 doi:10.20944/preprints202207.0071.v1

Article

Development of Global Impervious Surface Area Index for Automatic spatiotemporal Urban Mapping

Akib Javed [1,*], Zhenfeng Shao[1], Iffat Ara[2,3], Md. Enamul Huq[4], Md. Yeamin Ali[5], Nayyer Saleem[1] and Muhammad Nasar Ahmad[1]

[1] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei 43007, China; shaozhenfeng@whu.edu.cn
[2] Environmental Science and Geography, Islamic University, Kushtia 2003, Bangladesh; iffat@esg.iu.ac.bd
[3] Department of International Politics, Shandong University, Qingdao, Shandong 266237, China;
[4] College of Environmental Science, Hohai University, Nanjing, Jiangsu 210098, China; enamul.huq@hhu.edu.cn
[5] Caritas Bangladesh, Bandarban 4600, Bangladesh; yeaminiesru@gmail.com

* Correspondence: akibjaved@whu.edu.cn; Tel.: +880 172 111 2202

Abstract: Impervious surface area (ISA) is a crucial indicator for quantitative urban studies. It is also important for land use land cover classification, groundwater recharge, sustainable development, urban heat island effects, and more. Spectral ISA mapping suffers from mixed pixel problems, especially with bare soil. This study aims to develop an ISA index for spatiotemporal urban mapping from common multispectral bands by reducing soil signature better than in previous studies. This study proposed a global impervious surface area index (GISAI) enhancing ISA mapping accuracy using a temporal parameter of the remote sensing (RS) dataset. Bare soil spectral reflectance shows more fluctuation than urban ISA. Therefore, the study uses minimum composites of earlier urban indices to compile minimum soil signature. It is later improved by removing water, increasing the contrast between bare soil and urban ISA and reducing bright bare soil area. This study maps the ISA of all 12 megacities using the annual RS image collection from 2021. GISAI reduced the bare soil signature and achieved an overall accuracy of 87.29%, F1-score of 0.84, and Kappa coefficient of 0.75. However, it has some limitations with grey bare soil and barren hilly areas. By limiting bare soil signature, GISAI broadens the scope of inter-urban studies globally and lengthens potential urban time-series analysis using common multispectral bands.

Keywords: Urban Mapping; Impervious Surface Area; Google Earth Engine; GISAI; Spectral Index; Landsat.

1. Introduction

Urban impervious surface area (ISA) indexing is important for urban mapping, comparative urban studies, land use land cover change (LULCC) detection studies, urban dynamics, urban growth, urban flooding, groundwater recharge, sponge cities, urban heat island effects and etc. Urban areas are difficult to define. In contrast, the impervious surface area is a good indicator for urban areas. Usually, an urban area is a large concentration of ISA.

Supervised ISA classification methods have better accuracy for small study areas, but large-scale studies are time-consuming and prone to human error. Furthermore, supervised methods require sampling points to train the individual classifier, which increases with the number of image classes, study area and time. Manual selection of sample points can cause researchers to be biased and differences in mapping results. Unsupervised methods are rapidly applicable with spectral indices.

Manual RS imagery (RSI) selection also plays an important role in accuracy in land use land cover classification (LULCC). Atmospheric elements, such as clouds, dust, haze,

 doi:10.20944/preprints202207.0071.v1

cirrus clouds, and cloud shadows, can reduce the reflectance of the ground surface. Seasonal phenomena such as, changes in vegetation, water bodies, snow covers etc., also change LULCC. Summer seasons have brighter ground reflectance but are prone to more cloud coverage. Winter season scenes are less cloudy but with poor ground reflectance. Urban deciduous trees have a seasonal impact on urban ISA measurement. Especially, tree covered urban roads are difficult to map from RSI. Many urban dynamics studies use annual image composites instead of a single RSI. Additionally, annual image composites can be time-saving options for studying cloud-prone regions.

There is a difference in the mapping of urban areas and ISA too. The urban area can have both pervious and impervious areas. On the contrary, the ISA image class is opposite the pervious soil area where previous urban indices are often mixed with bare soil signature. Bright soil tends to mix with bright ISA, and dark soil tends to mix with shadow area. Their spectral similarity made classification a major challenge in multispectral remote sensing (RS) urban mapping.

Earlier attempts of index-based urban indexing examples are Index-based built-up index (IBI) [1], improved normalized difference of built-up index (INDBI) [2], vegetation index built-up index (VIBI) [3], built-up area extraction method (BAEM) [4], ratio normalized difference soil index (RNDSI) [5], normalized urban areas composite index (NUACI) [6], combinational built-up index (CBI) [7], combinational biophysical composition index (CBCI) [8], enhanced normalized difference impervious surface index (ENDISI) [9], and nighttime lights adjusted impervious surface index (NAISI) [10] are some of the examples.

Zhou, *et al.* [11] developed a binary built-up index (BBI) by adding two normalized difference built-up indices (NDBI). The two NDBI's are $NDBI_{Blue\text{-}Green,b}$ and $NDBI_{Red\text{-}Green,b}$. Positive values of these indices changed to 1, otherwise 0. Therefore, BBI ranges from 0 to 2. The study suggests that values 1 and 2 refer to built-up areas. Bai, *et al.* [12] modified it by excluding water areas as pre-processing steps and proposed a water eraser-normalized difference built-up index (WE-NDBI).

Many multi-index-based urban indices follow the elimination approach for better urban mapping, where the land surface comprises vegetation, impervious surfaces, and soil. The concept is popularized by Ridd [13] using V-I-S model. Usually, water areas were masked out in pre-processing stage to perform V-I-S model-based classification of impervious surfaces. Such as the biophysical composition index (BCI) [14]. Some urban indices show sensitivity in water bodies too. Therefore, removing waterbodies from RSI improves classification [12].

On the other hand, Zhang, *et al.* [15] proposed Water-Impervious-Pervious (W-I-P) model, where feature spaces have been used to develop an urban composition index (UCI). In the W-I-P model, water, Impervious surface and Pervious surfaces can be derived from a single composition index named UCI.

Both normalized difference water index (NDWI) [16] and modified normalized difference water index (MNDWI) have been used in urban indices. NDWI is used to add an urban feature in CBI [7]. Also, MNDWI is used several times in subtraction for urban indices. IBI, ENDISI, Modified Normalized Difference Impervious Surface Index (MNDISI) [17], and BAEM are some examples where MNDWI has been used to subtract in respective indices. MNDWI has reflectance towards ISA, which is second to the water area. Also, there is a slight difference in reflectance in-between ISA and bare soil. The first time, a significant attempt to distinguish the urban area from bare soil was found in the BCI, where tasseled cap transformation (TCT) was used. The separability among the two confusing classes was mild, however. Later, Deng, Wu, Li and Chen [5] proposed ratio normalized difference soil index (RNDSI) using the first band of TCT and normalized difference soil index (NDSI), which was also developed by Deng, Wu, Li and Chen [5]. Without TCT, Piyoosh and Ghosh [18] use Panchromatic (PAN) bands for soil-free urban mapping and proposed normalized ratio urban index (NRUI). Interestingly, this study also developed a soil index modified normalized difference soil index (MNDSI) like NDSI Deng, Wu, Li and Chen [5]. Dealing with bare soil's spectral similarity, urban researchers try to develop ISA-free soil indices to enhance ISA indexing.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 July 2022 doi:10.20944/preprints202207.0071.v1

This study uses multiple spectral indices to propose the global impervious surface area index (GISAI), which only uses common multispectral bands. The study also introduced a novel modified built-up and bare index (MBBI), which used SWIR1 and SWIR2 bands, and it is similar to the normalized difference built-up index (NDBI) [19] and urban index (UI) [20]. The study aims to develop a spectral index-based method to enhance ISA from multispectral RS datasets and extract ISA for urban studies with better separability from bare soil areas, spatial application with diverse geographical backgrounds, and longer temporal application. So that GISAI can compare cities with diverse climatic and environmental scenarios and be usable in time-series analysis with RS archives from the past decades.

2. Materials and Methods

Spatially and temporally diverse ISA mapping requires RS datasets with global coverage and archives with a long history. Landsat is most suitable to serve this purpose. Due to the temporal dimension of the proposed method, this study relies only on the most common spectral bands found in all the Landsat missions & also other multispectral sensors. This study uses only six primary visible and infrared bands, such as blue, green, red, near-infrared, short-wave infrared 1, and short-wave infrared 2 from Landsat naming of spectral bands.

2.1. Study area

The study chooses the top 12 megacities worldwide to run with this experiment. These 12 megacities are Tokyo & Osaka of Japan, Delhi & Mumbai of India, Shanghai & Beijing of China, Sao Paolo of Brazil, Mexico City of Mexico, Dhaka of Bangladesh, Cairo of Egypt, New York of USA, and Karachi of Pakistan.

2.2. Google Earth Engine

GEE is a cloud platform for remote sensing research developed and maintained by Google LLC [21]. It has a petabyte-scale of RS imagery (RSI) from numerous free RS data sources. The study used Landsat 8 OLI surface reflectance dataset for this study. The study script is added in Appendix B. This study required high computation, which is almost impossible without a cloud platform like GEE. It used an annual image collection of 10-100 images per city and generated composite images using various index statistics (minimum composites, median composites, etc.). The methodology is scripted on GEE from raw data to exporting the final index. The full script link of the study is given in Appendix C. This ensures higher reproducibility and flexibility spatiotemporally. Also, a temporal study is possible with historical Landsat datasets.

Furthermore, similar multispectral sensors can increase spatial accuracy. Ultimately, GISAI indexed images are exported from GEE for accuracy assessment and presentation purposes. It is worth mentioning that Landsat 7 ETM+ has scan line errors which are significantly minimized if used the annual image composites approach.

2.3. Study design

2.3.1. Pre-processing

This study uses annual image collection (AIC) from the USGS Landsat 8 Level 2, Collection 2, Tier 1 surface reflectance dataset using a date range from January 1st to December 31st instead of a single RSI. For atmospheric correction, Landsat provided quality band Bitmask has been used. Additionally, the study normalized the reflectance value to a 0 to 1 range using Landsat image properties.

2.3.2. Water masking

First, we separate the water bodies using two popular water indices. Xu [22] developed a modified normalized difference water index (MNDWI), and Yue and Liu [23] developed a deeply clear waterbody delineation index (DCWDI).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 July 2022 doi:10.20944/preprints202207.0071.v1

$$MNDWI = \frac{Green - SWIR1}{Green + SWIR1} \quad (1)$$

$$DCWDI = \sqrt{(Red^2 + NIR^2)} \quad (2)$$

Using both indices, we develop composite free water area (CWFA). To mitigate seasonal variation of water bodies median composite of both of the indices has been used. Then, thresholds 0 and 0.05 were used for medianMNDWI and medianDCWDI to extract water bodies.

$$CWFA = (medianMNDWI_T + medianDCWDI_T) < 1 \quad (3)$$

Here, subset 'T' refereeing to the threshold images. CWFA is a binary image of 1 and 0, where 0 is for water and 1 is for non-water areas.

2.3.3. Development of Composite ISA Index (CISAI)

The study starts from basic urban indices NDBI [19], UI [20], and novel Modified Bare and Built-up Index (MBBI) are, all using three infrared bands. This study used a simpler version of UI, that is NDUI.

$$NDBI = \frac{SWIR1 - NIR}{SWIR1 + NIR} \quad (4)$$

$$NDUI = \frac{SWIR2 - NIR}{SWIR2 + NIR} \quad (5)$$

$$MBBI = \frac{SWIR2 - SWIR1}{SWIR2 + SWIR1} \quad (6)$$

These indices are good in enhancing ISA pixels. The only problem is that they do not discriminate against bare soil areas. However, using temporal information, bare soil signature can be reduced. Bare soil reflectance is not as static as ISA in temporal variation. Therefore, soil signature will be minimized if a minimum composition of all three indices from AIC is taken.

The study added 1 to each of the three minimum composite urban indices and multiplied them to develop a composite ISA index (CISAI).

$$CISAI = (minNDBI + 1) * (minNDUI + 1) * (minMBBI + 1) * CWFA \quad (7)$$

Here, minNDBI, minNDUI and minMBBI refer minimum composites of each index. CISAI is the water-free composites of all the spectral differences of ISA from pervious surface areas.

CISAI (See Appendix A Figure A1) shows the combination of traditional urban indices, such as NDBI, UI, and MBBI. It works well for demarketing some cities but fails to map all the cities, especially cities surrounded by low vegetation.

2.3.4. Development of terraMNDWI

MNDWI also enhances ISA after waterbodies as a second sensitive LULC class. More importantly, MNDWI slightly differentiates ISA from bare soil areas, especially if it is a bright bare soil area. To get that, a minimum composition of MNDWI (minMNDWI) and added 1 with it. Then multiply with CWFA to get terraMNDWI.

$$terraMNDWI = (minMNDWI + 1) * CWFA \quad (8)$$

The function of terraMNDWI is to increase the separability of bare soil and ISA.

2.3.5. Development of short-wave infrared soil index (swirSoil)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 July 2022 doi:10.20944/preprints202207.0071.v1

SWIR1 and SWIR2 bands show very high reflectance for some bare soil areas, particularly desert sandy areas. The study proposes a simple yet effective soil index swirSoil to map the bright sandy area.

$$swirSoil = SWIR1 * SWIR2 * 4 \quad (9)$$

The purpose of swirSoil is to remove bright soil areas from RSI, such as desert and beach areas.

2.3.6. Development of the global ISA index (GISAI)

The study proposed that GISAI is taken advantage of CISAI, terraMNDWI and swirSoil indices.

$$GISAI = (CISAI * terraMNDWI) - swirSoil \quad (10)$$

GISAI dealt with the similarity of ISA and bare soil in two ways.

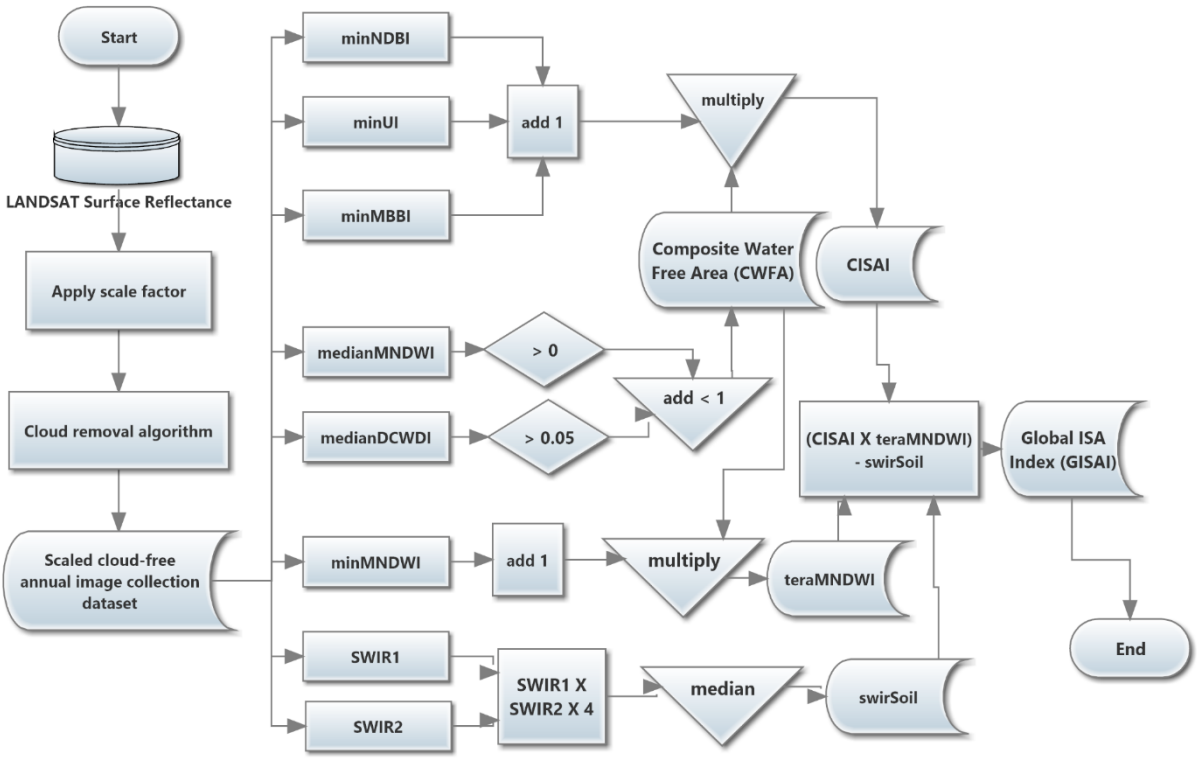

Figure 1. Complete methodology of GISAI.

2.4. Quality Assessment

2.4.1. Random sampling points

The study chooses 600 random sampling points per city to assess the accuracy of the GISAI. Sampling points were derived using ArcToolbox (ArcToolbox\Data Management Tools\Sampling\Create Random Points). The sample area was chosen manually, considering both urban and peri-urban areas. The study year is selected 2021 and for sampling locations and distribution, see appendix A Figure A3.

2.4.2. Ground truth reference points

Reference points were checked afterwards using higher spatial resolution latest Google Earth image. The reference data were collected in the last week of May 2022. For

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 July 2022 doi:10.20944/preprints202207.0071.v1

all 12 cities, Google Earth images were updated within 2021 and had 1m – 10m of ground spatial resolution.

2.4.3. Classified points

GISAI images are exported from GEE and reclassified as binary images using a threshold value of 0.1. 'Extract Values to Points' is a tool in ArcToolbox used to derive classified points using random sample points and binary images.

2.4.4. Accuracy assessment metrics

This study used only two classes, ISA and other areas. Therefore, the confusion matrix shows whether the classified points are ISA or not.

Table 1. Binary Confusion matrix

	Classified: No	**Classified: Yes**	
Reference: No	True Negative (TN)	False Positive (FP)	Total Reference No
Reference: Yes	False Negative (FN)	True Positive (TP)	Total Reference Yes
	Total Classified No	Total Classified Yes	

The binary confusion matrix is useful for dealing with only two classes; ISA and other classes. From the confusion matrix, the study derived all the accuracy assessment metrics.

Accuracy assessment metrics used in this study are producer accuracy, user accuracy and overall accuracy [24], F1 score, which is the balanced index for both producer accuracy and user accuracy [25,26], Cohen's Kappa coefficient, used as the classification has only two class, [27]. All the metrics formulas are described below. Where, OA= Overall Accuracy, TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative, PA = Producer Accuracy and UA = User Accuracy.

3. Results

$$Overall\ Accuracy\ (OA) = (TP + TN)/(TN + FP + FN + TP) \quad (11)$$

$$Error\ rate = 1 - OA \quad (12)$$

$$Omission\ error = FN/(FN + TP) \quad (13)$$

$$Comission\ error = FP/(FP + TP) \quad (14)$$

$$Producer\ Accuracy\ (PA) = TP/(TP + FN) \quad (15)$$

$$User\ Accuracy\ (UA) = TP/(TP + FP) \quad (16)$$

$$Kappa\ coeffieient = \frac{2(TP * TN - FP * FN)}{(TP + FP) * (FP + TN) + (TP + FN) * (FN + TN)} \quad (17)$$

$$F1\ score = (2 * PA * UA)/(PA + UA) \quad (18)$$

3.1. Enhancement of ISA with GISAI

The result shows significant improvement from traditional single index-based mapping. Significantly reduced soil signature. Some desert cities are demarcated with the index, such as Cairo of Egypt. Urban centers are more sensitive than urban fringes, such as Beijing. Hilly urban areas are also mapped differently from bare hills. Such as Karachi.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 July 2022 doi:10.20944/preprints202207.0071.v1

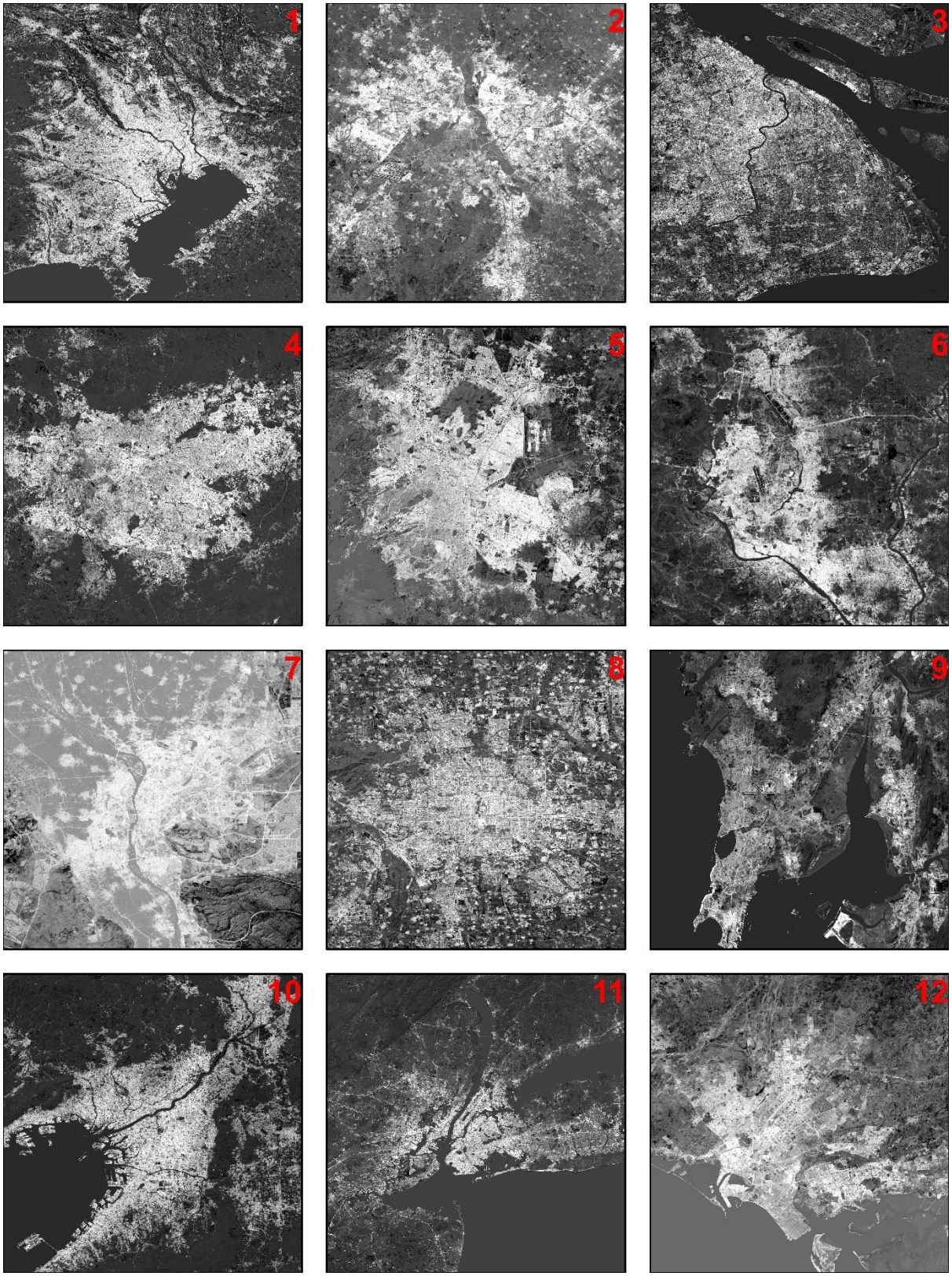

Figure 2. GISAI of all 12 megacities. 1. Tokyo, 2. Delhi, 3. Shanghai, 4. Sao Paolo, 5. Mexico City, 6. Dhaka, 7. Cairo, 8. Beijing, 9. Mumbai, 10. Osaka, 11. New York, and 12. Karachi

To understand the difference between infrared bands dependent composite urban indexed ISA enhancement (i.e. CISAI) and GISAI, compare Figure 2 with Figure A1 of Appendix A. Traditional indices are unable to separate ISA from bare soil areas. Bare soil

 doi:10.20944/preprints202207.0071.v1

areas included bare hills, bare mountains, river islands, sand-filled wetlands, desert areas and many more.

3.2. Extract ISA with GISAI

This study uses a common threshold value of 0.1 for all 12 cities and the result is shown in Figure 3. The result shows that GISAI extracted ISA areas for cities with diverse geographical settings. Using the Landsat 8 OLI dataset enables this study to reach 30m ground spatial resolution.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 July 2022 doi:10.20944/preprints202207.0071.v1

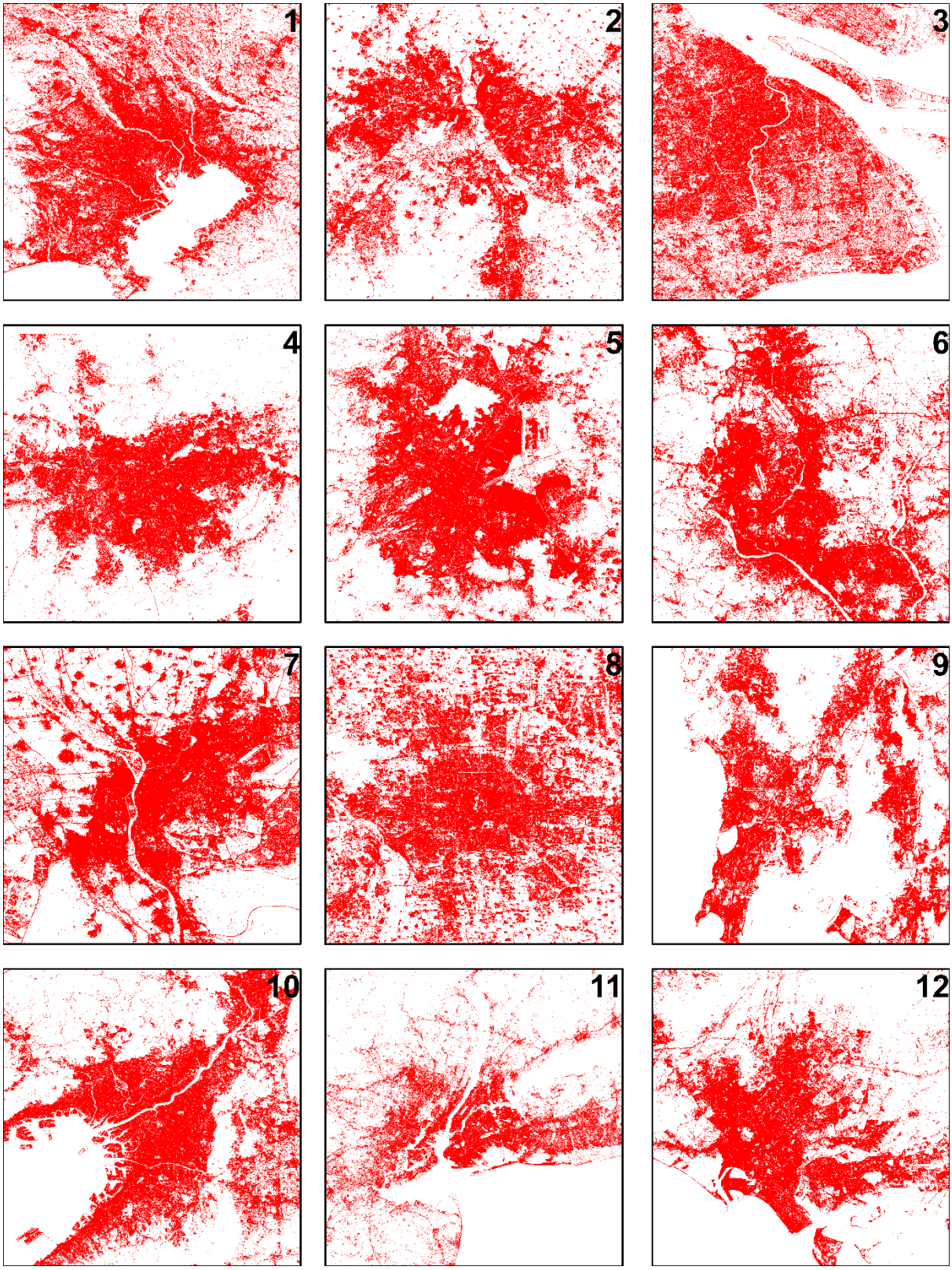

Figure 3. GISAI based urban mapping using threshold 0.1 for all 12 megacities. 1. Tokyo, 2. Delhi, 3. Shanghai, 4. Sao Paolo, 5. Mexico City, 6. Dhaka, 7. Cairo, 8. Beijing, 9. Mumbai, 10. Osaka, 11. New York, and 12. Karachi

GISAI shows good capabilities for mapping megacities and urban areas. The study tested various thresholds and found that threshold 0.1 works quite well. GISAI also has the potential to become a global urban mapping index, similar to NUACI. Most

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 July 2022 doi:10.20944/preprints202207.0071.v1

importantly, we use only one dataset, Landsat, to derive the result. Using additional datasets like NTL datasets can significantly improve the accuracy of the result. Furthermore, the multispectral bands are commonly found among previous Landsat missions and other multispectral earth observation satellites. Therefore, this index is also applicable to historical datasets easily.

The study also shows how effectively it can map various megacities with different environmental backgrounds and thus has the potential to compare cities in ISA measurement.

The spectral similarity of ISA and bare soil is dealt with in two ways. Primarily, the study took a minimum composite of all the urban indices from the annual image collection. That significantly reduces bare soil signature. Especially bare soil pixels that fluctuate in brightness or have any sort of vegetation. Secondly, the study uses a water-free version of the minimum MNDWI index composite to multiply with the existing ISA composite. Because, MNDWI is sensitive to ISA, which is less than water area. Furthermore, the water-free MNDWI index differentiates ISA from bare soil slightly. Finally, swirSoil removes the very bright sandy area from RSI.

The water-free area mask derived from the composite water index consists of MNDWI and DCWDI. The study also takes median water index composites from AIC for water area mapping. By doing so, only the most stable water areas are mapped, and seasonal variations are mitigated. Besides, most indices used in GISAI development are not sensitive to vegetation. Also, by using minimum and median composition, phenological effects are avoided.

Primarily Landsat provided Bitmask uses to remove cloud shadows. Although taking the minimum composition of urban indices has the potential to include building shadows.

Most importantly, this study follows an unsupervised classification approach, utilizes the cloud platform GEE, and uses freely available data. Therefore, this method is very robust and quickly reproducible by any researchers.

3.3. Quantitative Accuracy Assessment

The accuracy assessment in Table 2 shows user accuracy, producer accuracy, overall accuracy, omission error, commission error, error rate, precision, KAPPA coefficient and F1 score of all 12 megacities for GISAI. Dhaka city has the lowest overall accuracy, and Osaka city has the highest overall accuracy.

Table 2. Quantitative accuracy assessment of GISAI for top 12 megacities.

Megacity	Overall Accuracy	Error rate	Omission error	Commission error	Producer accuracy	User accuracy	Kappa	F1 score
Tokyo	0.900	0.100	0.056	0.163	0.944	0.837	0.798	0.887
Delhi	0.863	0.137	0.134	0.203	0.866	0.797	0.716	0.830
Shanghai	0.843	0.157	0.139	0.360	0.861	0.640	0.627	0.734
Sao Paolo	0.910	0.090	0.099	0.138	0.901	0.862	0.809	0.881
Mexico City	0.895	0.105	0.082	0.127	0.918	0.873	0.790	0.895
Dhaka	0.772	0.228	0.041	0.380	0.959	0.620	0.558	0.753
Cairo	0.875	0.125	0.105	0.142	0.895	0.858	0.750	0.876
Beijing	0.882	0.118	0.130	0.104	0.870	0.896	0.763	0.883
Mumbai	0.865	0.135	0.137	0.263	0.863	0.737	0.695	0.795
Osaka	0.938	0.062	0.077	0.054	0.923	0.946	0.876	0.934
New York	0.867	0.133	0.323	0.213	0.677	0.787	0.640	0.728
Karachi	0.865	0.135	0.130	0.220	0.870	0.780	0.714	0.823
Combined	0.873	0.127	0.113	0.193	0.887	0.807	0.738	0.845

4. Discussion

4.1. Significance of the study

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 July 2022 doi:10.20944/preprints202207.0071.v1

The study's most important contribution is the suppression of bare soil signature. In RS urban mapping classifying ISA and bare soil is always a major problem due to their spectral similarity. This study took various approaches to minimize the bare soil signature. So much so, even Cairo of Egypt, a desert surrounding city, is also possible to map with GISAI. Firstly, bright and highly reflective sandy areas were removed using SWIR1 and SWIR2 bands. These sandy areas reflect very high in both SWIR regions and more than any other class. Secondly, bare soil areas with vegetation or moisture are excluded using the temporal parameter.

To clarify, ISA and bare soil have similar spectral characteristics, but bare soil areas fluctuate more than ISA, especially if they have any seasonal vegetation presence. Therefore, if a minimum annual composite is generated of any urban indices (NDBI, UI, MBBI) or any soil indices, all those bare soil areas with any vegetation or moisture presence at any time of the year will be recorded with low value. In this way, any fluctuating reflectance of bare soil areas can be excluded or minimized. Thus, GISAI has better suited for urban comparison studies among cities.

Simultaneously, GISAI has used only five Landsat bands (Green, Red, NIR, SWIR1 and SWIR2) available in all the multispectral RS sensors. Especially, in Landsat TM, Landsat ETM+ and Landsat OLI missions. Therefore, GISAI can be used to map urban ISA since 1982 as the first launching of the Landsat 4 mission. It gave GISAI a huge temporal edge in urban studies over other urban indices. For example, Normalized Urban Areas Composite Index (NUACI) [6,28] used multispectral (MODIS and Landsat) and Defense Meteorological Satellite Program-Operational Linescan System (DMSP-OLS) NTL datasets. Both MODIS and NTL dataset does not have historical archives like Landsat missions. GISAI, on the other hand, only uses a single multispectral dataset with five bands to generate index values. GISAI can be enhanced as an index with an additional layer of NTL datasets but is not dependent on any NTL or MODIS datasets.

The study is independent of any sample regions, unlike NUACI [28]. Therefore, GISAI can be used for any city, big or small, without calculating sampling areas. Bypassing the need for sampling was a big motivator in this study.

At the same time, GISAI is better suited to comparison studies among cities across geographical backgrounds due to bare soil signature suppression. Usually, methodologically different geographical backgrounds can cause bias in urban mapping. For example, water bodies and vegetated areas are easy to map and bare soil, snow and mountainous region are difficult to map. The urban area from different geographical settings might have a different combination of the abovementioned features.

4.2. Comparison with similar studies

Developed in 2015, the Normalized Urban Areas Composite Index (NUACI2015) mainly depends on the annual maximum vegetation composite index [6]. Usually, urban areas cover a small portion of the LULC and behave like an anomaly where area is dominated by vegetation. Therefore, only vegetation composite indices are enough to map those urban areas surrounded by vegetation cover. However, the NUACI method will fail where vegetation is not a dominant background. This study has proven effective in those urban scenarios with a non-vegetated background.

Later in 2018, three popular normalized indices were used in another variant of $NUACI_{2018}$ [28]. Such as NDWI, Normalized Difference Vegetation Index (NDVI) [29] and NDBI. Carrying out $NUACI_{2018}$-based urban classification required several steps. Most importantly, sampling urban areas from various geographical settings are required for calibrating $NUACI_{2018}$. GISAI, on the other hand, is completely free from any form of sampling other than accuracy measurement. This sampling independence speed-up urban ISA measurement using GISAI.

Both variants of NUACI used NTL datasets. Although GISAI did not incorporate any NTL datasets, it is easy to add NTL datasets as additional layers. Adding NTL based layer has the potential to map global urban areas using GISAI. This index is better suited for urban and peri-urban area mapping. Most importantly, GISAI is more authentic in ISA

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 July 2022 doi:10.20944/preprints202207.0071.v1

mapping by utilizing only the Landsat multispectral dataset. Therefore, any ISA related studies can utilize this index.

On the other hand, Urban Composite Index (UCI) is very simplistic [15]. It uses a single cloud-free image, unlike NUACI and GISAI. On the other hand, this study uses annual median composite RSI to generate UCI images for all 12 cities and check the accuracy assessment. The accuracy assessment shows quite a balance results for accuracy assessment.

Table 3. Quantitative accuracy assessment of UCI for top 12 megacities.

Megacity	Overall Accuracy	Error rate	Omission error	Commission error	Producer accuracy	User accuracy	Kappa	F1 score
Tokyo	0.835	0.165	0.330	0.031	0.670	0.969	0.663	0.792
Delhi	0.808	0.192	0.410	0.075	0.590	0.925	0.585	0.720
Shanghai	0.778	0.222	0.463	0.264	0.537	0.736	0.470	0.621
Sao Paolo	0.702	0.298	0.750	0.079	0.250	0.921	0.273	0.393
Mexico City	0.825	0.175	0.286	0.072	0.714	0.928	0.652	0.807
Dhaka	0.727	0.273	0.326	0.192	0.674	0.808	0.458	0.735
Cairo	0.687	0.313	0.515	0.162	0.485	0.838	0.381	0.615
Beijing	0.830	0.170	0.254	0.104	0.746	0.896	0.660	0.814
Mumbai	0.765	0.235	0.531	0.219	0.469	0.781	0.436	0.587
Osaka	0.905	0.095	0.170	0.042	0.830	0.958	0.807	0.890
New York	0.868	0.132	0.451	0.151	0.549	0.849	0.589	0.667
Karachi	0.803	0.197	0.274	0.229	0.726	0.771	0.587	0.748
Combined	0.794	0.206	0.384	0.133	0.616	0.867	0.565	0.721

As a simplistic urban composite index, UCI has better user accuracy and commission error values but lacks in all other aspects of accuracy metrics. Although Zhang, Tian and Liu [15] found overall accuracy is 0.931 and a Kappa coefficient of 0.89, this study uses random sampling points. It gets overall accuracy of 0.830 and a Kappa coefficient of 0.660 for Beijing city. This study uses an annual cloud-free median image composite instead of a single image. That might play a role in declining accuracy assessment. Overall, UCI is a good urban index with limitations in dealing with bare soil.

5. Conclusion

This study developed GISAI with two aims and three characteristics. The two aims are enhancing ISA from RSI and extracting ISA for urban-related studies. It also has three characteristics that distinguish GISAI from previous indices. They are better separability of ISA from bare soil, better application with diverse geographic scenarios and longer temporal application with historical RSI.

GISAI enhanced ISA by using composite infrared band indices; secondly, by removing water areas; thirdly, enhancing separability of ISA and bare soil by multiplying water-free MNDWI layer; and subtracting bright bare soil area by using SWIR-based novel soil index. This method requires two novel indices derived from SWIR1 and SWIR2 bands. The first one is MBBI which used the normalized difference between them. The second one is swirSoil; a fourth-time multiplied version of the SWIR bands. This study uses 0.1 as a single threshold value to extract urban areas from all 12 megacities. Accuracy assessment shows that most cities have more commission errors than omission errors. Reducing the threshold value may gain better results. The whole method of the study is focused on increasing the separability of the two confusing classes. GISAI has significantly increased the separability of bare soil and ISA. Cairo and Karachi cities are strong proof of that (see Figure 2). The improved bare soil separability from ISA has enabled this method to apply to diverse geographical backgrounds. GISAI can achieve better ground spatial resolution with similar multispectral sensors, such as Sentinel-2 and ASTER. Although this study did not demonstrate any temporal application, limited uses of common spectral bands ensure the theoretical possibility of long-time series application. Landsat mission has the longest

 doi:10.20944/preprints202207.0071.v1

multispectral RSI archive. Landsat 4 TM, Landsat 5 TM, Landsat 7 ETM+, Landsat 8 OLI, and Landsat 9 OLI all these sensors have the bands GISAI uses.

GISAI is showing better accuracy than UCI in terms of overall accuracy, Error rate, omission error, producer accuracy, Kappa coefficient and F1 score. It also better represents the annual account of ISA estimation for urban areas by using annual image collection, unlike the single image-based approach of UCI. GISAI cannot only map all 12 top megacities but also classify desert-covered cities Cairo of Egypt and Karachi of Pakistan. Also, GISAI has shown significantly better results than UCI in mapping Sao Paolo of Brazil.

Limitations

The primary limitation is that the GISAI estimation method is highly computationally intensive and, therefore, dependent on cloud platforms like GEE. Furthermore, it has been almost impossible to deal with desktop computers until now. Besides increasing the separability of bare soil from ISA, some bare soil types are difficult to distinguish from ISA. Most importantly, gray bare soil and hilly regions. However, most of the barren gray regions are away from the study areas. Using the NTL dataset can eliminate most gray barren regions away from urban centers. Some glacial areas are misclassified as ISA and an additional snow index can be used to remove them. The snow index also helps mapping cities prone to snowfall.

Author Contributions: Conceptualization, A. J.; data curation, A. J.; funding acquisition, Z. S.; methodology, A. J. and M. N. A.; formal analysis, A.J., Z. S., I. A. and M. E. H.; supervision, Z. S.; validation, A. J. and I. A.; visualization, A. J.; writing – original draft, A. J.; writing – review and editing, I. A., M. Y. A. and N. S.; All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grants 42090012, Guangxi science and technology program (GuiKe2021AB30019, in part by 03 special research and 5G project of Jiangxi Province in China (20212ABC03A09); Zhuhai industry university research cooperation project of China (ZH22017001210098PWC); Key R & D project of Sichuan Science and technology plan (2022YFN0031); Zhizhuo Research Fund on Spatial–Temporal Artificial Intelligence (Grant No. ZZJJ202202).

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A: Related image results

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 July 2022 doi:10.20944/preprints202207.0071.v1

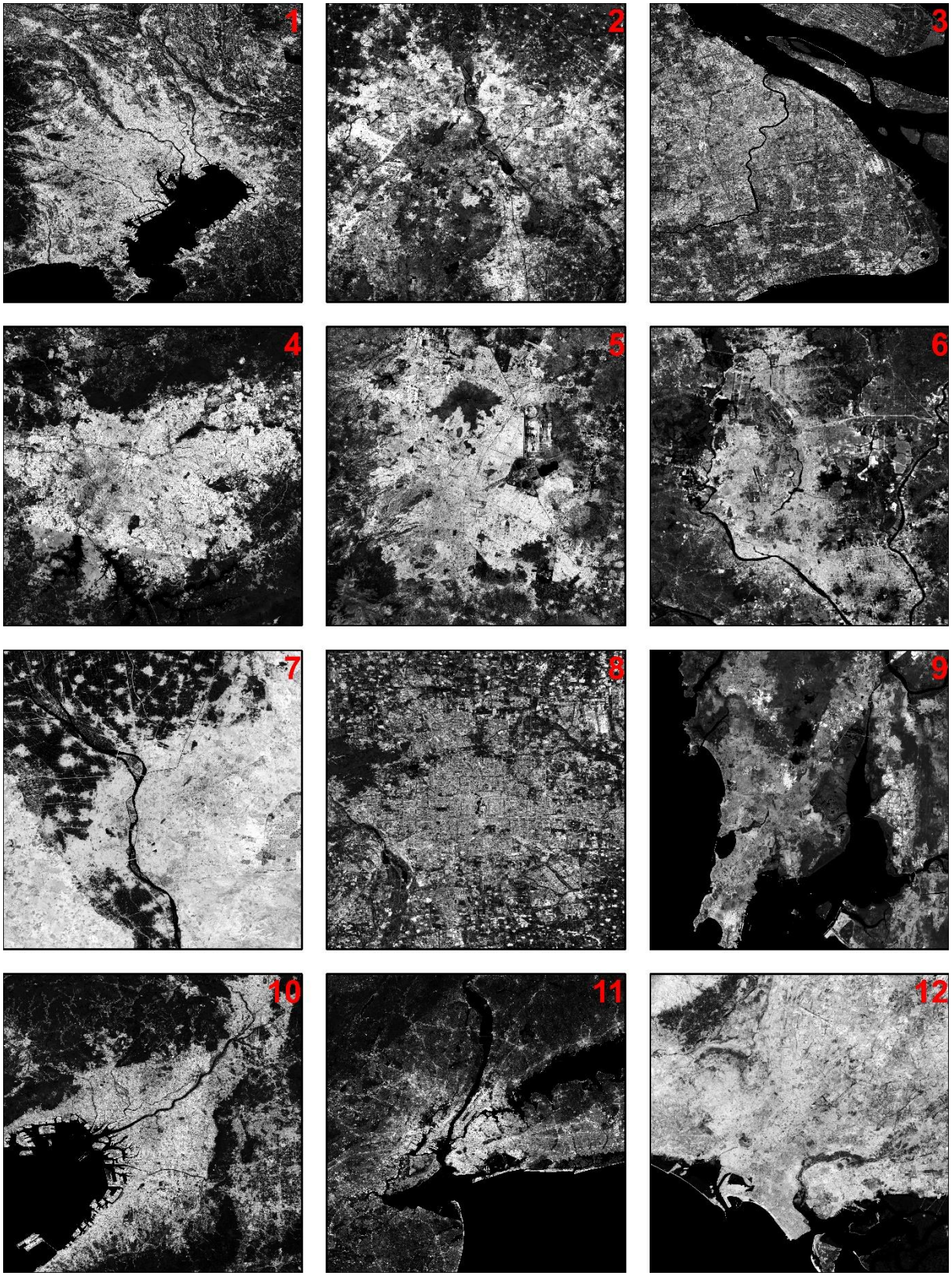

Figure A1. CISAI of all 12 megacities. 1. Tokyo, 2. Delhi, 3. Shanghai, 4. Sao Paolo, 5. Mexico City, 6. Dhaka, 7. Cairo, 8. Beijing, 9. Mumbai, 10. Osaka, 11. New York, and 12. Karachi

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 July 2022 doi:10.20944/preprints202207.0071.v1

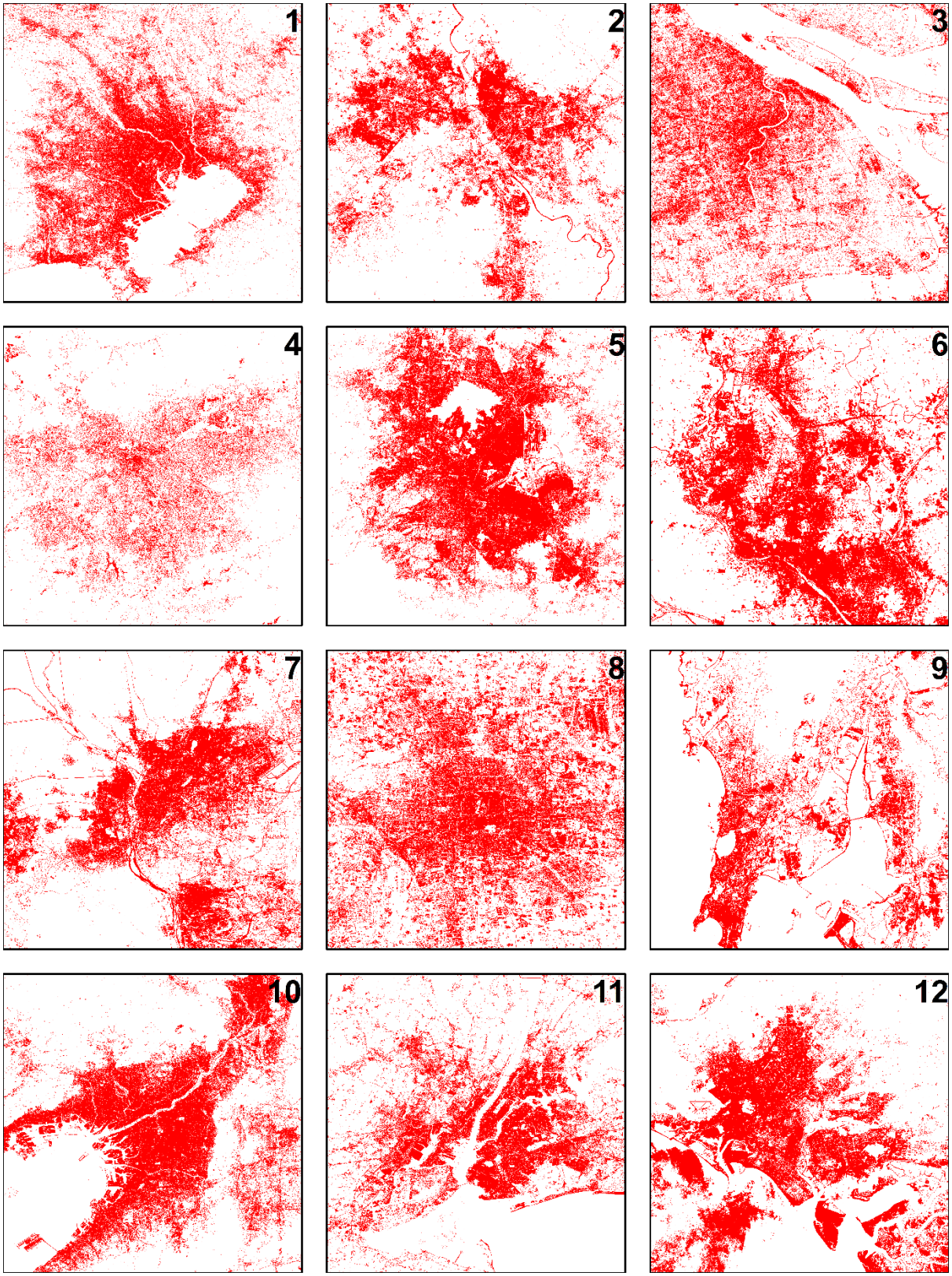

Figure A2. UCI based urban mapping using threshold (-0.414 – 0) for all 12 megacities. 1. Tokyo, 2. Delhi, 3. Shanghai, 4. Sao Paolo, 5. Mexico City, 6. Dhaka, 7. Cairo, 8. Beijing, 9. Mumbai, 10. Osaka, 11. New York, and 12. Karachi

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 5 July 2022 doi:10.20944/preprints202207.0071.v1

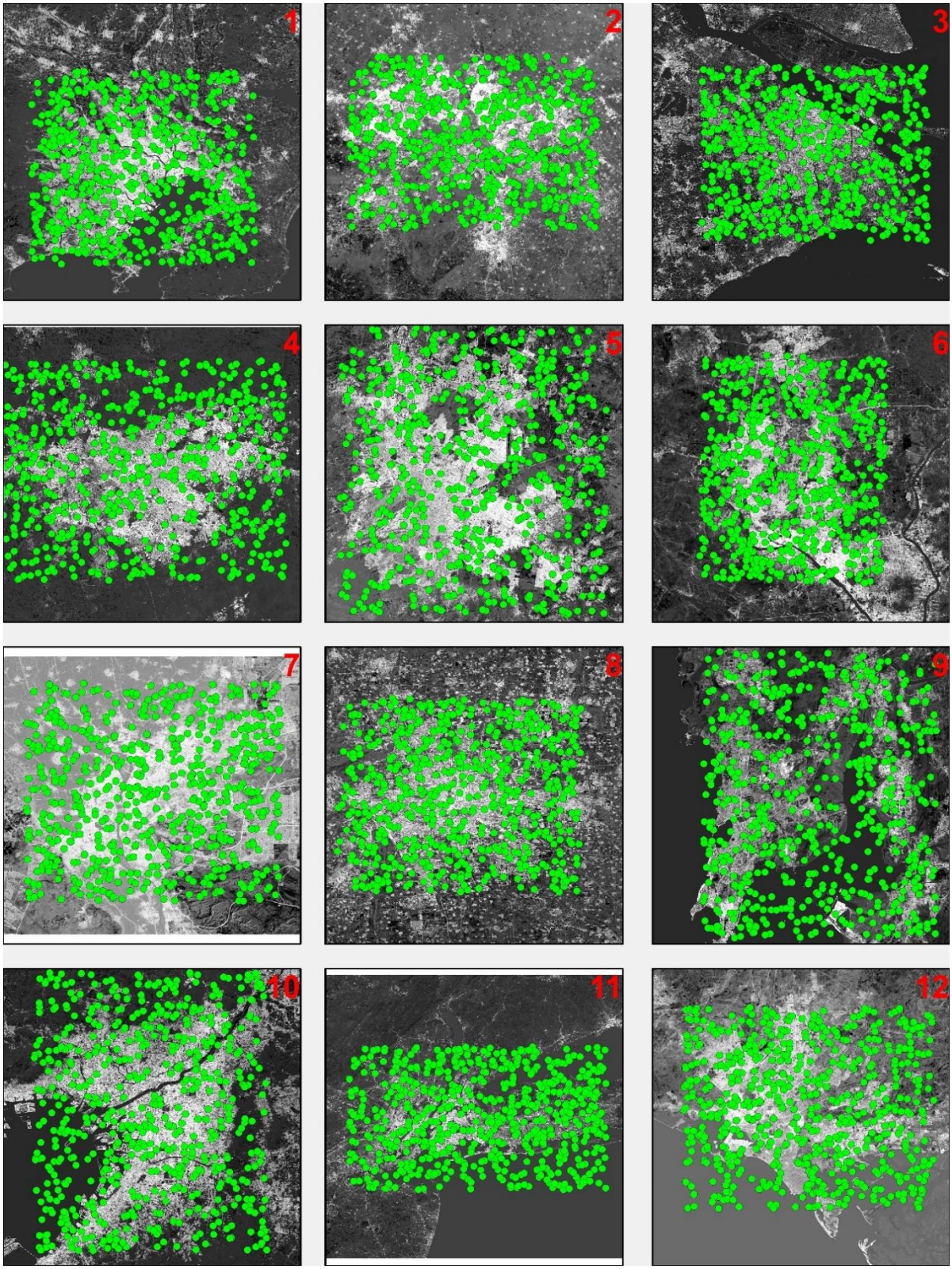

Figure A3. Sample points for all 12 megacities. 1. Tokyo, 2. Delhi, 3. Shanghai, 4. Sao Paolo, 5. Mexico City, 6. Dhaka, 7. Cairo, 8. Beijing, 9. Mumbai, 10. Osaka, 11. New York, and 12. Karachi

Appendix B: Landsat band names

Table A1. Landsat band names, numbers and used form in the text

 doi:10.20944/preprints202207.0071.v1

Band Name	Landsat 9 OLI	Landsat 8 OLI	Landsat 7 ETM+	Landsat 4 & 5 TM	Used in the text
Blue	2	2	1	1	Blue
Green	3	3	2	2	Green
Red	4	4	3	3	Red
Near Infrared	5	5	4	4	NIR
Short-wave Infrared 1	6	6	5	5	SWIR1
Short-wave Infrared 2	7	7	7	7	SWIR2

Appendix C: GEE script repository

Use the following link to access script used for the study. User required a Google Earth Engine account to access it.

https://code.earthengine.google.com/?accept_repo=users/AkibJaved/Manuscript

References

1. Xu, H. A new index for delineating built-up land features in satellite imagery. *Int. J. Remote Sens.* **2008**, *29*, 4269-4276, doi:10.1080/01431160802039957.
2. He, C.; Shi, P.; Xie, D.; Zhao, Y. Improving the normalized difference built-up index to map urban built-up areas using a semiautomatic segmentation approach. *Remote Sens. Lett.* **2010**, *1*, 213-221.
3. Stathakis, D.; Perakis, K.; Savin, I. Efficient segmentation of urban areas by the VIBI. *Int. J. Remote Sens.* **2012**, *33*, 6361-6377, doi:10.1080/01431161.2012.687842.
4. Bhatti, S.S.; Tripathi, N.K. Built-up area extraction using Landsat 8 OLI imagery. *GIScience remote sensing* **2014**, *51*, 445-467.
5. Deng, Y.; Wu, C.; Li, M.; Chen, R. RNDSI: A ratio normalized difference soil index for remote sensing of urban/suburban environments. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *39*, 40-48, doi:10.1016/j.jag.2015.02.010.
6. Liu, X.; Hu, G.; Ai, B.; Li, X.; Shi, Q. A Normalized Urban Areas Composite Index (NUACI) Based on Combination of DMSP-OLS and MODIS for Mapping Impervious Surface Area. *Remote Sens.* **2015**, *7*, 17168-17189, doi:10.3390/rs71215863.
7. Sun, G.; Chen, X.; Jia, X.; Yao, Y.; Wang, Z. Combinational Build-Up Index (CBI) for Effective Impervious Surface Mapping in Urban Areas. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2016**, *9*, 2081-2092, doi:10.1109/jstars.2015.2478914.
8. Zhang, S.; Yang, K.; Li, M.; Ma, Y.; Sun, M. Combinational biophysical composition index (CBCI) for effective mapping biophysical composition in urban areas. *IEEE Access* **2018**, *6*, 41224-41237.
9. Chen, J.; Yang, K.; Chen, S.; Yang, C.; Zhang, S.; He, L. Enhanced normalized difference index for impervious surface area estimation at the plateau basin scale. *J. Appl. Remote Sens.* **2019**, *13*, 19, doi:10.1117/1.Jrs.13.016502.
10. Chen, X.; Jia, X.; Pickering, M. A Nighttime Lights Adjusted Impervious Surface Index (NAISI) with Integration of Landsat Imagery and Nighttime Lights Data from International Space Station. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *83*, doi:10.1016/j.jag.2019.05.022.
11. Zhou, Y.; Yang, G.; Wang, S.; Wang, L.; Wang, F.; Liu, X. A new index for mapping built-up and bare land areas from Landsat-8 OLI data. *Remote Sens. Lett.* **2014**, *5*, 862-871.
12. Bai, Y.; He, G.; Wang, G.; Yang, G. WE-NDBI-A new index for mapping urban built-up areas from GF-1 WFV images. *Remote Sens. Lett.* **2020**, *11*, 407-415, doi:10.1080/2150704x.2020.1723171.
13. Ridd, M.K. Exploring a V-I-S (vegetation-impervious surface-soil) model for urban ecosystem analysis through remote sensing: comparative anatomy for cities† *Int. J. Remote Sens.* **1995**, *16*, 2165-2185, doi:10.1080/01431169508954549.

14. Deng, C.; Wu, C. BCI: A biophysical composition index for remote sensing of urban environments. *Remote Sens. Environ.* **2012**, *127*, 247-259, doi:10.1016/j.rse.2012.09.009.
15. Zhang, L.H.; Tian, Y.G.; Liu, Q.W. A Novel Urban Composition Index Based on Water-Impervious Surface-Pervious Surface (W-I-P) Model for Urban Compositions Mapping Using Landsat Imagery. *Remote Sens.* **2021**, *13*, 20, doi:10.3390/rs13010003.
16. McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **2007**, *17*, 1425-1432, doi:10.1080/01431169608948714.
17. Sun, Z.; Wang, C.; Guo, H.; Shang, R. A Modified Normalized Difference Impervious Surface Index (MNDISI) for Automatic Urban Mapping from Landsat Imagery. *Remote Sens.* **2017**, *9*, 18, doi:10.3390/rs9090942.
18. Piyoosh, A.K.; Ghosh, S.K. Development of a modified bare soil and urban index for Landsat 8 satellite data. *Geocarto Int.* **2017**, *33*, 423-442, doi:10.1080/10106049.2016.1273401.
19. Zha, Y.; Gao, J.; Ni, S. Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *Int. J. Remote Sens.* **2003**, *24*, 583-594, doi:10.1080/01431160304987.
20. Kawamura, M.; Jayamanna, S.; Tsujiko, Y. Quantitative evaluation of urbanization in developing countries using satellite data. *Doboku Gakkai Ronbunshu* **1997**, *1997*, 45-54.
21. Gorelick, N.; Hancher, M.; Dixon, M.; Ilyushchenko, S.; Thau, D.; Moore, R. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* **2017**, *202*, 18-27, doi:10.1016/j.rse.2017.06.031.
22. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2007**, *27*, 3025-3033, doi:10.1080/01431160600589179.
23. Yue, H.; Liu, Y. Method for delineating open water bodies based on the deeply clear waterbody delineation index. *J. Appl. Remote Sens.* **2019**, *13*, 16, doi:10.1117/1.Jrs.13.038504.
24. Story, M.; Congalton, R.G. Accuracy assessment: a user's perspective. *Photogramm. Eng. Remote Sens.* **1986**, *52*, 397-399.
25. Nasiri, V.; Deljouei, A.; Moradi, F.; Sadeghi, S.M.M.; Borz, S.A. Land Use and Land Cover Mapping Using Sentinel-2, Landsat-8 Satellite Images, and Google Earth Engine: A Comparison of Two Composition Methods. *Remote Sens.* **2022**, *14*, 1977.
26. Maxwell, A.E.; Warner, T.A. Thematic classification accuracy assessment with inherently uncertain boundaries: An argument for center-weighted accuracy assessment metrics. *Remote Sens.* **2020**, *12*, 1905.
27. Chicco, D.; Warrens, M.J.; Jurman, G. The Matthews correlation coefficient (MCC) is more informative than Cohen's Kappa and Brier score in binary classification assessment. *IEEE Access* **2021**, *9*, 78368-78381.
28. Liu, X.P.; Hu, G.H.; Chen, Y.M.; Li, X.; Xu, X.C.; Li, S.Y.; Pei, F.S.; Wang, S.J. High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform. *Remote Sens. Environ.* **2018**, *209*, 227-239, doi:10.1016/j.rse.2018.02.055.
29. Chen, J.M.; Cihlar, J. Retrieving leaf area index of boreal conifer forests using Landsat TM images. *Remote Sens. Environ.* **1996**, *55*, 153-162.