003.06505.
Erol, H., Akdeniz, F., 2005. A per-field classification method based on mixture distribution models and an application to Landsat Thematic Mapper data. Int. J. Remote Sens. 26, 1229–1244.
Evans, G., 2010. National elevation dataset. US Geological Survey Earth Resources Observation and Science (EROS) Center. VITA Webinar Series.
Frantz, D., Schug, F., Okujeni, A., Navacchi, C., Wagner, W., van der Linden, S., Hostert, P., 2020. National-scale mapping of building height using Sentinel-1 and Sentinel-2 time series. Remote Sens. Environ. 252, 112128.
Frias-Martinez, V., Soto, V., Hohwald, H., Frias-Martinez, E., 2012. Characterizing urban landscapes using geolocated tweets. In: Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom). IEEE, pp. 239–248.
Gao, J., O'Neill, B.C., 2020. Mapping global urban land for the 21st century with data-driven simulations and Shared Socioeconomic Pathways. Nat. Commun. 11, 2302.
Gong, P., Chen, B., Li, X., Liu, H., Wang, J., Bai, Y., Chen, J., Chen, X., Fang, L., Feng, S., 2020a. Mapping essential urban land use categories in China (EULUC-China): preliminary results for 2018. Sci. Bull. 65, 182–187.
Gong, P., Howarth, P.J., 1990. The use of structural information for improving land-cover classification accuracies at the rural-urban fringe. Photogramm. Eng. Remote Sens. 56, 67–73.
Gong, P., Howarth, P.J., 1992. Frequency-based contextual classification and gray-level vector reduction for land-use identification. Photogramm. Eng. Remote Sens. 58, 423–437.
Gong, P., Li, X., Wang, J., Bai, Y., Chen, B., Hu, T., Liu, X., Xu, B., Yang, J., Zhang, W., Zhou, Y., 2020b. Annual Maps of Global Artificial Impervious Area (GAIA) between 1985 and 2018. Remote Sens. Environ. 236, 111510.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27.
Grimm, N.B., Faeth, S.H., Golubiewski, N.E., Redman, C.L., Wu, J., Bai, X., Briggs, J.M., 2008. Global change and the ecology of cities. Science 319, 756–760.
Guo, W., Yang, W., Zhang, H., Hua, G., 2018. Geospatial object detection in high resolution satellite images based on multi-scale convolutional neural network. Remote Sens. 10, 131.
Haklay, M., 2010. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. Environ. Plan. B: Plan. Des. 37, 682–703.
Hu, T., Yang, J., Li, X., Gong, P., 2016. Mapping urban land use by using landsat images and open social data. Remote Sens. 8, 151.
Huang, X., Yang, J., Li, J., Wen, D., 2021. Urban functional zone mapping by integrating high spatial resolution nighttime light and daytime multi-view imagery. ISPRS J. Photogramm. Remote Sens. 175, 403–415.
Huang, X., Hu, T., Li, J., Wang, Q., Benediktsson, J.A., 2018. Mapping urban areas in China using multisource data with a novel ensemble SVM method. IEEE Trans. Geosci. Remote Sens. 56, 4258–4273.
Kennedy, C.M., Oakleaf, J.R., Theobald, D.M., Baruch-Mordo, S., Kiesecker, J., 2019. Managing the middle: A shift in conservation priorities based on the global human modification gradient. Glob. Change Biol.
Koppel, K., Zalite, K., Voormansik, K., Jagdhuber, T., 2017. Sensitivity of Sentinel-1 backscatter to characteristics of buildings. Int. J. Remote Sens. 38, 6298–6318.
Lansley, G., Longley, P.A., 2016. The geography of Twitter topics in London. Comput. Environ. Urban Syst. 58, 85–96.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521, 436–444.
Li, X., Gong, P., Zhou, Y., Wang, J., Bai, Y., Chen, B., Hu, T., Xiao, Y., Xu, B., Yang, J., 2020a. Mapping global urban boundaries from the global artificial impervious area (GAIA) data. Environ. Res. Lett.
Li, X., Hu, T., Gong, P., Du, S., Chen, B., Li, X., Dai, Q., 2021. Mapping Essential Urban Land Use Categories in Beijing with a Fast Area of Interest (AOI)-Based Method. Remote Sens. 13 (3), 477.
Li, X., Zhou, Y., Gong, P., Seto, K.C., Clinton, N., 2020b. Developing a method to estimate building height from Sentinel-1 data. Remote Sens. Environ. 240, 111705.
Liu, S., Qi, Z., Li, X., Yeh, A.G.-O., 2019. Integration of convolutional neural networks and object-based post-classification refinement for land use and land cover mapping with optical and sar data. Remote Sens. 11, 690.
Liu, X., He, J., Yao, Y., Zhang, J., Liang, H., Wang, H., Hong, Y., 2017. Classifying urban land use by integrating remote sensing and social media data. Int. J. Geogr. Inform. Sci. 31, 1675–1696.
Liu, X., Huang, Y., Xu, X., Li, X., Li, X., Ciais, P., Lin, P., Gong, K., Ziegler, A.D., Chen, A., 2020. High-spatiotemporal-resolution mapping of global urban change from 1985 to 2015. Nat. Sustain. 1–7.
Liu, X., Long, Y., 2016. Automated identification and characterization of parcels with OpenStreetMap and points of interest. Environ. Plan. B: Plan. Des. 43, 341–360.
Long, Y., Gong, Y., Xiao, Z., Liu, Q., 2017. Accurate object localization in remote sensing images based on convolutional neural networks. IEEE Trans. Geosci. Remote Sens. 55, 2486–2498.
Lu, D., Weng, Q., 2006. Use of impervious surface in urban land-use classification. Remote Sens. Environ. 102, 146–160.
Machado, M.R., Karray, S., de Sousa, I.T., 2019. LightGBM: An effective decision tree gradient boosting method to predict customer loyalty in the finance industry. In: 14th International Conference on Computer Science & Education (ICCSE), pp. 1111–1116.

Mills, S., Weiss, S., Liang, C., 2013. VIIRS day/night band (DNB) stray light characterization and correction. In: Earth Observing Systems XVIII, vol. 8866. International Society for Optics and Photonics, p. 88661.
Meijer, J.R., Huijbregts, M.A., Schotten, K.C., Schipper, A.M., 2018. Global patterns of current and future road infrastructure. Environ. Res. Lett. 13, 064006.
Myint, S.W., Gober, P., Brazel, A., Grossman-Clarke, S., Weng, Q., 2011. Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery. Remote Sens. Environ. 115, 1145–1161.
Pesaresi, M., Huadong, G., Blaes, X., Ehrlich, D., Ferri, S., Gueguen, L., Halkia, M., Kauffmann, M., Kemper, T., Lu, L., 2013. A global human settlement layer from optical HR/VHR RS data: concept and first results. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 6, 2102–2131.
Petropoulos, G.P., Kalaitzidis, C., Vadrevu, K.P., 2012. Support vector machines and object-based classification for obtaining land-use/cover cartography from Hyperion hyperspectral imagery. Comput. Geosci. 41, 99–107.
Sarzynski, A., Galster, G., Stack, L., 2014. Evolving United States metropolitan land use patterns. Urban Geogr. 35, 25–47.
Schmidhuber, J., 2015. Deep learning in neural networks: An overview. Neural Netw. 61, 85–117.
Schneider, A., Friedl, M.A., Potere, D., 2009. A new map of global urban extent from MODIS satellite data. Environ. Res. Lett. 4, 044003.
Seto, K.C., Shepherd, J.M., 2009. Global urban land-use trends and climate impacts. Curr. Opin. Environ. Sustain. 1, 89–95.
Shi, K., Huang, C., Yu, B., Yin, B., Huang, Y., Wu, J., 2014. Evaluation of NPP-VIIRS night-time light composite data for extracting built-up urban areas. Remote Sens. Lett. 5, 358–366.
Stevens, F.R., Gaughan, A.E., Linard, C., Tatem, A.J., 2015. Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. PLoS ONE 10.
Su, M., Guo, R., Chen, B., Hong, W., Wang, J., Feng, Y., Xu, B., 2020. Sampling Strategy for Detailed Urban Land Use Classification: A Systematic Analysis in Shenzhen. Remote Sens. 12, 1497.
Tatem, A.J., 2017. WorldPop, open data for spatial demography. Sci. Data 4, 1–4.
Theobald, D.M., 2014. Development and Applications of a Comprehensive Land Use Classification and Map for the US. PLoS ONE 9, e94628.
Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., Potin, P., Rommen, B., Floury, N., Brown, M., 2012. GMES Sentinel-1 mission. Remote Sens. Environ. 120, 9–24.
Tu, Y., Chen, B., Zhang, T., Xu, B., 2020. Regional Mapping of Essential Urban Land Use Categories in China: A Segmentation-Based Approach. Remote Sens. 12, 1058.
UNDESA, 2014. World urbanization prospects, the 2011 revision. Population Division, Department of Economic and Social Affairs, United Nations Secretariat.
USDA-FarmServiceAgency 2020. The National Agriculture Imagery Program (NAIP). In.
Wang, P., Huang, C., Brown de Colstoun, E.C., Tilton, J.C., Tan, B., 2017. Global Human Built-up And Settlement Extent (HBASE) Dataset From Landsat. In: Palisades. NASA Socioeconomic Data and Applications Center (SEDAC), NY.
Watts, N., Adger, W.N., Agnolucci, P., Blackstock, J., Byass, P., Cai, W., Chaytor, S., Colbourn, T., Collins, M., Cooper, A., 2015. Health and climate change: policy responses to protect public health. Lancet 386, 1861–1914.
Watts, R.D., Compton, R.W., McCammon, J.H., Rich, C.L., Wright, S.M., Owens, T., Ouren, D.S., 2007. Roadless space of the conterminous United States. Science 316, 736–738.
Weiss, K., Khoshgoftaar, T.M., Wang, D., 2016. A survey of transfer learning. J. Big Data 3 (1), 1–40.
Wheater, H., Evans, E., 2009. Land use, water management and future flood risk. Land Use Policy 26, S251–S264.
Yao, Y., Li, X., Liu, X., Liu, P., Liang, Z., Zhang, J., Mai, K., 2017. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec model. Int. J. Geogr. Inform. Sci. 31, 825–848.
Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P.M., 2018. An object-based convolutional neural network (OCNN) for urban land use classification. Remote Sens. Environ. 216, 57–70.
Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P.M., 2019. Joint Deep Learning for land cover and land use classification. Remote Sens. Environ. 221, 173–187.
Zhang, X., Du, S., Wang, Q., 2017. Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data. ISPRS J. Photogramm. Remote Sens. 132, 170–184.
Zhang, C., Ma, Y., 2012. Ensemble machine learning: methods and applications. Springer Science & Business Media.
Zhong, Y., Su, Y., Wu, S., Zheng, Z., Zhao, J., Ma, A., Zhu, Q., Ye, R., Li, X., Pellikka, P., Zhang, L., 2020. Open-source data-driven urban land-use mapping integrating point-line-polygon semantic objects: A case study of Chinese cities. Remote Sens. Environ. 247, 111838.
Zhou, Y., Li, X., Asrar, G.R., Smith, S.J., Imhoff, M., 2018. A global record of annual urban dynamics (1992–2013) from nighttime lights. Remote Sens. Environ. 219, 206–220.