A Decision Rule and Machine Learning-Based Hybrid Approach for Automated Land-Cover Type Local Climate Zones (LCZs) Mapping Using Multi-Source Remote Sensing Data

Md Didarul Islam, Liping Di, *Senior Member, IEEE*, Chen Zhang, *Member, IEEE*, Ruixin Yang, John J. Qu, Daniel Tong, Liying Guo, *Member, IEEE*, Li Lin, *Member, IEEE*, and Aran Pandey

***Abstract*—This article presents a streamlined, automated classification method to map land-cover-type local climate zones (LCZs). Using a two-phase hybrid approach, we first generated training samples through universal decision rules and subsequently, a machine learning (ML) algorithm was trained on the generated samples to classify LCZs. The proposed model harnesses plant height data, combined with spectral bands and remote sensing indices, to accurately classify various land-cover types, such as dense forest, scattered trees, bush/scrub, low plant/agricultural land, bare rocks/paved surface and bare soil/sands. Targeting global applicability, we tested our method across six diverse locations spanning four continents: Fresno (California), Central Michigan, Western Phoenix (Arizona), Khulna (Bangladesh), Lagos (Nigeria), and Western Sydney. In each location, after generating training samples with the decision rules, a random forest algorithm was employed for LCZ classification. Results showcase that data from sentinels 1 and 2, night-time light, and global ecosystem dynamics investigation relative height are effective in characterizing land-cover-type LCZs and decision-rules can be established. The decision-rules consistently auto-generate training samples, undeterred by varying geographical and climatic conditions. This automated system has achieved promising accuracy across all tested sites, suggesting its potential to map land-cover-type LCZs and vegetation globally with higher accuracy.**

***Index Terms*—Local climate zones (LCZs), machine learning (ML), multisource remotely sensed data, vegetation mapping.**

I. Introduction

LOCAL climate zones (LCZs) is a scheme of urban systems classification that classifies urban areas into discrete LCZs by morphological and land cover characters [1].

Fig. 1. Built types (1–10) and land-cover types (a)–(g) LCZs defined by Demuzere et al. [3].

This scheme identifies a total of 17 LCZs [2], [3], [4], [5]. LCZs one through ten are designated for built-up lands, encompassing various types of urban development and infrastructure. Conversely, LCZs A through G predominantly correspond to land cover types. These zones, characterized by their microclimatic features, include nonbuilt-up land covers, such as urban parks, forests, bare soil, and agricultural lands, among others. While these land cover classes can be found in both urban and rural settings, they are primarily associated with areas not dominated by built-up infrastructure (see Fig. 1) [2]. Forests and agricultural lands are both essential components

Manuscript received 11 October 2023; revised 28 January 2024, 27 February 2024, and 3 April 2024; accepted 4 April 2024. Date of publication 8 April 2024; date of current version 19 April 2024. This work was supported by IDS under Grant 80NSSC20K1262. *(Corresponding author: Liping Di.)*

Md Didarul Islam, Liping Di, Chen Zhang, Ruixin Yang, Liying Guo, and Li Lin are with the Center for Spatial Information Science and Systems (CSISS) and the Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA 22030 USA (e-mail: ldi@gmu.edu).

John J. Qu is with the Global Environment and Resources Institute and the Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA 22030 USA.

Daniel Tong is with the Center for Spatial Information Science and Systems (CSISS) and the Department of Atmospheric, Oceanic & Earth Sciences, George Mason University, Fairfax, VA 22030 USA.

Aran Pandey is with the Skyline High School, Sammamish, WA 98075 USA.

Digital Object Identifier 10.1109/JSTARS.2024.3386389

of our ecosystem, providing a wide range of benefits to both the environment and human society. Monitoring forests and agricultural lands is of paramount importance due to the numerous ecological, social, and economic benefits these landscapes provide. Regular and systematic monitoring helps to understand changes, identify problems, and make informed decisions for sustainable management. This study emphasizes mapping land-cover types, LCZs A-G, as illustrated in Fig. 1. Such mapping directly supports several United Nations sustainable development goals (SDGs) outlines by United Nations [6], [7]. It contributes to improved urban living conditions and overall city sustainability by identifying areas prone to heat stress that helps planners and decision-makers to promote urban green-infrastructures and urban parks targeting the selected areas and mitigate urban heat islands effect [1], [2], [8], [9]. This aligns the land-cover-type LCZ mapping with SDG 11: Sustainable cities and communities. By analyzing changes of different LCZ types, such as LCZ A (i.e., dense forest), LCZ B (i.e., scattered trees), LCZ C (i.e., bush/scrub), LCZ D (i.e., low plant/agricultural land), and LCZ F (i.e., bare soil/sand) it is possible to determine forest loss, loss of agricultural land, desertification and land degradation. These information helps the policymakers to adopt policy towards achieving SDG 15: Protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss. They are also essential for maintaining healthy ecosystem, biodiversity, climate change mitigation, disaster prevention, food security, and sustainable economic growth [8], [10], [11]. Considering the paramount importance of land-cover-type LCZs mapping, this article seeks to overcome the current classification limitations and enhance classification accuracy, specifically targeting those selected LCZ categories and provide data at a finer 25-meter spatial resolution.

The current LCZs classification approach are primarily based pixel-based supervised classification method using multisource remote sensing (RS) data [3], [4], [12]. Several limitations of current classification approach can be observed if we attempt to extend the classification workflow from local to regional or global level. The major limitation is collecting training samples manually for supervise classification [3], [12]. Several LCZ types among selected categories have overlapping spectral characteristics. For example, dense forest (LCZ A), scattered tress (LCZ B), bush/scrub (LCZ C), and low/plant or agricultural land (LCZ D) have similar spectral characteristics, and similar values in normalize difference vegetation index (NDVI). This spectral similarities, observed among dense forests, scattered forests, agricultural lands, scrubs/bushes, and others, can lead to misclassification during mapping [13], [14]. The difficulties for mapping those land-types requires large training samples, and the sample requirements, especially for mapping over large regional/global scales demand training sample collections beyond the costly manual approaches. In addition, an LCZ classification approach devised for a particular area may be inapplicable to another area due to differences in physical-geographical characteristics and global weather patterns [15], [16]. For example, equatorial regions experience a hot and humid climate throughout the year due to their proximity to the equator, whereas areas in the northern hemisphere can experience a wide range of temperatures from very cold in the winter to very hot in the summer. This affects tree-structure, forest cover and vegetation that grow in these regions, hence, variation of spectral signatures can be observed [16]. Consequently, the LCZ classification approach adopted for a specific region cannot be applied to a different location or on a global scale, as it fails to address these variations. Moreover, several studies have used commercial satellite/Lidar derived plant height data to distinguish low/high plant types which are not cost-effective and efficient solution and may not be adopted at the region/global scale [17], [18]. Another important limitation is—previous studies overlooked spatial dependence property of geospatial data [19], [20]. In other words, previous studies overlooked Tobler's first law of geography: "everything is related to everything else, but near things are more related than distant things" [21]. LCZs are unique homogeneous local micro-climate zones distinguished by morphological and land cover characters. All physical features in each homogeneous LCZ should exhibit similar morphological and land cover characteristics influenced by spatial dependence. The spatial information in terms of spatial coordinate in the machine learning (ML) model will address this spatial dependence issue. The goal of this article is to overcome these limitations and propose a classification workflow that will utilize freely available RS data and provide cost-effective solution for land-cover-type LCZ mapping.

RS allows us to monitor global forest cover, agricultural land and other land covers and provides a cost-effective and efficient way to map land-cover-type LCZs on a large scale [9], [22], [23], [24]. Sentinel 2 near-infrared (NIR) (B8), red-edge (B5, B6, B7) and red (B4) bands are particularly useful for agriculture, vegetation and, forest-type mapping [25]. Similarly, NIR (B8) and blue (B2) bands are particularly useful for extracting built-up features, such as buildings and other man-made infrastructures [26]. In addition, NIR (B8) region records low reflectance in the man-made features but high in vegetation, which makes it useful to separate man-made features from vegetation [26]. Several studies also used only red, green and blue (RGB) bands combination with deep learning model to classify land-cover [9], [27]. Therefore, incorporating different spectral bands from sentinel 2 or Landsat 8/9 is very useful to understand the characteristics of different vegetation types and classify them into distinct LCZs. Moreover, sentinel 1 synthetic aperture radar (SAR) image vertical transmit/vertical receive (VV) and vertical transmit/horizontal receive (VH) –bands also exhibit distinct characteristics on man-made surface and different vegetation types which makes it another important feature for LCZ mapping [28], [29]. And of course, NDVI is another important radiometric index for identifying healthy dense vegetation and separating them from man-made surface [9], [12]. Nighttime light data is also important to distinguish urban parks and vegetation cover from nonurban features and it has been widely used for LCZ classification [4], [16]. Nighttime light dataset captures artificial lighting produced by human settlements, infrastructure, and other

anthropogenic sources. Incorporating the nighttime light data in the classification model helps to separate urban parks/green-infrastructures, human settlements/infrastructures, forest covers and bare soil/sand. In addition, this article utilized monthly Level 2A processed data products from the global ecosystem dynamics investigation (GEDI) mission, a space-based laser mission designed to characterize Earth's ecosystems in three-dimensional (3-D). The GEDI Raster Canopy Height data contains relative height (RH) information of physical features. GEDI calculates RH, which determines the height of the vegetation canopy above the ground, as opposed to absolute height above sea level. Based on the RH information along with other spectral bands and indexes, each LCZs can be characterized, and decision rules can be established to classify GEDI shots into LCZ classes and the classified GEDI shots can be used as ground truth data in ML models.

Apart from factors selection for LCZ classification, selecting appropriate ML algorithms is also an important part of the classification process. Pixel-based LCZ classification using multisource RS data and ML is the most popular classification approach [12].Wide variety ML algorithms have been applied for local and global LCZ mapping including random forest (RF), support vector machine, neural network, Naïve bayers, and maximum likelihood classifier over the past years [16], [17], [30], [31], [32]. However, RF model standout as the most accurate and popular ML model for pixel-level LCZ classification because of its simplicity and effectiveness [12], [22], [24]. This article will develop a combined decision-rules and RF-based hybrid classification workflow for the land-cove- type LCZs mapping.

The aim of this article is to devise and validate a novel, automated classification technique that utilizes freely accessible RS data for mapping land-cover-type LCZs. In response to the challenges posed by expensive manual training sample collection and the variability in physical-geographical characteristics across different regions, we introduce a dual-phase hybrid classification process that integrates decision rules with a RF ML algorithm. This document will elaborate on our methodology, exhibit the outcomes from various test sites, explore the significance of our results, and propose future research directions in the context of LCZ mapping and its role in promoting sustainable urban development. Ultimately, this article explores how to effectively utilize freely available RS data for LCZ mapping, overcoming the limitations of expensive manual training and accommodating diverse physical-geographical characteristics and global weather patterns.

II. Materials and Methods

The entire workflow of the land-cover-type LCZ classification was implemented in Google Earth engine (GEE) with several steps including data preprocessing, combining multiple datasets, defining decision-rules, generating training samples, ML model development, and model performance evaluation. Fig. 2 displays the simplified workflow of automated land-cover-type LCZ classification model development process.

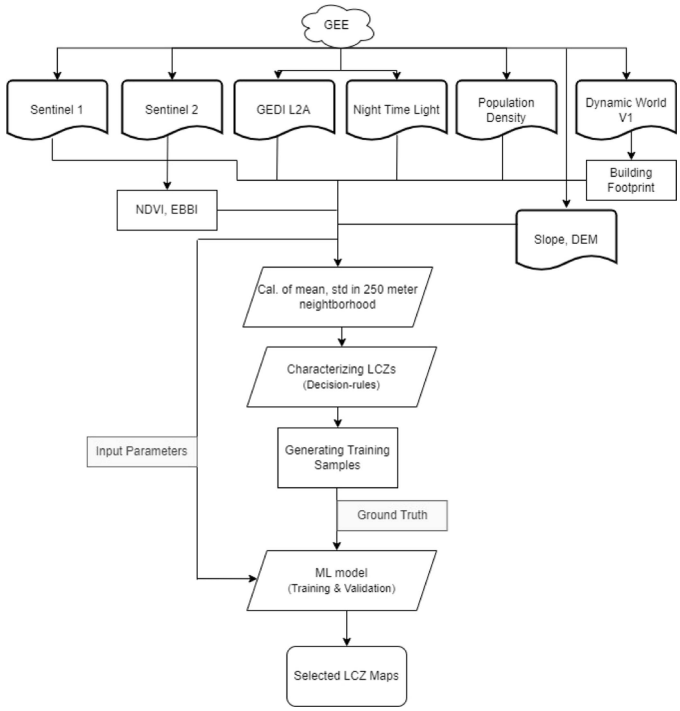

Fig. 2. Fully automated land-cover-type LCZ mapping workflow implemented in GEE.

A. Study Area

In this article, six experimental sites were selected across four continents, representing a diverse range of climatic regions, including temperate, equatorial/tropical, and subtropical areas. This diverse selection aims to evaluate the performance of the decision-rule based model across various geographical and climatic zones. The experimental sites include the north-western part of Fresno city in California, Central Michigan in Michigan, the western part of Phoenix city in Arizona, the surrounding region of Lagos city in Nigeria, part of the southern coastal districts in Khulna, Bangladesh, and the western part of Sydney city in Australia (see Fig. 3).

B. Data Collection

In this article, primarily GEDI Level 2A RH data, vertical-vertical (VV) and vertical-horizontal (VH) polarized bands from sentinel 1, B2-B8, and B11-12 bands from sentinel 2, visible infrared imaging radiometer suite (VIIRS) night-time day/night band (DNB) composite band, and several indices derived from spectral bands, such as NDVI, enhanced built-up and bareness index (EBBI) etc., were used. Sentinel 1, 2 and night-time light images for Fresno, Central Michigan, Phoenix, and Sydney region were taken from July to September in 2022 and their median image was calculated. However, for Lagos and Khulna region,

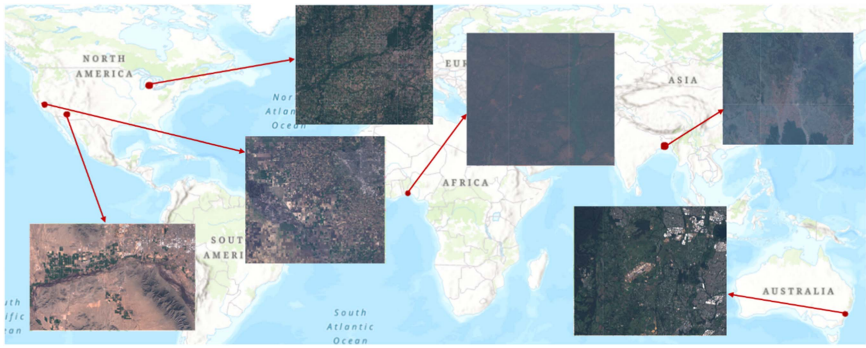

Fig. 3. Location of six experiment sites and sentinel 2 RGB composite bands.

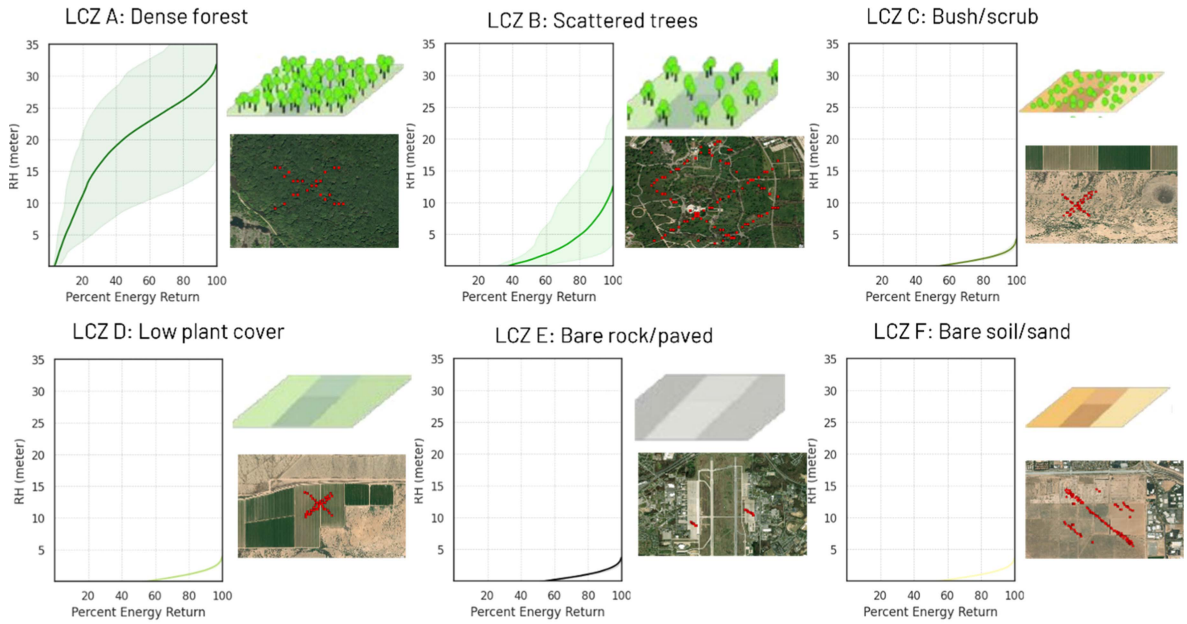

Fig. 4. GEDI RH profile for land-cover-type LCZ categories (LCZ A-F). LCZ G (i.e., waterbody is excluded from analysis).

the date range was taken in between January and February in 2022 in order to get cloud free images. In addition, waterbody image was extracted from Dynamic World V1 data from GEE [27]. All datasets were collected and processed using GEE platform and Google Colab.

1) GEDI Level 2A Data Product: This article used monthly level 2A processed data products derived from the GEDI mission, which is a space-based laser mission that aims to characterize the Earth's ecosystems in 3-D [33]. GEDI Level 2A data includes RH data at 25-m spatial resolution. RH data provides information about the vertical distribution of vegetation within the GEDI footprint. It represents the heights of various objects, such as the forest canopy, understory vegetation, and the ground surface, relative to a reference point. In this article, RH 100 (RH100) data was used. RH100 is a metric derived from GEDI level 2A data that represents the height at the 100th percentile of waveform energy relative to ground elevation. It is also referred to as the "maximum canopy height," which is the highest point within the vegetation canopy.

RH100 is a crucial parameter for characterizing forest structure and understanding ecosystem dynamics, particularly in estimating forest biomass and carbon stocks. Fig. 4 shows GEDI RH profile for land-cover-type LCZs.

While GEDI RH data provides valuable information about specific features, it is not sufficient on its own to fully understand the nature of those features and characterize each homogeneous LCZ. Information about neighboring features is also crucial for discerning both homogeneous and heterogeneous patterns, which in turn is essential for generating decision rules to identify homogeneous features and LCZ. For instance, if the RH of a GEDI shot in a forest area is 25 m and the average RH of the surrounding points within a neighborhood is only 3 m, this discrepancy suggests that the features likely belong to categories like agricultural land, low vegetation, or scattered trees. To address this, a 250-m circular neighborhood

was defined for each point, within which the average and standard deviation of the RH for all points were calculated. Additional attributes from GEDI level 2A data, including "urban proportion," help distinguish between built-up and nonbuilt-up areas by measuring the percentage of built-up land within a specific area around each GEDI shot. This measurement quantifies the level of urbanization in a given area. Furthermore, the proportion of built-up area and its average within a 250-m circular neighborhood were utilized to develop decision rules.

2) Sentinel 1 Satellite: Sentinel-1 is a satellite mission developed by the European Space Agency (ESA) that uses SAR to capture data about Earth's surface. Sentinel-1 carries a *C*-band SAR sensor, which operates at a wavelength of approximately 5.6 centimeters. The SAR sensor has two polarizations: VV and VH. The VV polarization refers to the transmission and reception of radar signals with a vertical orientation. It measures the backscatter intensity of radar waves that are polarized vertically. It is particularly useful for applications such as land cover classification, monitoring of agriculture and vegetation, and detection of changes in the environment. The VH polarization, on the other hand, refers to the transmission of radar signals with a vertical orientation and the reception of the backscattered radar waves with a horizontal orientation. VH polarization is sensitive to the orientation and characteristics of the surface, including the presence of man-made structures, buildings, and infrastructure [28]. It is often used for applications such as urban monitoring, identification of built-up areas, and detection of infrastructure changes. In this article, both VV, VH, and VV/VH bands were used in the decision rule and as input parameter in RF model.

3) Sentinel 2 Satellite: Sentinel-2 is a satellite mission operated by the ESA that provides high-resolution optical imagery of Earth's surface. The sentinel-2 satellite carries a multispectral sensor, which captures imagery in several spectral bands, each sensitive to different parts of the electromagnetic spectrum. These bands are designed to capture information about land cover, vegetation health, water quality, and other environmental parameters [25]. In this article, we selected only the bands with 10-m and 20-m spatial resolutions (e.g., B2, B3, B4, B5, B6, B7, B8, B11, and B12). These bands were then resampled to a 25-m resolution to make it consistent with GEDI 25-m spatial resolution data and used it as input parameters for the RF model.

4) Night-Time Light Data: The VIIRS Nighttime DNB Composites Version 1 data is a product derived from the VIIRS aboard the Suomi National Polar-orbiting Partnershipsatellite [34]. This dataset captures low-light imaging, allowing for observations of atmospheric features, city lights, and other nocturnal phenomena. The DNB capitalizes on moonlight, airglow, and other and anthropogenic light sources to produce detailed nighttime imagery. The night light data is useful to separate man-made features from features, such as built-up areas and forest cover or agricultural land [35]. As this article mainly targets mapping non-built-up land-cover-type LCZs, night-time light data plays a key role in separating man-made artificial and features and defining the decision-rules.

5) GEE Dynamic World Data: This article used Dynamic World dataset from GEE to extract specific land cover classes, such as built-up area, and waterbody [27]. Dynamic world provides near-real time global 10-m spatial resolution land-cover data. The built-up footprint was used to exclude built-up areas from the satellite image and limit the study areas only in nonurban areas targeting land-cover-type LCZs. The waterbody data was also extracted to exclude the corresponding pixels from all satellite data. It was used again with classified LCZs to replace waterbody pixels (i.e., LCZ F class).

C. Data Processing

1) Normalized Difference Vegetation Index: NDVI is a widely used vegetation index that quantifies the health and density of vegetation based on the reflectance of different wavelengths of light. It is calculated using the NIR and red (RED) reflectance bands of RS imagery. It is particularly useful to separate barren land, soil or sand from vegetation in the decision rules. The calculation of the NDVI raster employs a specific formula, denoted as

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{RED}}{\mathrm{NIR} + \mathrm{RED}}. \tag{1}$$

Here, NIR = Near-Infrared band, Red = Red band. The NDVI values are multiplied by 10 to make the scale -10 to 10. In addition, to remove noise in the NDVI data, the top and bottom 10% of NDVI values were trimmed. This process not only served as a method to reduce the impact of outliers and noise caused by clouds and other factors but also functioned as a separate parameter in decision-making rules. Trimming these extreme NDVI values is beneficial in diminishing the influence of such anomalies. Moreover, normalizing the data by eliminating these extremes leads to greater consistency, particularly when comparing different regions on a global scale. This approach ensures a more accurate and reliable analysis across diverse geographical areas.

2) Enhanced Built-Up and Bareness Index (EBBI): The EBBI is another RS index designed to effectively differentiate built-up areas from bare land in satellite images [36]. It makes use of the NIR, thermal infrared (TIR), and shortwave infrared (SWIR) bands to highlight these differences. The EBBI is especially useful in urban studies and land cover classification. The EBBI was calculated using

$$\mathrm{EBBI} = \frac{\mathrm{SWIR} - \mathrm{NIR}}{10 \times \sqrt{(\mathrm{SWIR} + \mathrm{TIR})}}. \tag{2}$$

Here, SWIR represents the reflectance in the SWIR band, and NIR represents the reflectance in the NIR. A higher EBBI value generally indicates built-up areas, while a lower value indicates bare land.

3) Defining Decision Rules: Decision rules for classifying GEDI shots into LCZ types were established by determining threshold values for each parameter. Several decision rules were generated for each land-cover-type LCZ category. The threshold values were manually determined by observing the relationship between each parameter for a specified LCZ class in the study areas. Example of decision-rules

LCZ A (Dense forest): urban_proportion_moving_average $<$ 1 And nighttime $<$ 10 And rh100_moving_average_in_250m_neighborhood $>$ 5 And NDVI10_90_moving_average$>$9.

In this decision rule, the "urban_proportion_moving_average" represents the average urban proportion of all points within a 250-meter circular neighborhood. The *"urban_proportion_moving_average"* values must be less than 1, indicating that the point is outside the urban area. Similarly, the "nighttime" value must be less than 10, indicating that the point is not in a built-up area. The *"rh100_moving_average_in_250m_neighborhood"* must be greater than 5, which helps in separating tall trees from short trees. The parameter "NDVI10_90_moving_average $>$ 9" represents the average of normalized NDVI values within a 250-m neighborhood surrounding a point. Setting the threshold at nine ensures that all neighboring trees are healthy, indicating dense vegetation, classified as LCZ A. For instance, if there is bare land within this 250-m neighborhood, the average NDVI value will fall below 9. In such cases, the area is classified as LCZ B, characterized by scattered trees or another LCZ type, rather than dense vegetation. Fig. 1 in Appendix section displays the RS parameters and thresholds values that are used to define the decision-rules.

Decision rules were defined by determining threshold values for various parameters to characterize each land-cover-type LCZ. These decision rules were then applied across the study areas to classify all GEDI shots within them. The main objective of these decision rules was to accurately classify GEDI shots into LCZ types and generate as many training samples as possible automatically as ground truth data for ML models. The classified GEDI shots were used as ground truth/training label data in a RF model, while the spectral bands from sentinels 1 and 2 and spectral indices were used as input parameters for training the LCZ classification model. Subsequently, the trained model was applied to the entire study area to classify LCZ categories. Therefore, the proposed classification approach is a two-stage hybrid approach, wherein the training samples are automatically generated using decision rules, and subsequently, an ML model is trained on these samples to classify LCZs across the entire study area. The decision to not develop a specific rule for classifying waterbodies (LCZ G) in our study was made because waterbodies represent a common land use and land cover (land-cover) class type, for which several accurate and globally available data sources already exist. Consequently, this article utilized the recently available waterbody layer from the Dynamic World data, ensuring both accuracy and consistency in our approach to land-cover classification. This workflow can be adopted to different geographic locations in order to obtain global LCZ map.

D. Random Forest Model

RF is a versatile and powerful ML algorithm used for both regression and classification tasks specially in RS field [8], [22], [24]. It operates by constructing multiple decision trees during training and outputs the mode of the classes (for classification) or mean prediction (for regression) of individual trees for a given input. Its ensemble nature makes it resistant to overfitting, as it aggregates the predictions of numerous trees. RF handles missing values effectively and can manage large datasets with higher dimensionality. It also provides insights into feature importance, helping in feature selection and understanding key drivers for predictions. The generated training samples of each study area were entirely used to train a RF model. Finally, the trained RF model was used to classify the entire study areas in land-cover-type LCZ zones.

E. Validation of Classified Maps

To validate the accuracy and reliability of the output maps generated from the LCZ classification model, a rigorous validation process was implemented for each study site. Initially, a set of 200 random points was generated for every study area. These points served as the basis for ground truth data collection, a critical step in the validation process.

For each of these random points, ground truth data were meticulously collected through manual inspection. This involved analyzing high-resolution Google satellite imagery and, where available, cross-referencing with additional sources, such as sentinel 2 (10-m RGB bands) satellite imagery and Google street view. The manual inspection process was designed to ensure a comprehensive and accurate understanding of the actual land cover at each point.

By comparing the model's classifications at these points against the manually collected ground truth data, a clear picture of the model's performance was obtained. This comparison allowed for the assessment of the model's precision and accuracy in correctly classifying each LCZ type. It also provided an opportunity to identify and understand any patterns of misclassification or areas where the model might struggle, such as distinguishing between classes with overlapping characteristics.

This validation process is crucial as it not only measures the model's effectiveness in classifying different LCZ types, but also ensures that the training and decision-making processes of the model are robust and reliable across varied geographical and climatic conditions. The outcome of this validation step is crucial for confirming the model's utility and applicability in real-world scenarios across various study sites. It is also vital for ultimately generating global LCZ data with a finer spatial resolution of 25 m.

F. Accuracy Assessment Metrics

For validation analysis, the following accuracy assessment metrics were used to validate the model's output and the classified maps-

1) Precision Score: Precision is a measure of the model's ability to correctly identify positive samples from the total predicted positive samples. It quantifies the proportion of true positives (correctly predicted positive samples) out of all samples predicted as positive. Precision is useful in scenarios where the cost of false positives is high, and we want to minimize the number of false positive predictions. Equation (3) was used to

TABLE I
GENERATED TRAINING SAMPLES FROM EACH STUDY AREA

Experimental sites	Total Area (sq.km.)	GEDI shots on non-built-up area	Training samples by decision-rules	Percentage
Fresno, CA	3422.45	172043	69603	40 %
Central Michigan, MI	5583.85	329722	135835	41 %
Phoenix, AZ	2220.88	92275	74774	81%
Khulna, Bangladesh	9121.71	133649	93549	70%
Lagos, Nigeria	3088.34	17614	14726	84%
Sydney, Australia	712.08	42906	25834	60%

Gedi shots that landed on built-up areas were excluded.

calculate precision scores

$$\text{Precision Score} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{3}$$

Here, TP represents true positives and FP represents false positives.

2) Accuracy Score: Accuracy is a measure of overall correctness in the model's predictions. It calculates the proportion of correctly predicted samples (both true positives and true negatives) out of the total number of samples. Accuracy provides an overall assessment of the model's performance across all classes. Equation (4) was used to calculate accuracy scores

$$\text{Accuracy score} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{4}$$

Here, TN and FN represent true negatives, and false negatives, respectively.

3) Recall: Recall, also known as sensitivity or true positive rate, is a performance metric used in classification tasks to measure the proportion of actual positives correctly identified by the model. Equation (5) was used to calculate recall

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{5}$$

4) F1 Score: The F1 score is a harmonic mean of precision and recall, providing a single metric that balances the tradeoff between the two. It is especially useful in scenarios where either false positives or false negatives have significant consequences. The F1 score ranges from 0 (worst) to 1 (best), with 1 indicating perfect precision and recall. Equation (6) was used to calculate F1scores.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{6}$$

III. RESULT AND DISCUSSION

Predefined decision rules for land-cover-type LCZs (i.e., LCZ A-F) were employed to classify GEDI points, aiming to automatically gather as many training samples as possible for all six study areas. Table I given the number of training samples generated using these decision rules. For Fresno in CA, the decision rules were able to classify 40% of the shots; for Phoenix, it was 81%; for Central Michigan, it was 41%; for Khulna, it was 70%; for Lagos, it was 84%; and for Sydney, it was 60%. The decision rules ensured that the points were correctly classified, and the ambiguous points, which were difficult to classify based on the decision rules, were left out. The results indicate that a substantial quantity of training samples can be automatically generated via the decision rules. While some noise emerged during this process, its impact is likely minimal given the vast number of correctly classified samples. It's assumed that the noise is counteracted by the overwhelming volume of accurately classified samples. To validate the accuracy of these generated training samples, the classified maps ware compared with validation data. The validation data was derived manually from the Sentinel's 10-mr RGB composite bands and Google Maps satellite imagery.

Table II gives the number of training samples generated by decision rules for land-cover-type LCZs in each study site. In the Fresno area, there were no training samples in LCZ E type as there is no bare rock. The dominant category is LCZ C, scrub/bush/sort woody trees. Based on observation of Google Map Street view imageries in May and October 2023, we found heavy presence of short woody trees, possibly fruits trees in the areas. The second most dominant class is dense forest (LCZ A). The average height of the LCZ A is greater than 3 m [2]. We have found an abundance of dense and tall fruit trees in the areas which belong to LCZ A. The classified points for each LCZ type are visualized in Fig. 5. It can be observed from Fig. 5 that the decision rules can classify each GEDI point into LCZ types. LCZ A image patch presents classified GEDI shots primarily for dense

TABLE II
NUMBER OF TRAINING SAMPLES GENERATED FOR EACH LCZ TYPE IN EACH STUDY STIES BY DECISION-RULES

Categories	Training samples					
	Fresno	Central Michigan	Phoenix	Khulna	Lagos	Sydney
LCZ A: Dense forest	21 314	93 414	2152	65 015	8748	17 741
LCZ B: Scattered trees	3429	653	920	8262	2036	5794
LCZ C: Scrub/ Bush/ Short woody trees	26 843	8065	39 491	2062	2944	302
LCZ D: Low plant/ agricultural land	10 937	9014	9595	18 061	944	1889
LCZ E: Bare rock or paved	-	-	1325	2	10	-
LCZ F: Bare soil or sand	7080	-	21 579	147	44	108

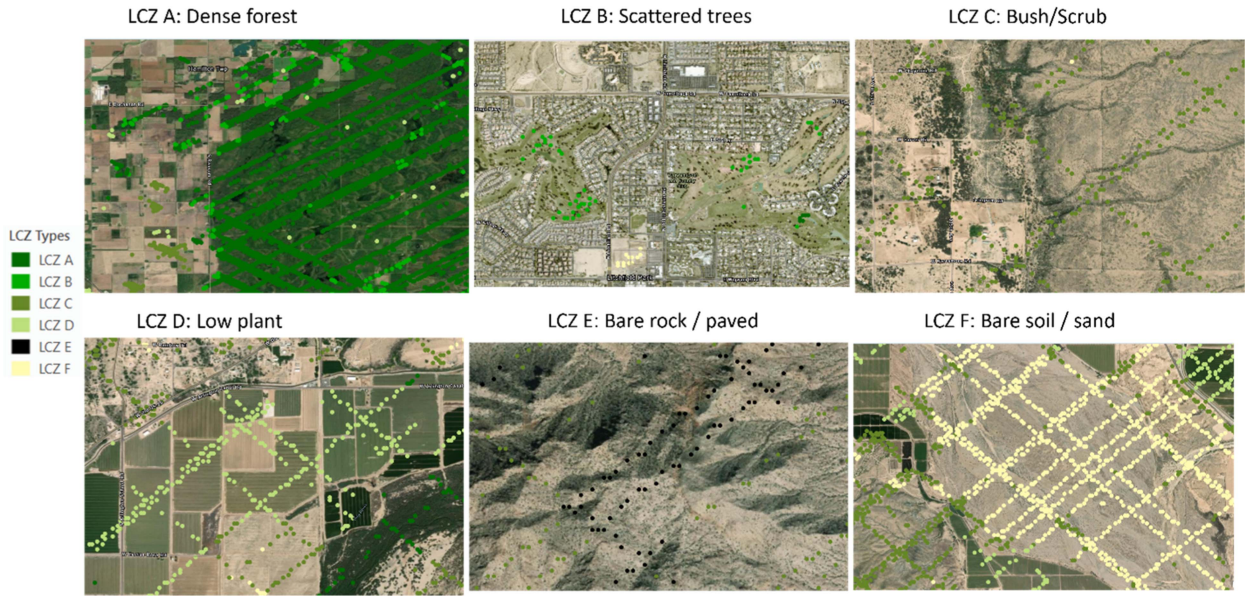

Fig. 5. Classified GEDI points for each LCZ type by decision-rules. LCZ G (waterbody) excluded.

forest, LCZ B presents classified urban parks, LCZ C presents bush/scrub/short woody trees, LCZ D presents classified shots for low plant/agricultural land, LCZ E presents bare rock and LCZ F present bare soil/sand. While there may be noise in the generated training samples, the overwhelming number of correctly generated training samples may suppress the noise in the ML model. Therefore, it can be concluded that the decision rules are effective in automatically generating training samples. These classified GEDI points can be used as ground truth data in the ML model to classify LCZs for the entire area.

Fig. 6 displays box plot of f1, precision and recall scores for each LCZ class. It can be seen that in all six study sites, the average f1, precision, and recall scores for LCZ A, C and F is above 85%.

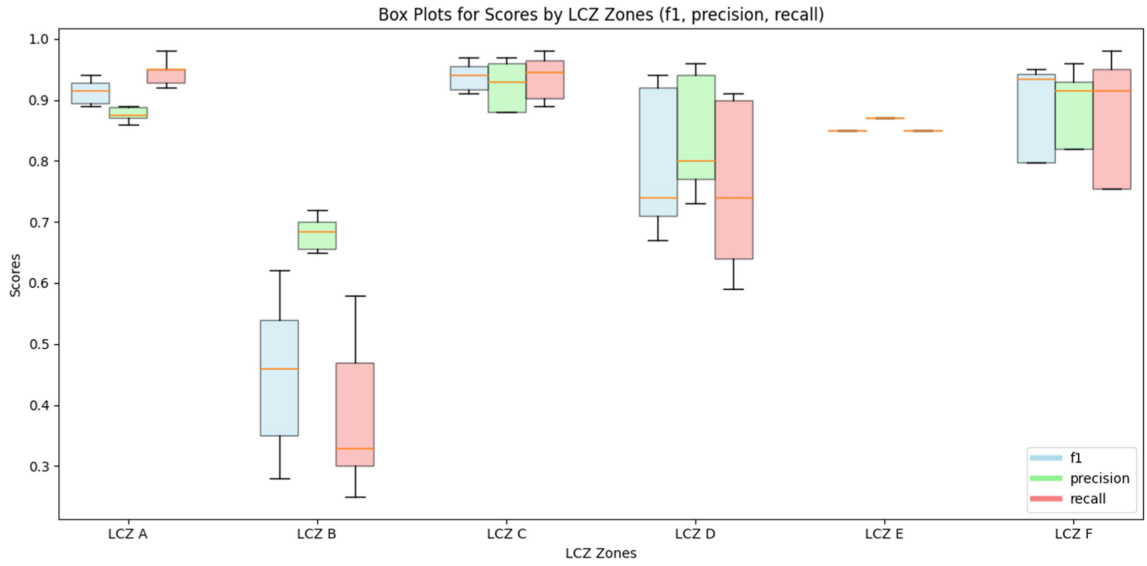

Fig. 6. Box plots of average, 25th and 75th percentile f1, precision and recall scores for each LCZ class. The scores are calculated based on ground truth data.

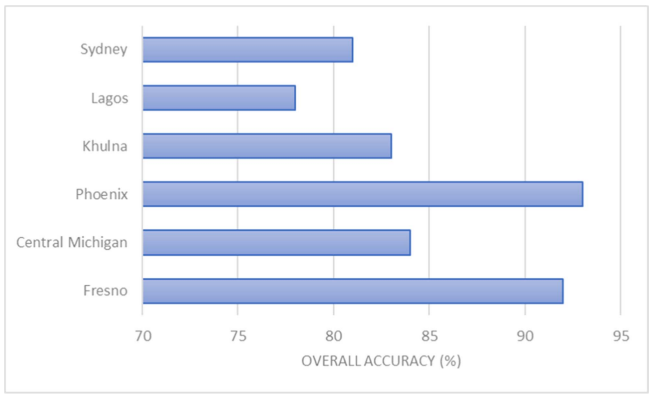

Fig. 7. Overall accuracy score for each study site.

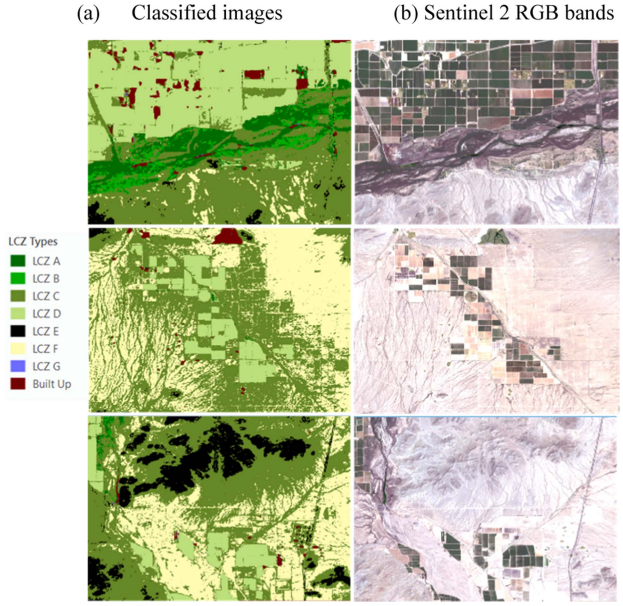

Fig. 8. (a) Classified LCZ map (left) along with (b) corresponding sentinel 2 RGB band composite images (right) for Phoenix, Arizona.

The scores for LCZ B are the lowest in all six study areas as large, scattered trees are usually mixed-up with low plant, vegetation and dense trees. The characteristics of LCZ B are also similar to LCZ C (i.e., bush/scrub/short woody trees) except for their height difference (see Fig. 1). Therefore, it is difficult to separate LCZ B using RS data. Among all six experimental sites, the LCZ E (bare rock) is only present in Phoenix. Therefore, scores for this category are represented only for Phoenix. We found that in the rocky hill, there is presence of bush/scrub (i.e., LCZ C). Therefore, it is difficult to assign those areas in either LCZ C or LCZ E.

However, high slope and low NDVI scores helped to accurately identify and classify LCZ E.

Fig. 7 displays overall accuracy scores for the six study sites. It can be seen that overall accuracy is lowest 78% in Lagos and highest 93% in Phoenix. Several reasons are behind the low accuracy in Lagos area including poor image quality and ambiguous surface features. Besides, it is difficult to acquire cloud free images over that region and we had to change the selected months (i.e., May to October) and choose only January and February to acquire cloud free images for Lagos. Overall, the accuracy scores of all six study areas are above 75% based on validation data which indicates the proposed method is capable of automatically generating land-cover-type LCZ map regardless of physical, geographical and climatic variation.

Fig. 8 displays the classified map for Phoenix, Arizona, alongside the sentinel 2 RGB band composite. Upon visual examination, the model appears to have successfully classified

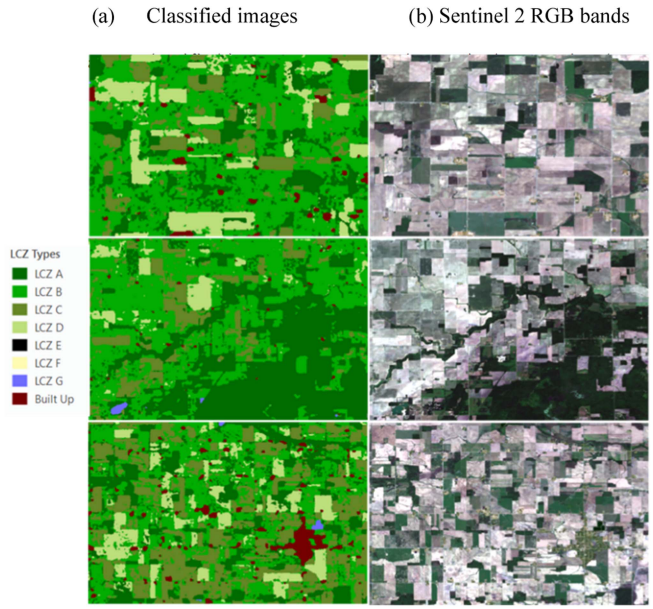

Fig. 9. (a) Classified LCZ map (left) along with (b) corresponding sentinel 2 RGB band composite images (right) for Central Michigan.

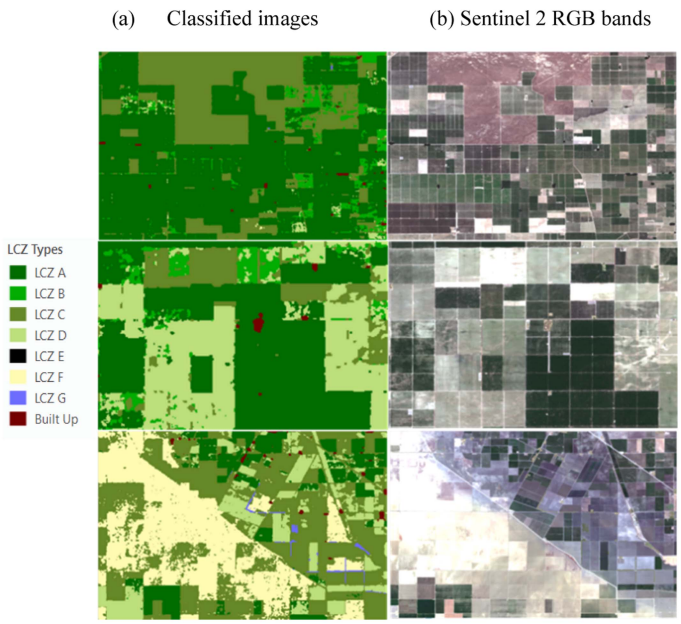

Fig. 10. (a) Classified LCZ map (left) along with (b) corresponding sentinel 2 RGB band composite images (right) for Fresno, CA.

all categories, particularly the agricultural land/low plant (i.e., LCZ D), rock/paved surface (LCZ E) and bush/scrub (i.e., LCZ F). Some minor misclassifications between bush/scrub (LCZ F) and bare soil/sand (LCZ G) can be observed. However, these classes often exhibit overlapping characteristics, which can lead to such misclassifications (referencing LCZ C and F in Fig. 1). LCZ B - open setting trees or scattered trees seems to be correctly classified by the model. Fig. 2 in Appendix section shows the classified map for the entire study area.

In Central Michigan (see Fig. 9), there's a significant presence of scattered trees (i.e., LCZ B). The dominance of this category might be attributed to the presence of woody trees or weeds within the agricultural land. Dense forest cover (i.e., LCZ A) is another dominant class as we can see the presence of dense forest cover in the RGB image. In addition, agricultural land stands out as another dominant class. Overall, the classification seems to accurately present the LCZ classes in the ground for that specific time-period. Fig. 3 in Appendix section shows the classified map for the entire study area.

In Fresno, CA (see Fig. 10), the dominant classes are dense forest (LCZ A) and scattered trees (LCZ B). As observed in the Phoenix and Central Michigan regions, the prominence of short woody trees or weeds within agricultural lands might account for the prevalence of this category in the Fresno area. In addition, it appears that some agricultural land (LCZ D) has been misclassified as dense Tree (LCZ A) or scattered Trees (LCZ B). This situation is particularly perplexing because, based on the satellite image, these areas seem to be agricultural plots with crops. Nevertheless, upon a visual inspection using Google satellite imagery and Google Street View (street view image: May, 2023), we have identified a significant presence of dense and tall fruit trees in the areas, which closely resemble either dense trees (LCZ A) or scattered trees (LCZ B). Despite these plants initially appearing to be agricultural crops (LCZ D) in the satellite image, their actual height exceeds 3 m, categorizing them as either LCZ A or B (representing dense or scattered trees with an average height greater than 3 m). Therefore, these areas have been correctly classified into their respective LCZ types based on the criteria (see Fig. 1). Fig. 4 in Appendix section shows the classified map for Fresno study area.

In Khulna (see Fig. 11), satellite images were captured between January and February in 2022. In contrast, the water-body imagery from the Google dynamic worldview database was taken from July to September, a period that aligns with the rainy season. Consequently, a significant portion of the area appeared as a waterbody. The predominant classes are low plant/agricultural Land (i.e., LCZ D) and dense trees (i.e., LCZ A). In this region, small settlements are enveloped by large trees, complicating the differentiation of settlements. As a result, some settlements might have been mistakenly classified under the dense tree category. Nevertheless, the classification appears reliable based on visual assessments of satellite imagery. Fig. 5 in Appendix section shows the classified map for Khulna study area.

Similar to Khulna, the satellite images for the Lagos area were captured between January and February 2022. However, the waterbody imagery sourced from the Google dynamic worldview database dates from July to September, coinciding with the rainy season. As a result, a substantial portion of the area was identified as a waterbody (see Fig. 12). Upon closer examination of the satellite images, we observed an abundance of both

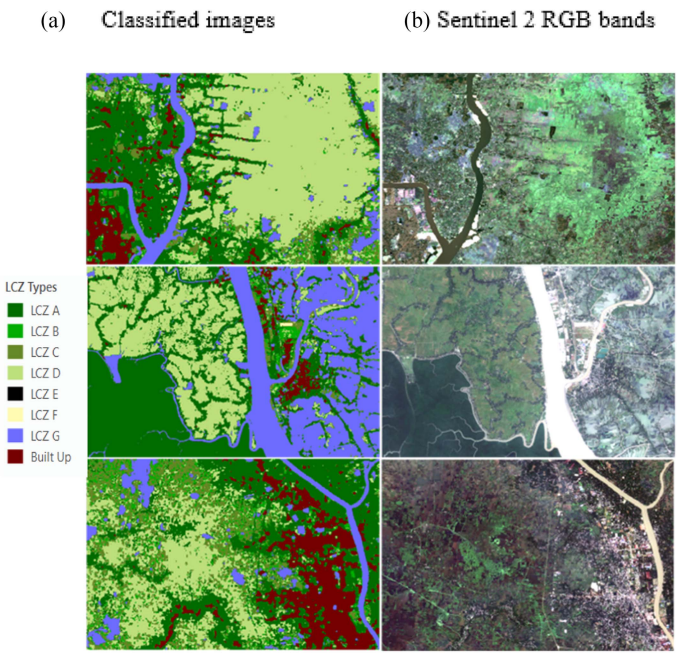

Fig. 11. (a) Classified LCZ map (left) along with (b) corresponding sentinel 2 RGB band composite images (right) for Khulna.

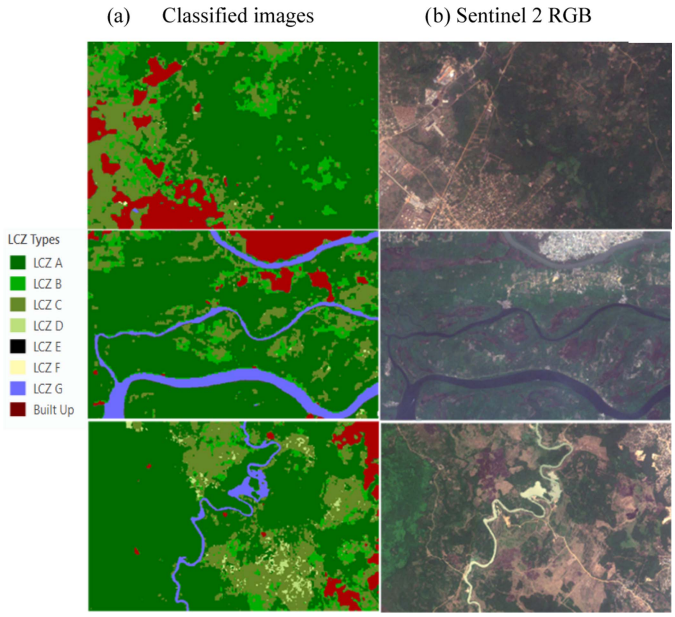

Fig. 12. (a) Classified LCZ map (left) along with (b) corresponding sentinel 2 RGB band composite images (right) for Lagos.

large and small woody trees that contribute to dominance of LCZ A & B class. Overall, the classification appears to be reliable based on a visual assessment of the satellite imagery. Fig. 6 in Appendix section shows the classified map for Lagos area.

In the Western Sydney region (see Fig. 13), Dense Trees (i.e., LCZ A) and Scattered Trees (i.e., LCZ B) are the predominant classes. A significant portion of the area is covered

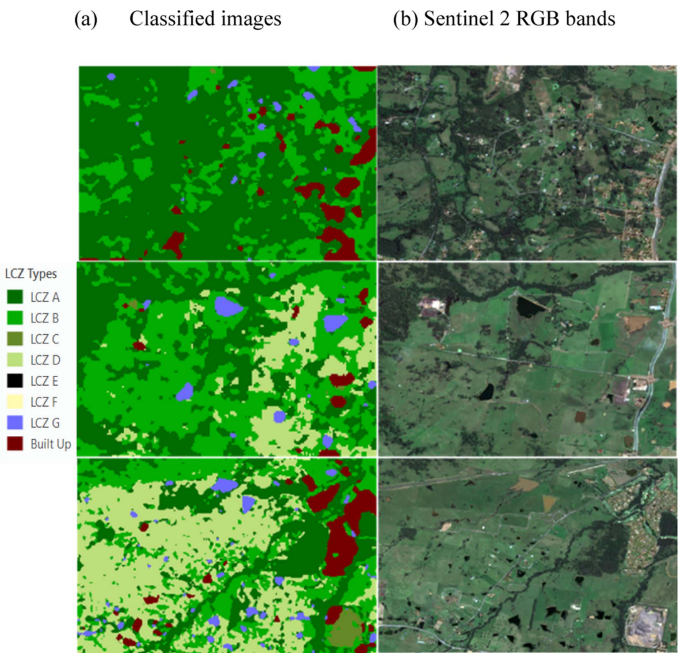

Fig. 13. (a) Classified LCZ map (left) along with (b) Corresponding Sentinel 2 RGB band composite images (right) for Sydney.

by settlements, which were excluded from our analysis, but included at the post classification stage from dynamic world land-cover data. A detailed examination of the satellite imagery reveals an abundance of both large and small woody trees in this region. Overall, the classification appears accurate based on a visual assessment of the satellite images. Fig. 7 in Appendix section shows the classified map for Sydney area.

IV. Conclusion

This article introduces an automated mapping approach using decision rules and ML to identify select LCZ classes, such as dense tree, scattered tree, bush/scrub/short woody trees, low plant/agricultural land, bare rock/paved, and bare soil/sand. Decision rules were formulated by manually selecting thresholds for various parameters, such as RH and NDVI. After generating training samples, a RF model was trained to categorize these chosen LCZ classes. Six experimental sites were selected spanning both the Northern and Southern hemispheres, as well as tropical/equatorial regions. The experiments yielded promising outcomes across the six sites, with an average accuracy exceeding 75%. A two-stage validation by ground truth data and visual inspection of high-resolution imageries affirmed the efficacy of the automated model, demonstrating its capability to accurately classify detailed vegetation covers and the chosen LCZ classes. However, sourcing high-quality satellite images can be challenging in some parts of the world, which can impact the classification process, leading to relatively poor results. In the future, this model could be expanded to classify other LCZ classes and contribute to

generating global scale 25-m spatial resolution LCZs maps. This hybrid model, combining automated decision rules with ML, presents a viable alternative to supervised classification, where manually collecting training data can be both tedious and error prone.

APPENDIX

In Fig. 14, the orange lines represent the threshold values selected for various parameters, which are instrumental in defining the decision rules for categorizing areas into different LCZs. For example, in the case of LCZ C, which typically includes bush or scrub in desert areas, night-time light data is utilized as a crucial parameter. The rationale behind this is that such areas are expected to be devoid of artificial lights. Conversely, in LCZ B, which encompasses areas with scattered trees, such as urban parks, golf courses, and cemeteries, artificial lights may be present. Therefore, in these zones, night-time light data is not a primary parameter.

The selection of these threshold values is a meticulous process, involving manual examination of various characteristics observed in different RS bands and indices. This detailed analysis is essential to establish accurate decision rules that effectively classify GEDI shots into their respective LCZ zones. Each orange line signifies the specific threshold established for a parameter, aiding in the differentiation and classification of areas based on their unique characteristics and RS data. For LCZ A, there are total three distinct decision-rules that were defined to accurately detect medium large and dense trees (LCZ A_1), large trees (LCZ A_2) and large dense trees (LCZ A_3).

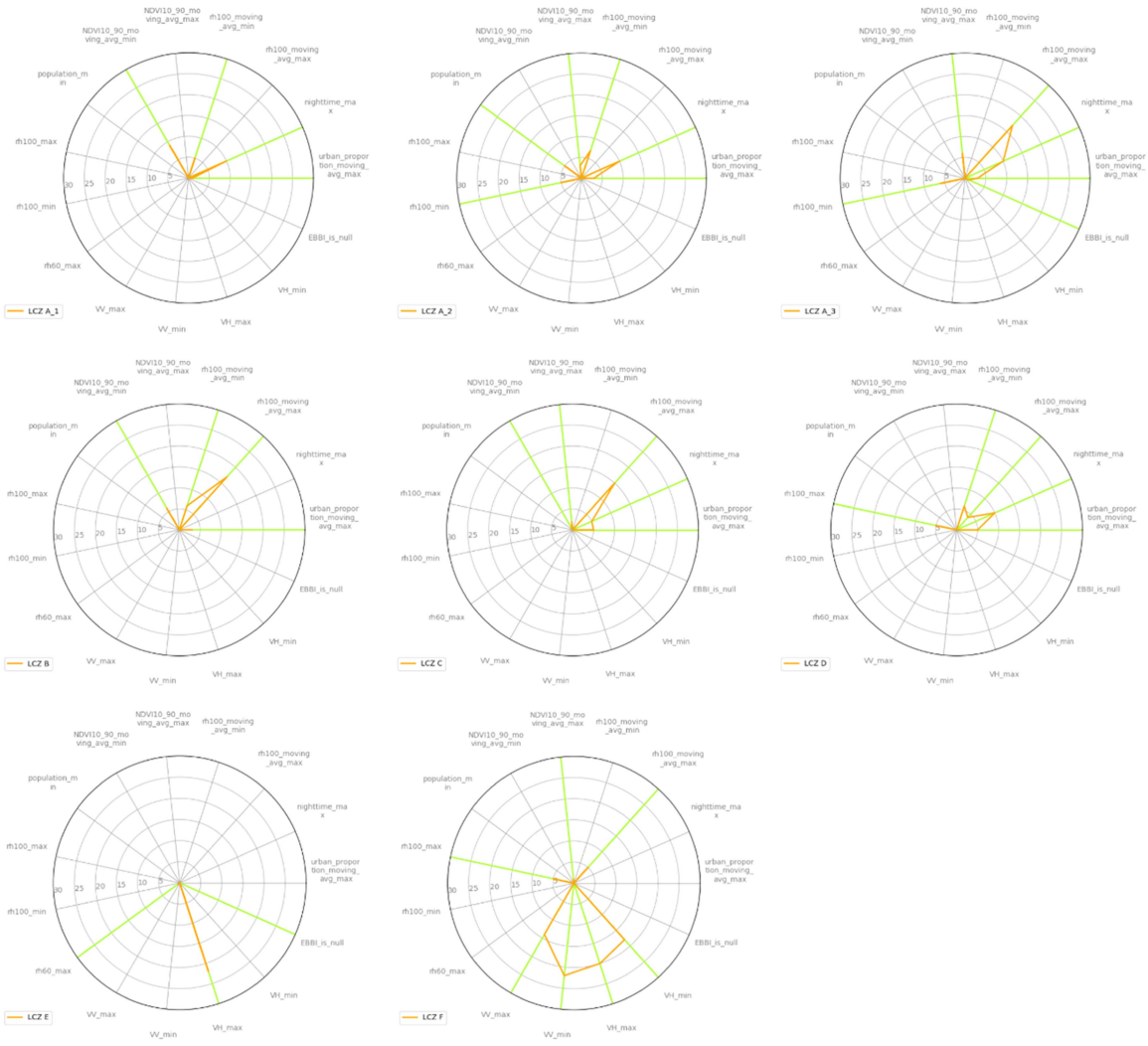

Fig. 14. Display the parameters and thresholds used to define the distinct decision-rules for each land-cover-type LCZs. The orange line indicates the threshold values and the highlighted lime color lines indicates which parameters are used to define the decision-rules for each LCZs. N.B. the parameters are not converted into common scale. The legend of each figure indicates the type of LCZ.

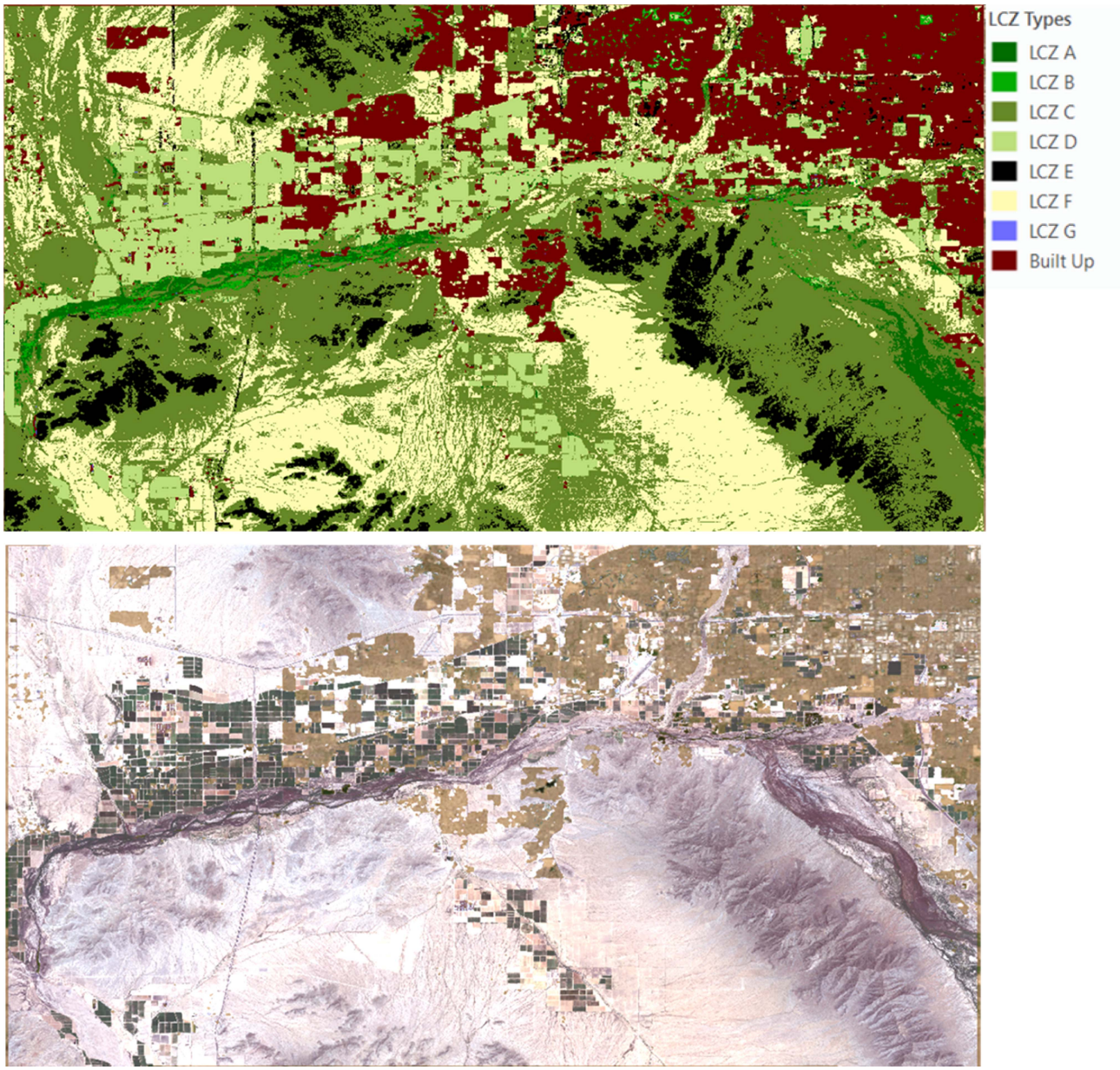

Fig. 15. Classified land-cover type-LCZ map of Phoenix, AZ and sentinel 2 RGB composite bands.

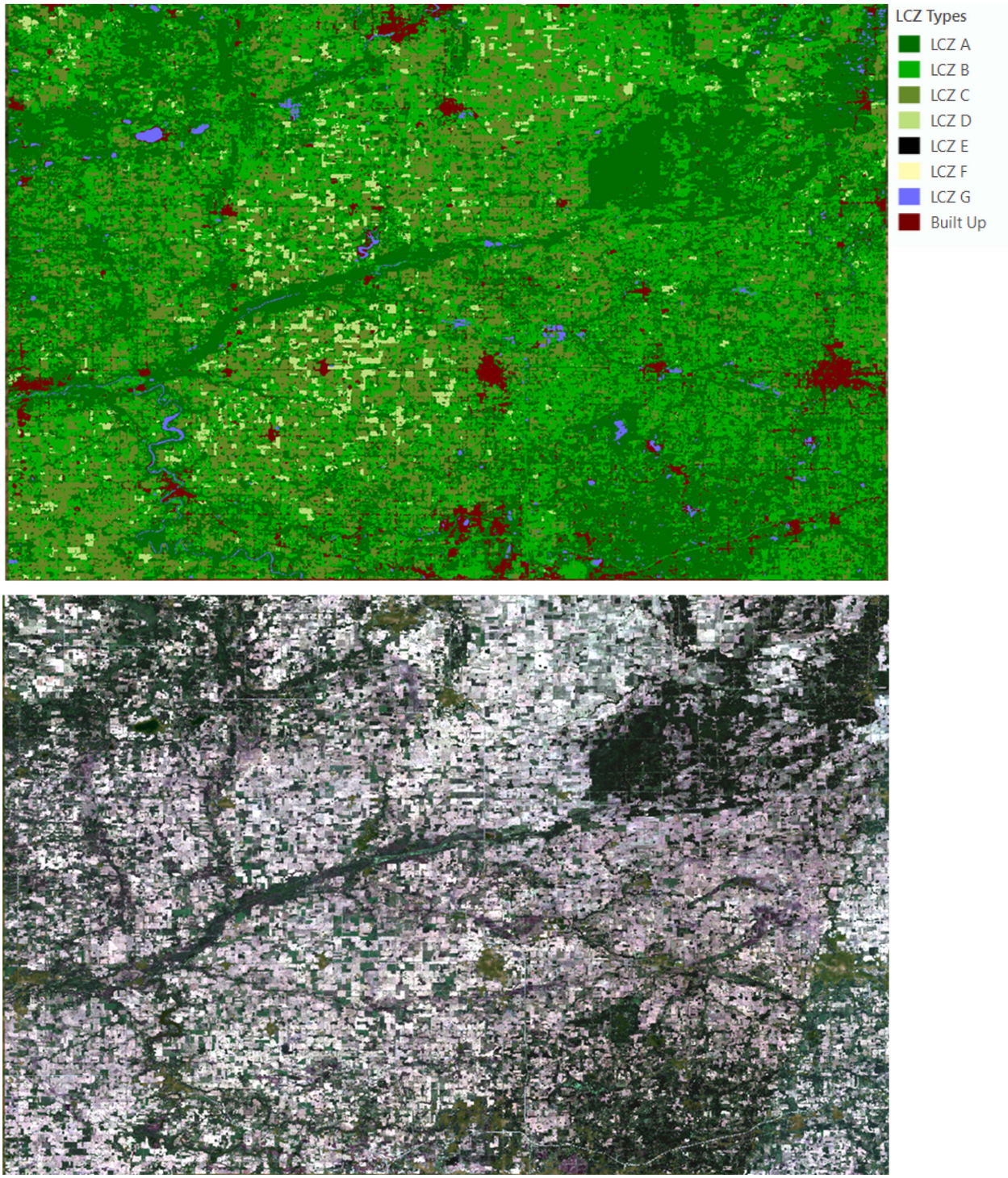

Fig. 16. Classified land-cover type-LCZ map of Central Michigan and sentinel 2 RGB composite bands.

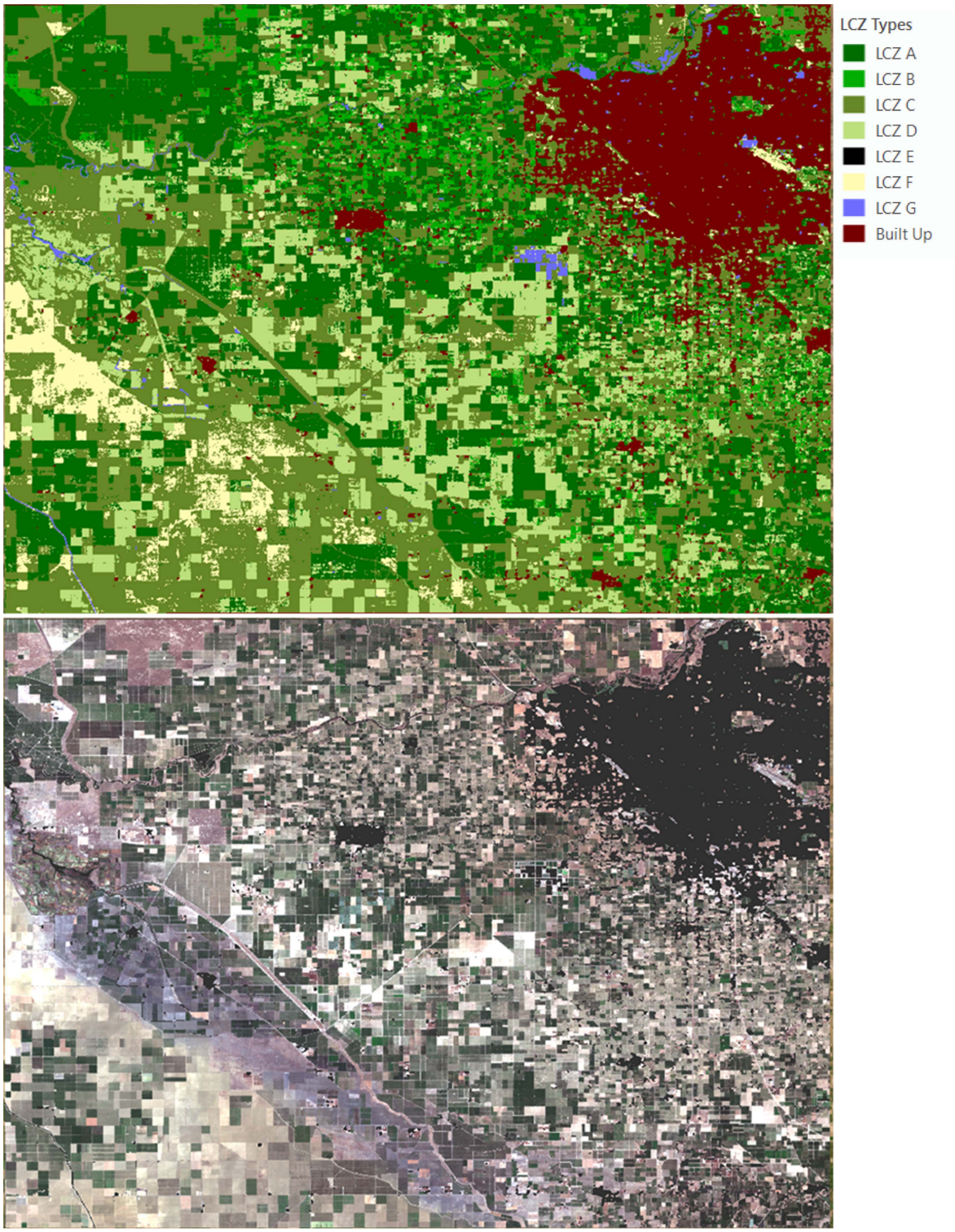

Fig. 17. Classified land-cover type-LCZ map of Fresno, CA and sentinel 2 RGB composite bands.

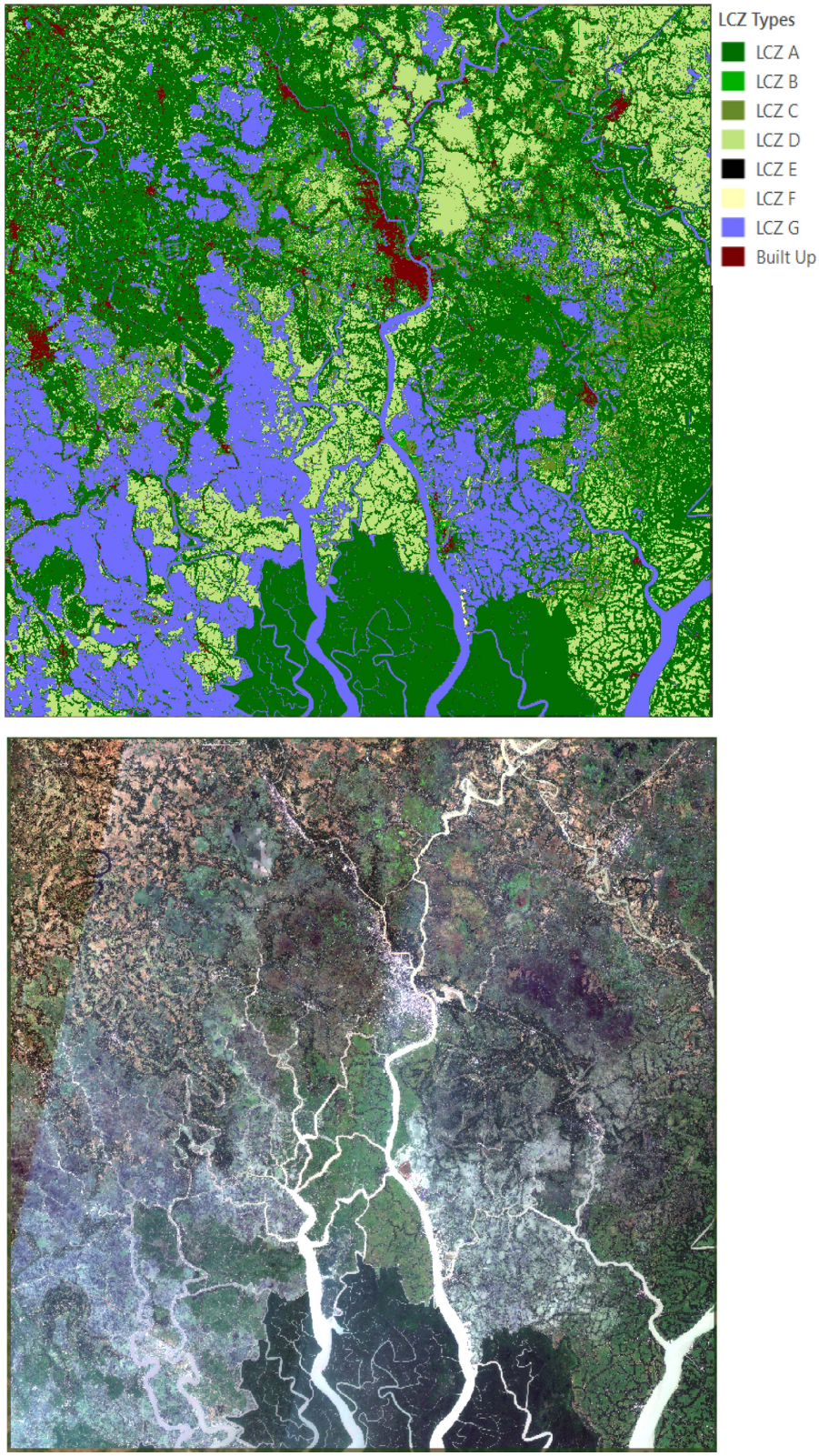

Fig. 18. Classified land-cover-type-LCZ map of Khulna, Bangladesh and sentinel 2 RGB composite bands.

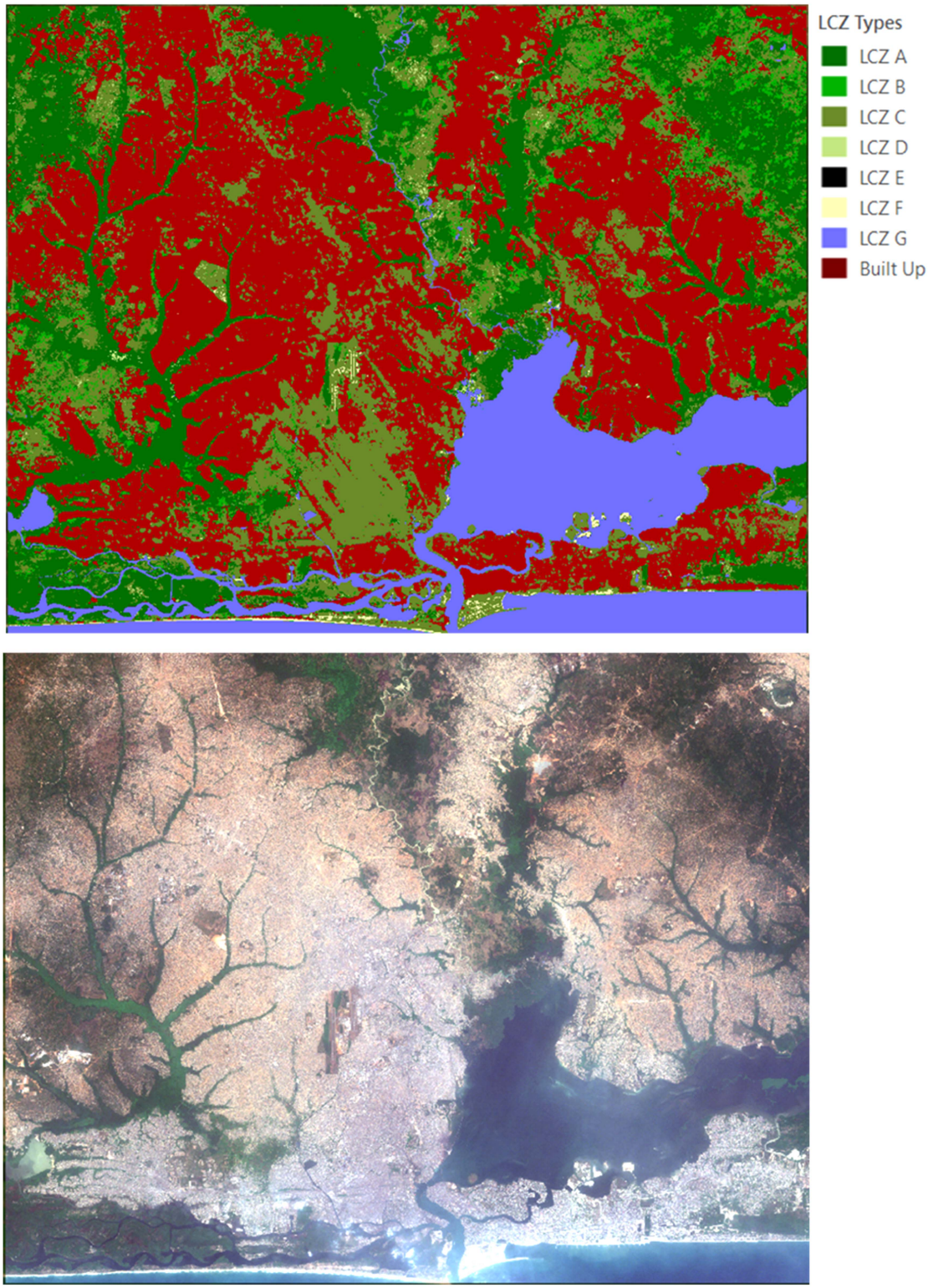

Fig. 19. Classified land-cover-type-LCZ map of Lagos, Nigeria and sentinel 2 RGB composite bands.

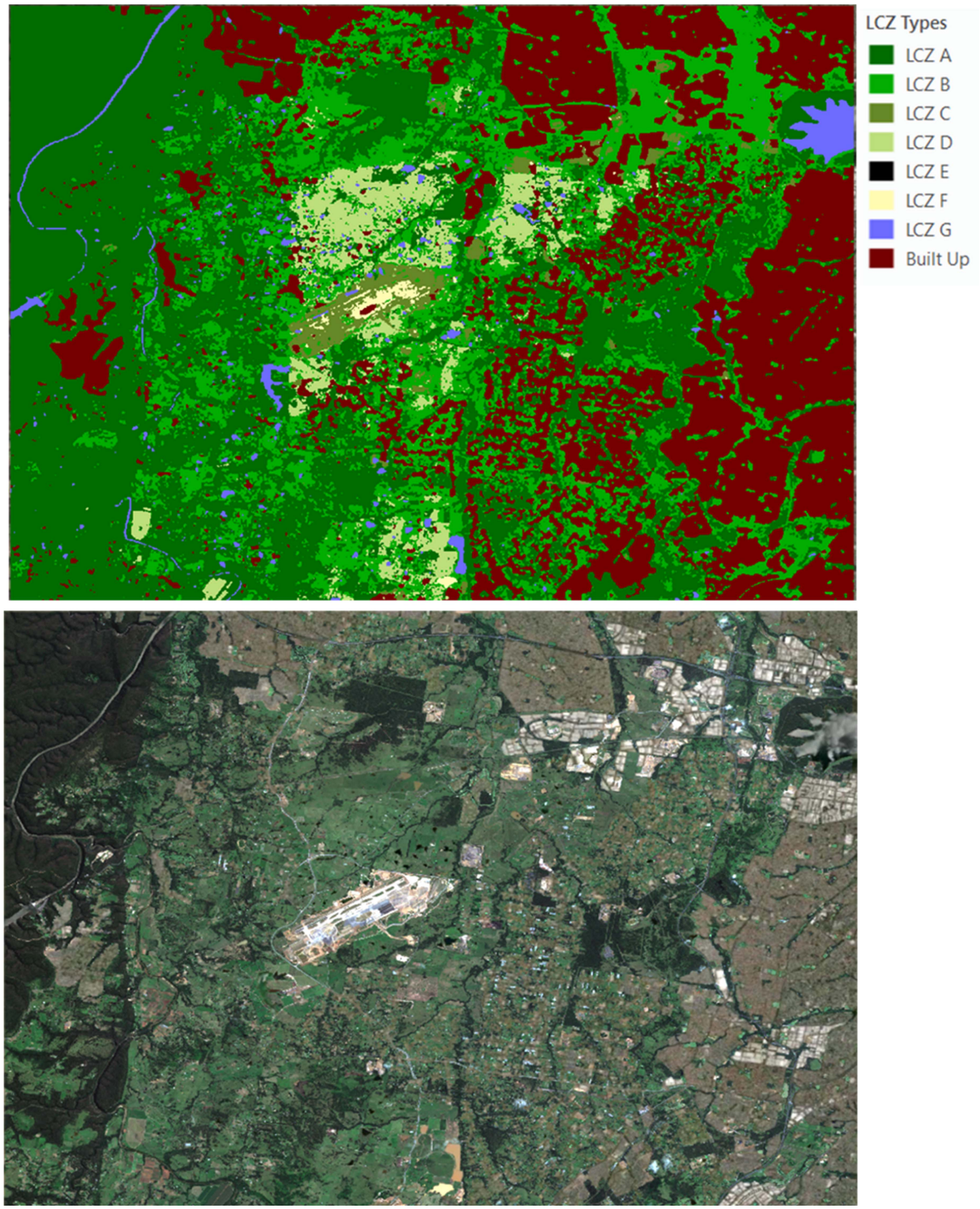

Fig. 20. Classified land-cover type-LCZ map of Sydney, Australia and sentinel 2 RGB composite bands.

REFERENCES

[1] N. Kaloustian and B. Bechtel, "Local climatic zoning and urban heat island in Beirut," *Procedia Eng.*, vol. 169, pp. 216–223, Jan. 2016.

[2] I. D. Stewart and T. R. Oke, "Local climate zones for urban temperature studies," *Bull. Amer. Meteorol. Soc.*, vol. 93, no. 12, pp. 1879–1900, Dec. 2012.

[3] M. Demuzere, J. Kittner, and B. Bechtel, "LCZ generator: A web application to create local climate zone maps," *Front. Environ. Sci.*, vol. 9, Apr. 2021, Art. no. 637455.

[4] M. Demuzere et al., "A global map of local climate zones to support earth system modelling and urban-scale environmental science," *Earth Syst. Sci. Data*, vol. 14, no. 8, pp. 3835–3873, Aug. 2022.

[5] B. Bechtel et al., "Mapping local climate zones for a worldwide database of the form and function of cities," *ISPRS Int. J. Geo-Inf.*, vol. 4, no. 1, pp. 199–219, Feb. 2015.

[6] J. S. Mukherjee, *The Millennium Development Goals and Sustainable Development Goals*, vol. 1. Oxford, U.K.: Oxford Univ. Press, 2017.

[7] S. Hawkes and K. Buse, "Sustainable development goals," in *Public Health*. Oxford, U.K.: Oxford Univ. Press, 2019.

[8] E. G. Yu et al., "Rice modeling using long time series of high temporal resolution vegetation indices in Nepal," in *Proc. 10th Int. Conf. Agro-Geoinformatics*, 2022, pp. 1–6.

[9] M. D. Islam, L. Di, M. R. Mia, and M. S. Sithi, "Deforestation mapping of Sundarbans using multi-temporal sentinel-2 data & transfer learning," in *Proc. 10th Int. Conf. Agro-Geoinform.*, 2022, pp. 1–4.

[10] A. W. Komba, T. Watanabe, M. Kaneko, and M. B. Chand, "Monitoring of vegetation disturbance around protected areas in central Tanzania using landsat time-series data," *Remote Sens.*, vol. 13, no. 9, Jan. 2021.

[11] B. E. Taiwo et al., "Monitoring and predicting the influences of land use/land cover change on cropland characteristics and drought severity using remote sensing techniques," *Environ. Sustainability Indicators*, vol. 18, Jun. 2023, Art. no. 100248.

[12] F. Huang et al., "Mapping local climate zones for cities: A large review," *Remote Sens. Environ.*, vol. 292, Jul. 2023, Art. no. 113573.

[13] G. P. Asner and R. E. Martin, "Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels," *Remote Sens. Environ.*, vol. 112, no. 10, pp. 3958–3970, Oct. 2008.

[14] Y. Wu, Y. Wu, B. Wang, and H. Yang, "A remote sensing method for crop mapping based on multiscale neighborhood feature extraction," *Remote Sens.*, vol. 15, no. 1, Jan. 2023.

[15] J. - F. Bastin et al., "Understanding climate change from a global analysis of city analogues," *PLoS One*, vol. 14, no. 7, Jul. 2019, Art. no. e0217592.

[16] B. Bechtel et al., "Mapping local climate zones for a worldwide database of the form and function of cities," *ISPRS Int. J. Geo-Inf.*, vol. 4, no. 1, pp. 199–219, Feb. 2015.

[17] C. Chen, H. Bagan, and T. Yoshida, "Multiscale mapping of local climate zones in Tokyo using airborne LiDAR data, GIS vectors, and Sentinel-2 imagery," *GISci. Remote Sens.*, vol. 60, no. 1, Dec. 2023, Art. no. 2209970.

[18] K. Y. Fung, Z. - L. Yang, and D. Niyogi, "Improving the local climate zone classification with building height, imperviousness, and machine learning for urban models," *Comput.Urban Sci.*, vol. 2, no. 1, Dec. 2022.

[19] M. D. Islam, B. Li, C. Lee, and X. Wang, "Incorporating spatial information in machine learning: The Moran eigenvector spatial filter approach," *Trans. GIS*, vol. 26, no. 2, pp. 902–922, Apr. 2022.

[20] J. P. LeSage, "An Introduction to spatial econometrics," *Revue d'économie industrielle*, vol. 123, pp. 19–44, Sep. 2008.

[21] W. R. Tobler, "A computer movie simulating urban growth in the Detroit region," *Econ. Geography*, vol. 46, no. sup1, pp. 234–240, Jun. 1970.

[22] A. R. Phalke et al., "Mapping croplands of Europe, middle east, Russia, and central Asia using landsat, random forest, and Google earth engine," *ISPRS J. Photogramm. Remote Sens.*, vol. 167, pp. 104–122, Sep. 2020.

[23] Q. Jiang et al., "Mapping paddy rice planting area in dongting lake area combining time series Sentinel-1 and Sentinel-2 images," *Remote Sens.*, vol. 15, no. 11, Jan. 2023, Art. no. 11.

[24] A. R. Phalke and M. Özdoğan, "Large area cropland extent mapping with Landsat data and a generalized classifier," *Remote Sens. Environ.*, vol. 219, pp. 180–195, Dec. 2018.

[25] P. E. Osgouei, S. Kaya, E. Sertel, and U. Alganci, "Separating built-up areas from bare land in Mediterranean cities using sentinel-2a imagery," *Remote Sens.*, vol. 11, no. 3, Feb. 2019, Art. no. 345.

[26] A. Lefebvre, C. Sannier, and T. Corpetti, "Monitoring urban areas with sentinel-2a data: Application to the update of the copernicus high resolution layer imperviousness degree," *Remote Sens.*, vol. 8, no. 7, Jul. 2016.

[27] C. F. Brown et al., "Dynamic world, near real-time global 10 m land use land cover mapping," *Sci. Data*, vol. 9, no. 1, Jun. 2022.

[28] K. Clauss, M. Ottinger, and C. Kuenzer, "Mapping rice areas with sentinel-1 time series and superpixel segmentation," *Int. J. Remote Sens.*, vol. 39, no. 5, pp. 1399–1420, Mar. 2018, doi: 10.1080/01431161.2017.1404162.

[29] M. Gašparović and D. Dobrinić, "Comparative assessment of machine learning methods for urban vegetation mapping using multitemporal sentinel-1 imagery," *Remote Sens.*, vol. 12, no. 12, Jun. 2020, Art. no. 1952.

[30] Y. La, H. Bagan, and Y. Yamagata, "Urban land cover mapping under the local climate zone scheme using Sentinel-2 and PALSAR-2 data," *Urban Climate*, vol. 33, Sep. 2020, Art. no. 100661.

[31] A. Nurwanda and T. Honjo, "Analysis of land use change and expansion of surface urban heat island in bogor city by remote sensing," *ISPRS Int. J. Geo-Inf.*, vol. 7, no. 5, Apr. 2018, Art. no. 165.

[32] Z. Xu, J. Chen, J. Xia, P. Du, H. Zheng, and L. Gan, "Multisource earth observation data for land-cover classification using random forest," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 789–793, May 2018.

[33] R. Dubayah, H. Tang, J. Armston, S. Luthcke, M. Hofton, and J. Blair, "GEDI L2B canopy cover and vertical profile metrics data global footprint level V002," in *NASA EOSDIS Land Processes Distributed Active Archive Center*. Washington, DC, USA: NASA, 2021.

[34] C. D. Elvidge, M. Zhizhin, T. Ghosh, F.-C. Hsu, and J. Taneja, "Annual time series of global VIIRS nighttime lights derived from monthly averages: 2012 to 2019," *Remote Sens.*, vol. 13, no. 5, Mar. 2021, Art. no. 922.

[35] B. Yu et al., "Urban built-up area extraction from log- transformed NPP-VIIRS nighttime light composite data," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 8, pp. 1279–1283, Aug. 2018, doi: 10.1109/LGRS.2018.2830797.

[36] A. R. As-syakur, I. W. S. Adnyana, I. W. Arthana, and I. W. Nuarsa, "Enhanced built-up and bareness index (EBBI) for mapping built-up and bare land in an urban area," *Remote Sens.*, vol. 4, no. 10, Oct. 2012.

Md Didarul Islam received the M.S. degree in geographic information system from Central Michigan University, Mt Pleasant, MI, USA, in 2021. and the Ph.D. degree in Earth systems and geoinformation sciences from George Mason University, Fairfax, VA, USA, in 2024.

He is currently with the Center for Spatial Information Science and Systems, George Mason University. He is currently a Research Assistant with the Center for Spatial Information Science and Systems, George Mason University. His research interests include machine learning, spatial statistics, urban science, remote sensing, and geographic information systems.

Liping Di (Senior Member, IEEE) received the Ph.D. degree in remote sensing/GIS (geography) from the University of Nebraska–Lincoln, Lincoln, NE, USA, in 1991.

He is currently with the Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, USA. He is currently a Professor and the Founding Director of the Center for Spatial Information Science and Systems and a Professor with the Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA, USA. He has engaged in geoinformatics and remote sensing research for more than 30 years, authored or coauthored more than 550 publications, and received research grants of more than $70 million.

Dr. Di has actively participated in the activities of a number of professional societies and international organizations, such as IEEE GRSS, ISPRS, CEOS, ISO TC 211, OGC, INCITS, and GEO. He chaired INCITS/L1, a U.S. national committee responsible for setting U.S. national standards on geographic information and representing USA at ISO Technical Committee 211 (ISO TC 211) from 2010 to 2016, and was the Convenor of ISO TC 211 Working Group 9 from 2019 to 2022. He was elected as the Inaugural President of the International Society of Agromatics in 2022. He served as the Co-Chair for the Data Archiving and Distribution Technical Committee (DAD TC) of IEEE GRSS from 2002 to 2005 and the Chair for DAD TC from 2005 to 2009. He was one of the founders of the Earth Science Informatics Technical Committee of IEEE GRSS. In 2012, he founded the annual International Conference on Agro-geoinformatics, an IEEE GRSS-sponsored specialty conference, and has served as the General Conference Chair since 2012. He is also the General Chair of IGARSS 2026. He was the recipient of many awards for his work, including the R&D100 Award from R&D Magazine in 2008, the Secretary's Honor Award from USDA in 2011, the Merit Award from the InterNational Committee for Information Technology Standards in 2016, and the Lifetime Achievement Award from INCITS in 2022.

Chen Zhang (Member, IEEE) received the Ph.D. degree in earth systems and geoinformation sciences from George Mason University, Fairfax, VA, USA, in 2021.

He is currently a Research Assistant Professor with the Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, USA. He has authored or coauthored more than 80 peer-reviewed journal articles, book chapters, conference papers, and engineering reports. He leads the development of many geospatial data service systems to serve the geosciences and remote sensing community. His research interests include geographic information science and systems, remote sensing, agro-geoinformatics, geospatial information interoperability and standards, and land use and land cover change.

Ruixin Yang is currently with Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, USA. He is currently an Associate Professor with the Department of Geography and Geoinformation Science, College of Science, George Mason University, Fairfax, VA, USA. His research areas range from fluid dynamics, to astrophysics and general relativity, to data sciences, and then to data information systems, data analysis, and earth systems science. In recent years, he led a software developer team in developing several prototypes for Geoinformation systems for supporting online data search, access, and analysis. He has been the PI for several projects sponsored by the USDA on commodity data processing and integration, farm income and wealth data and model support, dissemination of global food security and model, etc. He has helped to organize the past Agro-Geoinformatics conferences and was co-PI for NIFA support of Agro-Geoinformatics 2017, 2018. His current research interest includes Earth system science, and advanced data analysis techniques, data mining for intensity prediction of tropical cyclones, and governmental agriculture data integration and analysis.

John Qu received the Ph.D. degree in remote sensing from Colorado State University, Fort Collins, CO, USA, in 1997.

He is currently with Global Environment and Natural Resources Institute, George Mason University, Fairfax, VA, USA. He is currently a Professor with the Department of Geography and GeoInformation Science, College of Science, George Mason University. He is the Founder and the Director of The Global Environment and Natural Resources Institute and Environmental Science and Technology Center, GMU. His major research areas are satellite remote sensing, atmospheric radiation transfer, environment and climate sciences, fire sciences, sustainable energy and geographical information system applications.

Daniel Tong is currently with the Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, USA. He is currently a Research Professor with the Center for Spatial Science and Systems, a research unit affiliated with the College of Science, George Mason University. He is interested in improvements to the National Air Quality Forecast Capability.

Liying Guo (Member, IEEE) is currently with the Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, USA. He is currently a Research Professor with the Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, USA. She has engaged in geospatial information research for more than 25 years. Her research interests include geospatial science applications, food security, ecosystem sustainability, public health, climate change, and land use and land cover change. She has actively participated in the activities of a number of IEEE GRSS, ISO TC211, INCITS, and OGC, and served as the PI and the Co-PI on multiple federal and national research grants. She is the Secretary of the International Society of Agromatics (ISAM), the Co-Chair of the Organizing Committee of the annual International Conference on Agro-Geoinfomatics, and the Manager of the NSF-funded multi-institutional and multidisciplinary CropSmart project.

Li Lin (Member, IEEE) received the Ph.D. degree in earth systems and geoinformation sciences from George Mason University, Fairfax, VA, USA, in 2022.

He is currently with the Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, USA. He is currently a Research Assistant Professor with the Center for Spatial Information Science and Systems, George Mason University. He has authored or coauthored more than 50 peer-reviewed journal articles, book chapters, conference papers, and engineering reports. He leads the development of many geospatial data service systems to serve the geosciences and remote sensing community. His research interests include agro-geoinformatics, remote sensing, urban development, geospatial metadata, and catalog federation.

Aran Pandey is currently a Senior with the Skyline High School, Sammamish, WA, USA.