

# Automated Method for Artificial Impervious Surface Area Mapping in Temperate, Tropical, and Arid Environments Using Hyperlocal Training Data With Sentinel-2 Imagery

Kristofer Lasko  and Francis D. O'Neill

**Abstract**—This study presents an automated methodology to generate training data to map up-to-date artificial impervious surface (AIS) extent maps using two dates (winter and nonwinter) of a Sentinel-2 granule across six international sites (Egypt, India, Qatar, U.K., Eastern USA, and Western USA). It uses a series of spectral, textural, and distance decision functions combined with an outdated AIS layer to create nontarget and target binary masks from which to generate a balanced set of training data applied to a random forest classifier. Two outdated global AIS layers (GMIS-2010 and GISA-2016) were evaluated within the framework to create AIS maps from more recent years (e.g., 2020). For the decision functions, stepwise threshold adjustments applied to normalized difference vegetation index (NDVI) and Euclidean distance layers were evaluated on the binary masks (low-density AIS, high-density AIS, and nontarget land covers) with 729 permutations and 115 permutations for global impervious surface area (GISA) and global manmade impervious surface (GMIS), respectively. The optimal thresholds were determined globally (all six scenes), individually (scene) and grouped by climate for adaptive thresholds. The accuracy assessment found both GMIS-output and GISA-output with global thresholds can accurately map current AIS with 86.9% ( $\pm 1.7\%$ ) (GISA) and 82.7% ( $\pm 2.3\%$ ) (GMIS) accuracy. Adaptive climate thresholds yielded slightly higher accuracies for temperate, tropics, and arid scenes. A novel beach bare ground sampling mask and annual NDVI standard deviation were also evaluated for performance and improved the accuracy in 5/6 sites. Lastly, the global GISA output was compared with a manually labeled deep learning model (Esri) with slightly lower overall accuracy (86.9% vs 88.6%).

**Index Terms**—Built-up, global impervious surface area (GISA), global manmade impervious surface (GMIS), image classification, impervious surface, machine learning, random forest, Sentinel-2, spectral indexes.

## I. INTRODUCTION

**T**HE mapping of artificial impervious surfaces (AIS) is key for understanding a variety of environmental and social

Manuscript received 7 July 2023; revised 11 September 2023; accepted 24 October 2023. Date of publication 30 October 2023; date of current version 23 November 2023. This work was supported by the U.S. Army Corps of Engineers, Engineer Research and Development Center, Geospatial Research business area. (Kristofer Lasko and Francis D. O'Neill are co-first authors.) (Corresponding author: Kristofer Lasko.)

The authors are with the Geospatial Research Laboratory, Engineer Research and Development Center, Alexandria, VA 22315 USA (e-mail: Kristofer.D.Lasko@erdc.dren.mil; oneill@erdc.dren.mil).

Digital Object Identifier 10.1109/JSTARS.2023.3328301

phenomena. AIS includes manmade material that is mostly impervious to precipitation and is associated with human settlements [1]. Accurate mapping of AIS is critical for urban demographics and population growth estimates, impacts on land cover and land use changes, and more. AIS and associated urban land uses represent a small proportion of the global land area; however, their effects are wide-reaching with impacts on hydrology due to increased run-off, air pollution emissions, and associated health impacts due to focused commerce and more [2], [3], [4], [5]. Accurate and up-to-date data on the extent and change of AIS are key for gaining a better understanding of the above impacts and broader issues such as climate change mitigation and adaptation.

There have been many efforts to map AIS extent and change using remotely sensed imagery. Studies have applied AIS spectral indexes or devised new ones. The built-up area extraction index, for example, is a band ratio consisting of SWIR-1, green, red, and an arithmetic constant from Landsat 8 imagery which provided better accuracy than earlier methods [6]. Other spectral indexes included the new built-up index, normalized-difference built-up index, band ratio for built-up area, normalized built-up area index, built-up land features extraction index, and VBISWIR1 among others [7], [8], [9], [10], [11], [12], [13], [14], [15]. Accuracies have been mixed and ranged from approximately 70% to 95% with manual thresholds created on a per-site basis. AIS is difficult to map across certain landscapes, especially bare ground due to similar spectral signatures [16]. AIS can also be confused with cropland, grassland, and wetlands [17], [18]. Generation of an algorithm to accurately map across varied geographies can introduce challenges related to high spectral complexity in central business districts attributed to tall buildings and shadows, as well as varied AIS construction styles and materials which result in significantly varied spectral signatures [19]. These include rocky material used for construction in arid and semi-arid regions resulting in misclassification between bare ground and AIS, as well as cropland confused for low-density AIS (villages and small towns) due to spectral similarity of vegetation mixed with high reflectance surfaces [20], [21]. In fact, some studies have creatively generated separate workflows for arid and nonarid regions to account for these difficulties [22], [23].

To overcome issues with similar spectral signatures researchers began using nighttime lights (NTL) imagery. Early studies leveraged Defense Meteorological Satellite Program (DMSP)- Operational Linescan System (OLS) NTL imagery for mapping AIS using a time series, considering that built-up areas will yield a persistent signal through time as compared with natural sources like fires [24], [25], [26]. The NTL datasets have increased in accuracy and resolution with the use of VIIRS [27]. One recent study used VIIRS NTL data and Landsat 8 with adaptive thresholding and automated training data yielding 93% accuracy [28]. Other studies have mapped megacities with VIIRS NTL time series yielding overall accuracies between 85% and 95% [29], [30], [31].

More recently, studies have used moderate resolution multispectral imagery such as Landsat 8 or Sentinel-2 for mapping AIS extent or change by leveraging grey-level co-occurrence matrixes (GLCMs), spectral unmixing, or image fusion [32], [33], [34]. Robust studies capturing AIS extent and change over multiple years have leveraged machine learning models in conjunction with temporal filtering and other heuristics to exclude false positive AIS found in a single date [22], [23]. Newer studies have gained the ability to better characterize human settlement structure by harnessing deep neural networks and manually or automatically selected training data. For example, one recent study used a U-NET based convolutional neural network (CNN) with manually drawn labels and 10 m SAR imagery to map AIS over a large portion of China with relatively strong accuracy ranging from 76% to 93% depending on the scene [35]. Unlike some earlier studies, they assessed accuracy across a wide range of challenging landscapes. Other studies have used various CNNs with moderate or fine resolution imagery and accuracies above 90%–95% [36], [37], [38]. However, some of these studies are geographically limited and the models may require further training and tedious manual data collection for broader application. To alleviate this issue of a lack of training data, models can be pretrained in the large ancillary dataset, and then finetuned at the local scale with limited labels; furthermore, reference data can be generated based on center points derived from CenterNet and guided by transfer learning [39], [40].

Several studies have harnessed ancillary training datasets with multispectral, SAR, or a combination of imagery. This approach does not require tedious training data collection and can result in models that are applicable to broad regions. Two recent studies harnessed OpenStreetMaps to generate training data for supervised classification on Landsat 8 with accuracy as high as 91% in tropical Brazil [41], or across multiple international human settlements using Sentinel-2 with F1-scores averaging from 0.75 to 0.94 [42]. Some studies have used NTL with normalized difference vegetation index (NDVI) time series composites for accurate training data selection and supervised classification with accuracy above 91% across varied regions [43], [44]. Mapping of rural–urban gradients in Europe was done using automatically created training data from spectral indexes [45].

To overcome cloud cover, some studies have leveraged an SAR time series. One study used C-band ENVISAT ASAR and Sentinel-1 [46]. Another study demonstrated that operational built-up area mapping was possible across urban and rural

areas at 30 m scale with C-band SAR, GLCM texture, and other layers integrated into a supervised classifier [47]. Texture metrics can capture spatial heterogeneity of human settlements and improve accuracy as characterized in a recent study with one-class random forest [48]. Other studies have successfully mapped AIS using Sentinel-1 backscatter [49], [50] or interferometric coherence [51], [52]. Lately, studies have leveraged the joint use of ALOS PALSAR, Sentinel-1 SAR and/or Sentinel-2 multispectral imagery and applying unsupervised clustering, supervised classification, or deep learning [53], [54], [55], [56], [57].

Global AIS datasets have been produced from sensors such as Landsat 8 or Sentinel-1. For example, a 2015 AIS was produced using multispectral imagery [58], and a very accurate operational 30 m dataset has also been made recently available [59], [60]. Lastly, a Landsat-based 30 m dataset was created with moderately high accuracy [61]. These datasets were generated based on automated or semi-automated supervised classification schemes with multisource satellite imagery and ancillary datasets.

Using AIS data layers from prior dates as training data can save significant time to a user in the data collection process, further, many of the summarized studies use dense time series imagery from an entire year which also can be intensive without cloud computing resources. While previous studies have leveraged ancillary datasets, they typically did not leverage datasets that were more than ten years outdated (e.g., GMIS produced in 2010). In this study, we create a framework to automatically map current AIS extent by leveraging just two dates of Sentinel-2 20 m multispectral imagery (winter and nonwinter) with ancillary datasets including digital elevation models (DEMs) and a globally produced AIS layer from prior dates considering that AIS does not typically revert to other land cover classes [62]. Logical decision functions based on spectral indexes, band reflectance, Euclidean distance rules, and texture are created to determine possible AIS training data pixels and possible nontarget training. The training data are selected automatically and there is no required manual input in the framework. The study seeks to quantify how accurately this automated method can map AIS and evaluates the use of two different AIS layers from different time periods. The approach uses balanced training data selection and subsequent image classification using hyperlocal training data fed into a random forest classifier. Furthermore, the thresholds for the logical decision functions are evaluated and optimized using a stepwise evaluation procedure grouped by different climate regions. The AIS mapping algorithm is assessed across six different Sentinel-2 scenes representing varying degrees of urbanization with challenging terrain characteristics including arid environments, agricultural environments with fallow fields.

The novelty of the proposed research and its contribution is described as follows.

- 1) Evaluation of AIS workflow incorporating 30 m global manmade impervious surface (GMIS) AIS dataset from more than ten years prior (2010) against recent Sentinel-2 imagery containing 20 m imagery with different pixel alignment.
- 2) Evaluation of the same AIS workflow incorporating more recent Sentinel-2 based AIS dataset from 2016 (GISA)

and how the output differs when using the GMIS AIS from 2010.

- 3) Creation of a workflow using only two dates of imagery within a year (winter and non-winter image) with shallow machine learning classifier (random forest) as compared against an Esri-hosted land cover layer based on manually labeled training data with a dense time series and CNN method.
- 4) Evaluation of a novel beach bare ground sampling method and time series NDVI standard deviation included in the workflow.
- 5) Use of automated sampling scheme based on the outdated AIS layers and logical decision functions optimized by stepwise threshold testing.
- 6) Creation of adaptive thresholds based on climate region types and determination of impact on accuracy.

## II. OBJECTIVES, STUDY AREA, AND DATASETS

### A. Objectives

While supervised classification with manual training data collection is common, it can be tedious. Previous studies have produced accurate global AIS datasets, but these data layers eventually become outdated over time. For example, GMIS was generated on imagery from 2010 and global impervious surface area (GISA) was generated on 2016 imagery, and while accurate for their respective years, they have not been updated to a more recent year. The primary objectives of this study are as follows.

- 1) To create a framework based on logical decision functions and two existing AIS datasets from prior years (2010 and 2016) with supervised classification for automated mapping of AIS from recent moderate resolution satellite imagery in order to obtain accurate, timely, and current spatial extent of AIS.
- 2) Evaluate different stepwise thresholds on the logical decision functions (NDVI thresholds, Euclidean distance thresholds) to map AIS and compare the accuracy of thresholds across climate-based groups via adaptive thresholding.
- 3) Create a novel beach bare ground sampling method to reduce commission errors on bare ground adjacent to water, create and apply an NDVI texture metric, and include an annual NDVI standard deviation composite and evaluate the impact that these inputs have on accuracy.
- 4) Compare the accuracy of this framework to a more advanced and data intensive deep learning model maintained by Esri.

### B. Study Area

For evaluation and testing, we selected six unique Sentinel-2 granules spanning multiple ecological and climate zones. The first site covers granule ID T36QVM featuring Aswan, a moderately-populated city of 1.5 million located along the Nile River in the arid tropics of Egypt. It includes seasonal water, wetlands, croplands, and large swaths of bare ground, exposed rock, and varied topography yielding a very challenging site

for mapping AIS. The second site covers granule ID T10TDP, which includes Eugene, Oregon USA, a town of low population (<175K inhabitants) located adjacent to large swaths of forestry. The site also includes the Willamette Valley with agricultural areas which can be erroneously mapped as AIS. This granule lies in a Marine West Coast climate zone with Mediterranean characteristics. The third site covers granule ID T18TYM, featuring Providence, USA (approximately 200k inhabitants) and more than ten small towns with populations less than 100k. It is a humid continental climate and includes forest, mixed land use, and suburban development. The fourth site covers granule ID T30UXB and features Portsmouth, U.K., which is the most densely populated U.K. city with a population of approximately 250k. It also includes the port city of Southampton. The site includes a large coastal area, along with small tracts of forest, wetlands, rivers, and agriculture including field crops such as wheat, grasses, and fallow fields. The fifth site covers granule ID T45QXE, featuring the megacity of Kolkata, India, which has a metropolitan population in excess of 13 million and rapid peri-urban expansion. It has a tropical wet-and-dry climate with monsoonal influence. It also includes portions of the Sundarbans National Park and agricultural activity, including Rice fields and shrimp farms. It contains dozens of small towns and villages within the state of West Bengal, many with populations under 5k inhabitants, and a small portion of Bangladesh (including the town of Basantapur). The sixth site cover granule ID T39RWJ, featuring Doha, the capital city of Qatar, whose population exceeds 2 million. It features peri-urban expansion, coastal areas along the Persian Gulf, and zones of irrigated agriculture. The climate is hot desert with minimal precipitation, and the scene contains large swaths of sand.

### C. Datasets

1) *Sentinel-2 MSI*: The Sentinel-2 multispectral instrument (MSI) imagery is provided by the European Space Agency [63]. We acquired the bottom of atmosphere reflectance, Level 2A (L2A), from the Copernicus Open Access Hub. The L2A product contains geometrically- and radiometrically corrected imagery, along with a quality assurance layer called the scene classification layer (SCL) that contains information on clouds, cloud shadows, snow, water, and ice, which we use to mask the input imagery and obtain a snow mask. Sentinel-2 MSI imagery includes 13 spectral bands at three spatial resolutions (10, 20, and 60 m). For this study, we leveraged the 20 m dataset to utilize the two shortwave-infrared (SWIR) bands and red-edge bands. For each site, a winter season and nonwinter season image was acquired. The processing algorithm for Sentinel-2 L2A imagery applied through Sen2cor underwent changes for imagery acquired in late 2021 and after. The most notable difference is that, post-2021, L2A surface reflectance values have an offset value of 1000 added to them. Note that all spectral band values reported in the methods have had this offset removed (where applicable), and reported values correspond to pre late 2021 ranges.

Spectral indexes such as the NDVI, modified normalized-difference water index (MNDWI) were derived from the



TABLE I  
SENTINEL-2 IMAGERY USED IN THIS STUDY

| Site Name                       | Sentinel-2 Granule ID | Imagery Date (Winter) | Imagery Date (Summer) |
|---------------------------------|-----------------------|-----------------------|-----------------------|
| Aswan, Egypt (EG)               | T36QVM                | 2021_12_26            | 2021_08_18            |
| Kolkata, India (IN)             | T45QXE                | 2018_02_04            | 2018_10_22            |
| Doha, Qatar (QA)                | T39RWJ                | 2022_03_10            | 2022_09_11            |
| Portsmouth, United Kingdom (UK) | T30UXB                | 2019_11_18            | 2020_07_30            |
| Eugene, USA (US-OG)             | T10TDP                | 2020_12_28            | 2020_08_28            |
| Providence, USA (US-MA)         | T18TYM                | 2020_12_03            | 2020_05_27            |

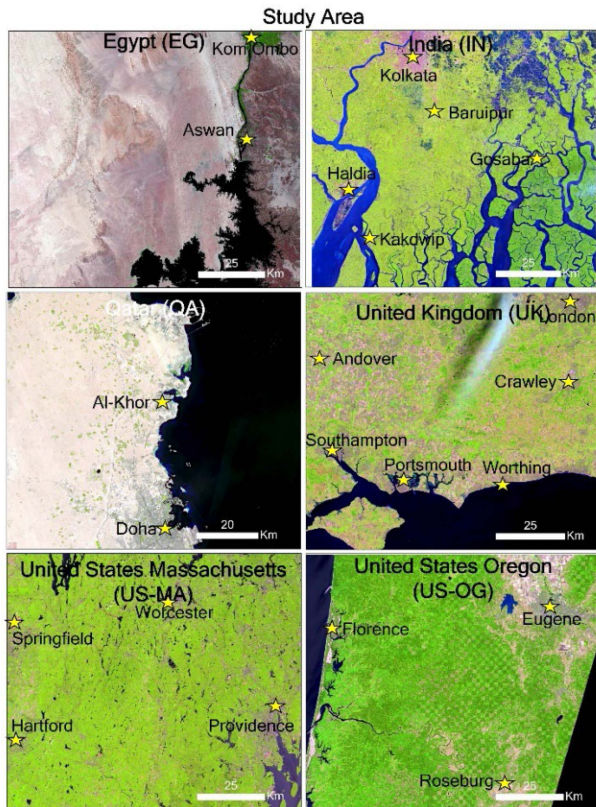


Fig. 1. Sentinel-2 imagery of the six study sites.

Sentinel-2 imagery. Table I shows a list of input imagery used for each site and Fig. 1 visualizes the study sites.

2) *SRTM Digital Elevation Model*: DEM data from the shuttle radar topography mission (SRTM) were acquired for the one scene with significant topographic relief: Aswan, Egypt (Sentinel-2 granule ID T36QVM), for the purpose of correcting false positive AIS in hilly terrain. The SRTM DEM is provided at 1 arcsec (30 m) with void-filled data. The dataset was downloaded from the USGS EarthExplorer website. As with most DEMs, the SRTM DEM suffers some artifacts due to terrain shadow, vegetation canopy, or other dense obstructions [64]. The SRTM DEM was downsampled to 20 m resolution using nearest neighbor resampling and aligned to the Sentinel-2 imagery.

AIS training pixels with slope greater than  $18^\circ$  are converted to bare ground training. Any remaining AIS training pixels with slope greater than  $5^\circ$  are masked out.

3) *Landsat 8 Global Manmade Impervious Surface (GMIS) Layer*: The GMIS layer was previously generated for the year 2010 from 30 m Landsat imagery and was downloaded from the Socioeconomic Data and Applications Center of NASA [61]. It was derived from Global Land Survey imagery, which contains both Landsat 5 TM and Landsat 7 ETM data. The distributed layer includes a “percent impervious” variable, ranging from 0 to 100 indicating the percentage of impervious surface within a given pixel. Using GDAL, we downsampled the GMIS layer to 20 m resolution using nearest neighbor and ensured pixels were properly aligned considering the discrepancy between this derived Landsat dataset and the Sentinel-2 imagery it will be applied upon. We apply logical decision functions to increase the accuracy of the training data selection for our model as described in the methodology section.

4) *Global Impervious Surface Area (GISA)*: The GISA layer was generated for the year 2016 using a full time series of imagery from Sentinel-2 at 10 m and Sentinel-1 SAR at 10 m resolution [59]. The dataset is a binary layer produced at a global scale with an overall accuracy of about 86%. It was generated from a previous 30 m Landsat-based layer [60], as well as training data from Open Street Maps and other sources combined with a complex set of spectral, spatial, temporal, and geometric rules.

5) *Landsat 8 Annual NDVI Composites*: The Landsat 8 Level 2, Collection 2 dataset (surface reflectance) was used to create annual composites of the NDVI. One year of cloud-free pixels were used to obtain the standard deviation of NDVI for each 30 m pixel using Google Earth Engine. The 30 m pixels were resampled to 20 m scale and used for improving separation of AIS and non-AIS pixels. The AIS extent mapping framework was evaluated both with and without this layer.

6) *Sentinel-2 10 m Land Use/Land Cover*: The Sentinel-2 10 m land use/land cover product was produced on an annual basis for each year of 2017–2022. It was developed using the UNet CNN model trained on millions of manually labeled points across the world [65]. Nine land cover classes were produced. For this study, we used only the “built area” class to correspond with AIS and remapped the remaining values to zero. The dataset was resampled to 20 m pixels (majority reclassification scheme) to correspond with our Sentinel-2 20 m imagery. We acquired the year of imagery corresponding to the date of our Sentinel-2 imagery for each study area. In this study we refer to this dataset as Esri Land Cover because it is freely available for download on their website (<https://livingatlas.arcgis.com/landcover/>).

### III. METHODOLOGY

Some current methods for AIS mapping rely on an imagery time series, computationally intensive deep learning, or manually collected training data with supervised classification. The method proposed in this study is designed to be efficient and automated while leveraging the results of previous studies and using only two dates of multispectral imagery within a single year. While previous studies have produced accurate AIS extent maps, these maps are sometimes produced via regeneration of new manually labeled training points which can be a tedious



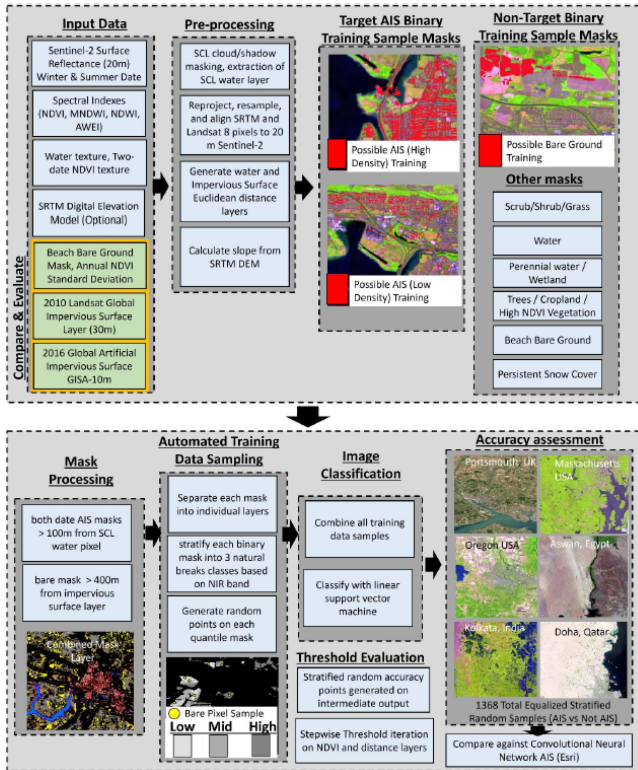


Fig. 2. Flowchart of the methodology evaluated in this study.

process. The framework proposed in this study does not require manual labeling, which would enable AIS maps to be generated for any given proceeding year via an easily reproducible workflow. This is in contrast to the currently publicly available AIS datasets which are generated only for a previous year and not regularly updated (e.g., GMIS in 2010, and GISA in 2016). A high-level overview of the proposed workflow is shown in Fig. 2.

Accordingly, this approach uses two dates (winter and non-winter) of Sentinel-2 L2A imagery to capture vegetation phenology and intra-annual dynamics, but without the intensity of a full time series. From this imagery, spectral indexes, texture, etc., are generated for each date and included in the feature set. An intermediate set of AIS extent maps is then generated separately using GISA and GMIS as inputs combined with logical decision functions and a sampling scheme which selects training points for the random forest classifier for various target (low-density AIS, high-density AIS) and nontarget land covers (e.g., wetlands, water, etc.). After the intermediate maps are generated, a stepwise threshold method is used to determine optimal NDVI and Euclidean distance thresholds for each scene and dataset. Pixel frequency count maps are generated to demonstrate how these thresholds impact the output. The differences between GISA (Sentinel-2 based AIS from 2016) and GMIS (Landsat 8 based AIS from 2010) are also evaluated. The most accurate threshold for each scene is then recorded as well as the most accurate threshold across all scenes which is referred to as the global threshold. Climate region adaptive thresholds are also generated where the study sites are grouped into tropics, temperate, arid, and nonarid. The study also evaluates the

TABLE II  
BINARY LAND COVER MASK CREATION FOR SUBSEQUENT TRAINING DATA SAMPLING

| Mask                  | Impervious Layer                     | MID-Infrared (MIR) Reflectance | NDVI | NDVI Texture | NDWI | MNDWI | AWFI | Distance to Water |
|-----------------------|--------------------------------------|--------------------------------|------|--------------|------|-------|------|-------------------|
| <i>Target</i>         |                                      |                                |      |              |      |       |      |                   |
| Built-up (Dense)      | ✓                                    |                                | ✓    | ✓            |      |       |      | ✓                 |
| Built-up (Light)      | ✓                                    |                                | ✓    | ✓            |      |       |      | ✓                 |
| <i>Non-Target</i>     |                                      |                                |      |              |      |       |      |                   |
| Water (Persistent)    | ✓ (distance)                         | ✓                              |      |              | ✓    | ✓     | ✓    |                   |
| Water (Temporary)     | ✓ (distance)                         | ✓                              |      |              | ✓    | ✓     | ✓    |                   |
| Wetland               | ✓ (distance)                         |                                | ✓    |              |      | ✓     |      | ✓                 |
| Vegetation (Light)    | ✓ (distance)                         | ✓                              | ✓    |              |      |       |      |                   |
| Vegetation (Vigorous) | ✓ (distance)                         |                                | ✓    |              |      |       |      |                   |
| Bare Ground           | ✓ (distance)                         | ✓                              | ✓    |              |      |       |      | ✓                 |
| Snow                  | <i>Internal Sentinel-2 Snow Mask</i> |                                |      |              |      |       |      |                   |

accuracy of a novel beach bare ground binary mask and an annual time series NDVI standard deviation composite. Lastly, the AIS maps generated from this study are evaluated against a robust manually trained deep learning based global land cover layer hosted by Esri. The accuracy assessment was guided by a robust method using 228 equalized random points per scene as described subsequently in more detail.

#### A. Ancillary Data Generation

An NDVI texture metric is created from the two dates of Sentinel-2 NDVI. It is calculated using the standard deviation of the absolute difference of the leaf-off and leaf-on NDVI images with a 3x3 pixel window (1). We expect AIS will have higher standard deviation than other land cover types because AIS NDVI often varies significantly within a landscape due to varying vegetation levels, mixed pixels, different surface and structure types (e.g., pavement, flat roofs, sloped roofs, etc.) which yield significant spectral heterogeneity evident in the texture layer. Other studies have had success with different types of multitemporal texture layers [48].

A shoreline water texture layer was also created based on the Sentinel-2 SCL water mask: water-flagged pixels are assigned a value of 1, while nonwater pixels are reclassified to 0. The texture layer is a convolved sum of these pixels within a 5x5 pixel neighborhood. Pixels surrounded by water would have a value of 25 and pixels adjacent to, but not surrounded by water would have lower values. This layer was designed to assist the image classification to reduce erroneous AIS on shoreline or water areas

$$\text{texture}_{\text{ndvi}} = \text{SD}_{\text{kernel} = 3} (|\text{ndvi}_{\text{summer}} - \text{ndvi}_{\text{winter}}|). \quad (1)$$

#### B. Binary Mask Layer Generation

After the input data are processed, binary raster masks are created for target and nontarget land cover types using the input data including spectral indexes, texture layer, GMIS or GISA layer, and NDVI annual composite layer as shown in Table II.

These masks are used only for generating training data for the target and nontarget land cover types used to fit the supervised classifier. The following masks are created: low-density AIS, high-density AIS, bare ground, beach bare ground, low NDVI vegetation, high NDVI vegetation, persistent water, snow, and wetland and perennial water. The two AIS masks (low density and high density) are designed to have minimal erroneous pixels due to commission errors, while omitting some pixels which may be borderline or have uncertain land cover type.

The masks are designed to capture representative training points from each class, not to capture all possible pixels for each class, which will be the task for the classifier. Subsequently, the supervised classifier would determine the appropriate class for such pixels based on the combination of training data.

### C. Intermediate AIS Binary Mask Generation

The preceding sections describe the creation of each of the binary masks. The procedure is run with different AIS masks based on the GISA and GMIS datasets. For generation of the AIS mask with GMIS, the 2010 GMIS layer containing percent impervious pixels (0%–100%) from 30 m Landsat imagery is used. Two separate masks are created prior to supervised classification and combined postclassification: 1) low-density AIS, and 2) high-density AIS. Intermediate threshold values were assigned for each one for subsequent optimization after accuracy points are generated, the high-density AIS contains GMIS pixels with 50% or greater impervious surface, while low-density AIS contains GMIS pixels between 25% and 50% impervious surface. Pixels with less than 10% impervious surface are excluded from the mask due to errors in the dataset for some scenes. Moreover, for data quality purposes and to generate a mask with minimal erroneous pixels, an NDVI threshold maximum of 0.40 is applied to the low-density AIS training points, and 0.30 maximum is applied to the high-density AIS training points. The two-date NDVI texture layer has a threshold minimum of 3.5 to reduce false errors on bare ground, which typically has a more homogenous spatial texture than AIS. The annual time series NDVI standard deviation composite layer was used to mask out training pixels with annual standard deviation above the 95th percentile. The logic behind this is that AIS pixels have significantly less spectral variation through time than other land cover classes such as wetlands, forest, or cropland which vary seasonally. The use and removal of this layer is evaluated in the study as well.

The intermediate GISA-based AIS mask is created through a different procedure because it only contains two values: 0 for non-AIS and 1 for AIS. In this manner, the initial, nonoptimized AIS mask is generated for low-density and high-density AIS. The low-density AIS contains AIS pixels with  $NDVI < 0.80$  and  $> -0.50$  for winter ( $< 0.75$  and  $> -0.50$  for nonwinter), while the high-density AIS mask contains AIS pixels with  $NDVI < 0.35$   $> -0.50$  for winter ( $< 0.60$  and  $> -0.50$  for nonwinter).

### D. Nontarget Binary Training Data Mask Generation

The bare ground/sparse vegetation binary mask is created by the following logical decision functions with threshold

values guided by the literature [66], [67], [68] and through data exploration: 1) winter NDVI must be  $\geq 0$  and  $\leq 0.35$ , non-winter NDVI must be  $\geq -0.20$ , 2) with a minimum SWIR-1 reflectance of 600 to exclude possible wetland and water pixels. To reduce confusion with AIS, the bare area pixels must be at least 600 m from pixels in the AIS training layer.

A similar binary mask was generated for representing bare ground on beaches which are commonly incorrectly classified as AIS. This mask has the same thresholds as the bare ground mask except a Euclidean distance minimum of 40 m from AIS pixels and a maximum of 90 m Euclidean distance from surface water so that points are focused on this area.

Two low-moderate NDVI vegetation layers are generated; the first captures pixels which may have lower NDVI in winter, and the second captures pixels which may have lower NDVI in nonwinter. The first mask requires the following.

- 1) Winter NDVI between 0.10 and 0.70, and nonwinter NDVI between 0.35 and 0.70.
- 2) A minimum SWIR-1 reflectance of 800 to reduce confusion with wetland pixels.
- 3) A minimum Euclidean distance of 250 m from the AIS layer to avoid erroneous and mixed pixels.

The second layer requires the following.

- 1) Winter NDVI between 0.35 and 0.70, and nonwinter NDVI between 0.10 and 0.70.
- 2) A minimum SWIR-1 reflectance of 700 in winter and 1200 in nonwinter.
- 3) A minimum Euclidean distance of 250 m from the AIS layer.

Two high NDVI layers are also generated, with the first capturing pixels with low NDVI in winter, and second capturing pixels with low NDVI in nonwinter. The first mask requires the following: 1) winter NDVI between 0.10 and 0.7, and nonwinter NDVI greater than 0.7, 2) a minimum Euclidean distance of 200 m from the AIS layer. The second mask requires the following: 1) winter NDVI greater than 0.7, and nonwinter NDVI between 0.1 and 0.70, and 2) a minimum Euclidean distance of 200m from the AIS layer.

A wetland mask is generated based on the following logic.

- 1) Both dates of NDVI must be between  $-0.05$  and  $0.65$ .
- 2) Both dates of MNDWI must be between  $-0.20$  and  $0.60$ .
- 3) A minimum Euclidean distance of 200 m from the AIS layer.
- 4) A maximum Euclidean distance of 200 m from the Sentinel-2 SCL water layer from both dates.

The water mask was produced following a published method [69] and requires the following.

- 1) MNDWI, AWEL, and NDWI values greater than  $-0.03$ .
- 2) SWIR-1 reflectance below 1000.
- 3) A minimum Euclidean distance of 100 m from the AIS layer.

A second, perennial water layer was created which required only one date of imagery to meet all three of the previously listed requirements.

For any of the above layers, if a snow pixel is detected in the winter date of imagery, then only the nonwinter date criteria must be met. Lastly, a snow mask is created based on the provided

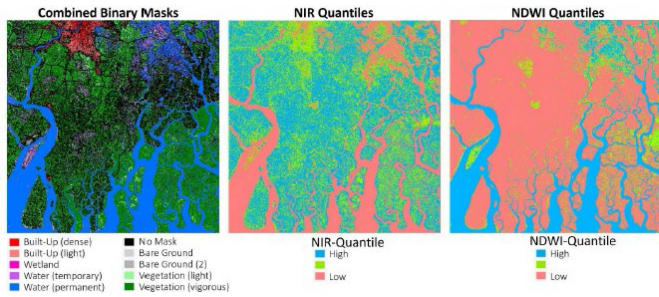


Fig. 3. Binary masks, NIR, and MNDWI quantile maps visualized. Spectrally balanced training samples from the binary masks are generated in conjunction with the quantile layers.

Sentinel-2 SCL snow mask if both dates of imagery contain snow.

Subsequently, training points are randomly selected from these binary masks and then used for classifying the imagery into AIS and not-AIS pixels.

#### E. Training Data Selection and Image Classification

After the target AIS and nontarget binary land cover masks are successfully created, then training data are selected from each of the masks to fit a supervised classifier. Each binary mask is converted into three quantiles based on the near-IR band with each group containing one-third of the pixels as shown in Fig. 3. The purpose of the quantiles is to ensure selection of a spectrally balanced set of training data. Subsequently, points are generated on each mask with equalized stratified random sampling across the three quantiles. The following numbers of points are generated for each class: persistent water (450), low NDVI vegetation (250), wetland (450), dense AIS (400), light AIS (400), bare ground (450), high NDVI vegetation (300), seasonal water (250), persistent snow (450), beach/coastal bare ground (50). Two training point datasets are created for each mask: one “no-snow” set, in which both the nonwinter and winter criteria must be met; and one “snow” set, in which the nonwinter criteria must be met and the winter scene must be labeled as snow by the SCL layer. Previous work has shown that accuracy is not substantially affected after a minimum training point count is achieved [70].

The training data are used to fit a random forest classifier containing the following parameters: 500 trees, max\_depth = 30, max\_samples = 500 as implemented in scikit-learn [71], [72]. The classifier then predicts the land cover type using the two dates of Sentinel-2 imagery, 2-date NDVI texture, shoreline water texture layers, spectral indexes, annual standard deviation of NDVI (results were evaluated with and without this layer), and optionally the DEM (if within a mountainous environment, only applied to EG scene). After image classification is complete, the nontarget classes are condensed into a single class with value “0,” and both AIS classes are condensed into a single AIS class with value “1.” The AIS classes are combined due to the potential for mistakes in the accuracy assessment when differentiating between the subjective low- and high-density AIS.

#### F. Accuracy Assessment and Stepwise Threshold Adjustment

After the intermediate set of AIS extent maps are created for each site, an accuracy assessment is conducted across each of the six study sites. AIS and non-AIS pixels are selected via equalized stratified random sampling with 228 points per scene (114 AIS, 114 non-AIS) and 1368 points total, following best practices in accuracy assessment and an estimated standard error of 0.15 for AIS and 0.05 for non-AIS [73], [74]. A 200 m minimum distance between points was specified to reduce spatial bias in the assessment. The ground-truth for each point was interpreted using a combination of high-resolution worldview imagery, the original Sentinel-2 imagery, and spectral indexes. The overall accuracies, user’s accuracies, producer’s accuracies, F1-scores, and adjusted overall accuracies are reported. A 95% confidence interval of uncertainty is computed for comparison purposes and the adjusted overall accuracy was computed based on error rates and the weighting from the proportion of land area for each class [73].

After ground-truth has been interpreted for the 1368 points across the six scenes, the points are then used for stepwise determination of optimal thresholds for the logical decision binary masks which could improve upon model accuracy during training sample selection. Several key variables were evaluated separately for GISA and GMIS.

For GMIS, the study first evaluated the minimum required Euclidean distance between bare ground training data pixels and the AIS layer’s pixels for both leaf-on and leaf-off imagery. Ranges from 0, 200, 400, and 600 m were evaluated. The second variable was the minimum Euclidean distance between AIS training pixels and water pixels, with distances including 0, 100, 200, 300, 400, and 500 m. The third variable evaluated was the lower bound value for the low-density AIS training data. That is, the minimum percent AIS from the Landsat 8 GMIS layer which would be considered as AIS training pixels. Similarly, the upper bound for low-density AIS, along with both bounds for high-density AIS, were also evaluated. These thresholds were used to inform the overall model for both data layers.

Unlike GMIS, the GISA layer is binary and lacks percentage AIS; therefore, different stepwise thresholds were analyzed. Specifically, the study evaluated different combinations of the minimum and maximum NDVI thresholds for both light AIS and dense AIS. This included maximum NDVI values for dense AIS of 0.15, 0.20, and 0.25 on both the leaf-on and leaf-off imagery, minimum NDVI values for light AIS of 0.20, 0.25, and 0.30 for both dates, and maximum NDVI values for light AIS of 0.45, 0.55, and 0.65 for both dates. This resulted in a total of 729 different permutations for evaluation on each scene. Ultimately, these stepwise comparisons are applied during training data selection and are designed to select a threshold which improves accuracy of the model.

The final AIS maps are created and evaluated based on the optimal thresholds determined from this analysis.

## IV. RESULTS

After the two sets of intermediate AIS extent maps were generated (GISA- and GMIS-based output) from the initial set



of thresholds, ground-truth points were created on each of the six sites as described in the methods. These 1368 equalized random sample points were then used to evaluate stepwise threshold changes as described in the methods.

#### A. Stepwise Thresholds: GMIS

The stepwise threshold adjustments were evaluated on 11 variables for the GMIS-trained output, as shown in Fig. 4(a) and (b). After evaluating the minimum Euclidean distance from impervious surface when creating the bare ground training mask, 4/6 sites yielded strong overall accuracy improvements when using values greater than 0 m, whereas 2 sites (US-OR and EG) yielded minimal change and decreased accuracy with thresholds greater than 0 m (respectively) for both winter and nonwinter imagery. The minimum Euclidean distance from water for AIS pixels found minimal accuracy variation for non-arid sites, but generally any threshold above 0 m performed best. For the two arid sites the accuracy was highest when the distance was reduced to 0 m. This is likely due to spatial location of many AIS pixels in the desert regions being located adjacent or near water pixels, more than the other sites. The dense AIS percentage minimum and maximum thresholds that were tested did not seem to have a noticeable impact on accuracy. Whereas the light density AIS minimum percent impervious in both winter and nonwinter resulted in values less than 45% yielding the highest accuracy, except for EG where 45% had the highest accuracy, the QA where 45% had similar accuracy to the other values.

The most accurate threshold combination for the GMIS-trained output based on overall accuracy was found to be the following.

- 1) 400 m minimum distance of bare ground training from GMIS AIS based on Euclidean distance.
- 2) 300 m minimum distance of water training from SCL water layer based on Euclidean distance.
- 3) 50%–90% impervious pixel area from GMIS layer for dense AIS class.
- 4) 15%–60% impervious pixel area from GMIS layer for light density AIS class.

#### B. Stepwise Threshold: GISA

Unlike GMIS (which contains percent impervious surface per pixel), GISA is a binary layer; therefore, other factors were evaluated to determine the impact on the overall accuracy. The optimal thresholds from GMIS for Euclidean distance of AIS training pixels from water, along with Euclidean distance of bare pixels from AIS, were transferred into the GISA-based analysis. The GISA-based analysis is shown in Fig. 5 and involved different NDVI threshold minimums and maximums for low- and high-density AIS layers. Specifically, the stepwise threshold evaluation found that NDVI maximums of 0.15, 0.2, and 0.25 in winter and nonwinter yielded minimal discernable differences in accuracy. The same is also true for the minimum NDVI for light density AIS. However, for the NDVI maximum on light density AIS, moderate variation in accuracy was observed when the winter NDVI threshold was changed for the three temperate climate sites of U.K., US-OR, and US-MA. For these three sites,

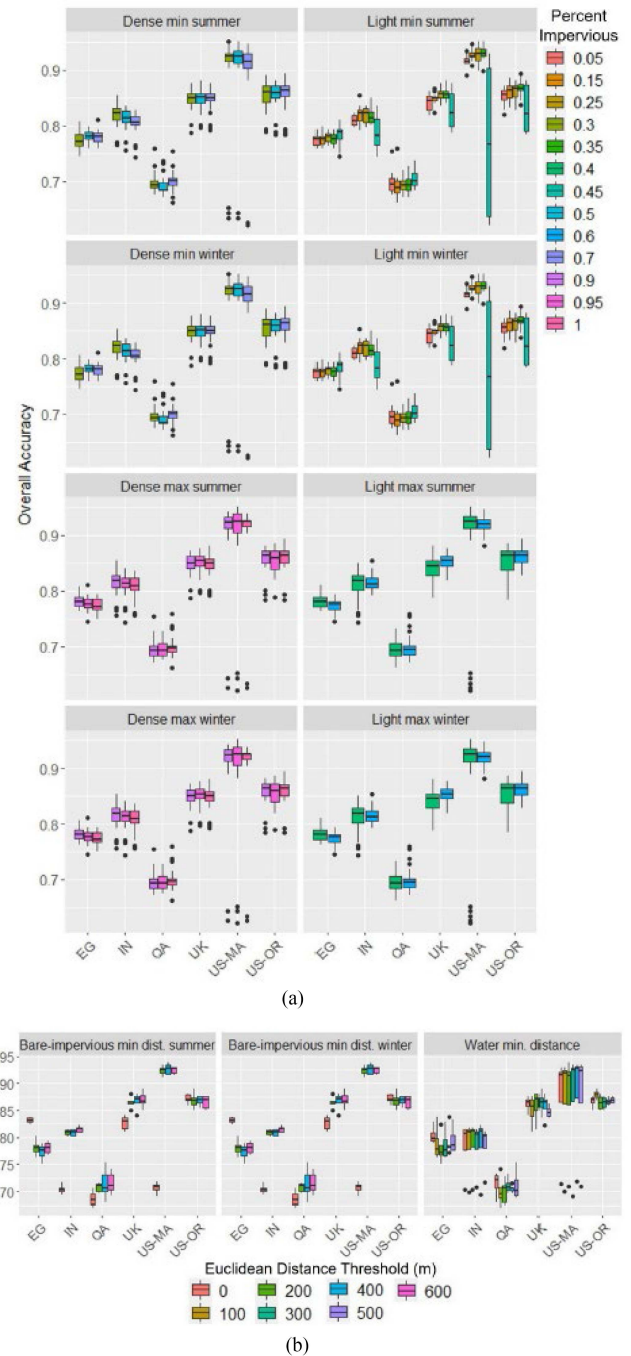


Fig. 4. (a) Percent impervious (0–100) thresholds evaluated for low-density AIS and high-density AIS while using the 2010 GMIS layer as AIS training. Variables include for example: Dense min. AIS summer which is the minimum AIS percentage from the GMIS layer for a pixel to be a candidate for the “Dense AIS” class. (b) Euclidean distance thresholds evaluated against the GMIS-trained output. The first facet evaluates the minimum distance of nontarget bare ground training samples from the GMIS AIS layer (to reduce false positive AIS).

a maximum NDVI threshold of 0.45 performed with the lowest accuracy. Unsurprisingly, less variation was observed in the two arid sites (EG and QA) and subtropics site (IN) as they contain less seasonal vegetation senescence.

The most accurate threshold combination for GISA-trained output based on overall accuracy was dense AIS winter NDVI

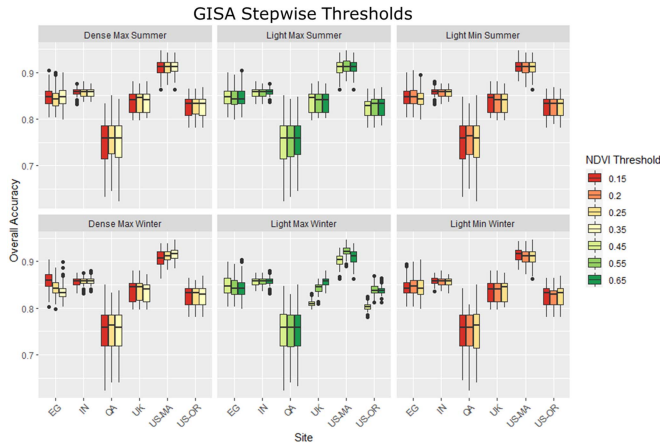


Fig. 5. Evaluation of different NDVI thresholds tested using GISA AIS as training data. Light density AIS and high-density AIS classes were evaluated with minimal variation in overall accuracy observed based on the changes.

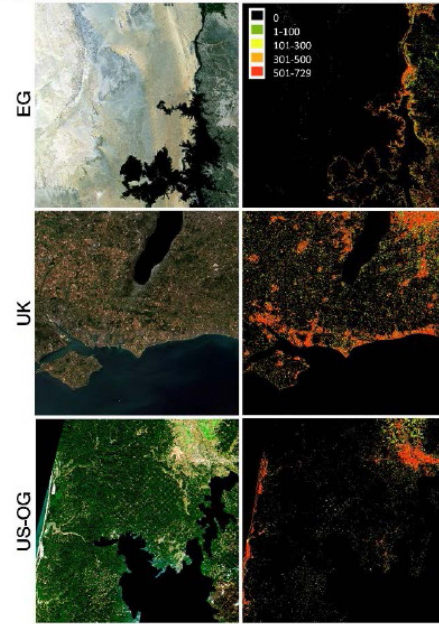
max (0.15), dense AIS nonwinter NDVI Max (0.25), light AIS winter minimum NDVI (0.15), light AIS nonwinter min NDVI (0.25), light AIS winter max NDVI (0.65), light AIS non-winter max NDVI (0.45).

### C. AIS Pixel Frequency Maps (Stepwise Thresholding)

The AIS pixel frequency count maps resulting from the 729 different threshold combinations applied to the GISA-trained output are shown in Fig. 6(a) and (b). They highlight that the threshold changes accounted for noticeable variation in AIS pixel labeling in some locations, but not all. The most variation was observed in the U.K. scene, with 13.9% of pixels yielding values between 1 and 728, indicating that altering the NDVI thresholds impacted whether the pixel was classified as AIS or not AIS. The India scene had 10.8% of pixels in that range, followed by US-MA (9.8%), QA (6.3%), US-OG (5.3%), and EG (4.0%). From the figure, it is apparent that much of the AIS pixel variation occurs in hilly bare terrain and shoreline regions along Nasser Lake in the EG scene. For the U.K. scene, much of the variation is found in suburban areas in the Northeast quadrant, for the US-OG scene most variation is found in agricultural areas in the northern portion of the scene, and for the remaining scenes the variation is more spatially varied.

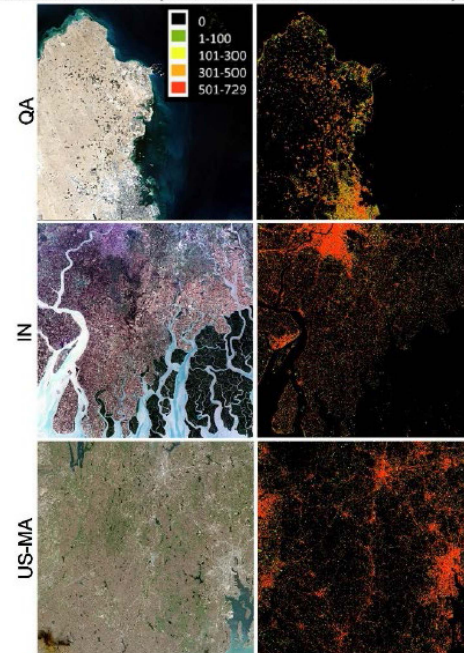
Analysis of the 90th percentile of the most accurate NDVI threshold combinations applied per site using the GISA-based framework revealed some interesting patterns. Frequency plots are shown in Fig. 7. For the two arid sites, a dense AIS winter image NDVI threshold of 0.15 was found in a majority of the most accurate outputs. Interestingly, when evaluating accuracy across all six sites combined, this 0.15 minimum threshold was also key suggesting that there is flexibility in the thresholds for the nonarid sites. For the nonarid sites and temperate sites, light density AIS winter NDVI threshold max of 0.65 was key for maintaining high accuracy. Importance of threshold values for the other variables was not as clear.

AIS Pixel Frequency Counts for stepwise threshold adjustments



(a)

AIS Pixel Frequency Counts for stepwise threshold adjustments



(b)

Fig. 6. (a), (b) Number of times a pixel was classified as AIS based on the 729 different NDVI threshold combinations evaluated using the GISA-based training layer.

### D. Overall Accuracy (Global Versus Scene GMIS Versus GISA Versus Esri)

After evaluating the 729 different threshold combinations based on the GISA-trained output, and the 115 different threshold combinations based on the GMIS-trained output, we determined the most accurate threshold values for each. Because



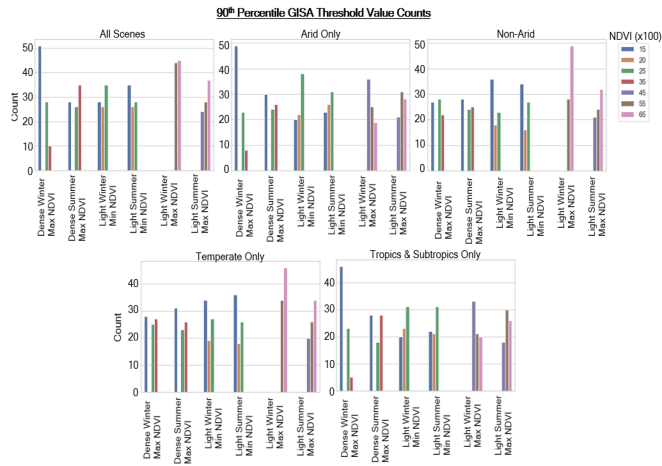


Fig. 7. Occurrence of different GISA-based NDVI threshold values in the 90th percentile of the most accurate threshold combinations. Plots are grouped by the respective climate regions, as well as all sites combined.

TABLE III  
ACCURACY METRICS FOR EACH MODEL EVALUATED

| Scene-Optimized Thresholds (Sen-2)    |                 |         |       |   | Non-Arid scene-optimized Thresholds (Sen-2) |       |                       |       |   |
|---------------------------------------|-----------------|---------|-------|---|---|-------|-----------------------|-------|---|
| REFERENCE                             |                 |         |       |   | REFERENCE                                   |       |                       |       |   |
| Class                                 | AIS             | Not AIS | Total | User's Accuracy                                 | Class                                       | AIS   | Not AIS               | Total | User's Accuracy                                 |
| AIS                                   | 385             | 109     | 490   | 78.1% ( $\pm 0.2\%$ )                           | AIS   | 294   | 97                    | 391   | 93.9% ( $\pm 0.2\%$ )                           |
| Not AIS                               | 45              | 830     | 873   | 94.9% ( $\pm 0.1\%$ )                           | Not AIS                                     | 43    | 518                   | 561   | 92.3% ( $\pm 0.1\%$ )                           |
| Total                                 | 428             | 935     | 1363  |   | Total                                       | 327   | 585                   | 912   |   |
| Producer's accuracy                   | 89.5%           | 88.2%   |       | F1 Score 83.4%                                  | Producer's accuracy                         | 89.9% | 88.9% ( $\pm 0.1\%$ ) |       | F1 Score 83.8%                                  |
|                                       | ( $\pm 0.2\%$ ) |         |       | Overall Accuracy 88.8%                          |   |       |                       |       | Overall Accuracy 87.9%                          |
|                                       |                 |         |       | Adjusted Overall Accuracy 93.8% ( $\pm 1.4\%$ ) |   |       |                       |       | Adjusted Overall Accuracy 91.4% ( $\pm 2.2\%$ ) |
| Globally-optimized Thresholds (Sen-2) |                 |         |       |   | Arid Optimized Thresholds (Sen-2)           |       |                       |       |   |
| REFERENCE                             |                 |         |       |   | REFERENCE                                   |       |                       |       |   |
| Class                                 | AIS             | Not AIS | Total | User's Accuracy                                 | Class                                       | AIS   | Not AIS               | Total | User's Accuracy                                 |
| AIS                                   | 363             | 111     | 474   | 76.6% ( $\pm 0.2\%$ )                           | AIS   | 88    | 59                    | 127   | 82.4% ( $\pm 0.2\%$ )                           |
| Not AIS                               | 68              | 826     | 894   | 92.4% ( $\pm 0.1\%$ )                           | Not AIS                                     | 5     | 294                   | 299   | 98.3% ( $\pm 0.2\%$ )                           |
| Total                                 | 431             | 937     | 1368  |   | Total                                       | 93    | 353                   | 446   |   |
| Producer's accuracy                   | 84.2%           | 88.2%   |       | F1 Score 80.2%                                  | Producer's accuracy                         | 95.1% | 89.3% ( $\pm 0.1\%$ ) |       | F1 Score 75.40%                                 |
|                                       | ( $\pm 0.2\%$ ) |         |       | Overall Accuracy 86.9%                          |   |       |                       |       | Overall Accuracy 86.0%                          |
|                                       |                 |         |       | Adjusted Overall Accuracy 91.4% ( $\pm 1.7\%$ ) |   |       |                       |       | Adjusted Overall Accuracy 90.9% ( $\pm 1.8\%$ ) |
| Globally-optimized Thresholds (L8)    |                 |         |       |   | Temperate Optimized Thresholds (Sen-2)      |       |                       |       |   |
| REFERENCE                             |                 |         |       |   | REFERENCE                                   |       |                       |       |   |
| Class                                 | AIS             | Not AIS | Total | User's Accuracy                                 | Class                                       | AIS   | Not AIS               | Total | User's Accuracy                                 |
| AIS                                   | 411             | 132     | 543   | 75.7% ( $\pm 0.2\%$ )                           | AIS   | 226   | 55                    | 281   | 80.6% ( $\pm 0.3\%$ )                           |
| Not AIS                               | 105             | 720     | 825   | 87.3% ( $\pm 0.1\%$ )                           | Not AIS                                     | 24    | 379                   | 403   | 94.0% ( $\pm 0.1\%$ )                           |
| Total                                 | 516             | 852     | 1368  |   | Total                                       | 250   | 434                   | 684   |   |
| Producer's accuracy                   | 78.7%           | 84.5%   |       | F1 Score 77.6%                                  | Producer's accuracy                         | 90.4% | 87.5% ( $\pm 0.1\%$ ) |       | F1 Score 85.1%                                  |
|                                       | ( $\pm 0.2\%$ ) |         |       | Overall Accuracy 82.7%                          |   |       |                       |       | Overall Accuracy 88.5%                          |
|                                       |                 |         |       | Adjusted Overall Accuracy 95.5% ( $\pm 2.3\%$ ) |   |       |                       |       | Adjusted Overall Accuracy 92.9% ( $\pm 2.2\%$ ) |
| Esr Deep Learning Model (Sen-2)       |                 |         |       |   | Tropics Optimized Thresholds (Sen-2)        |       |                       |       |   |
| REFERENCE                             |                 |         |       |   | REFERENCE                                   |       |                       |       |   |
| Class                                 | AIS             | Not AIS | Total | User's Accuracy                                 | Class                                       | AIS   | Not AIS               | Total | User's Accuracy                                 |
| AIS                                   | 350             | 71      | 421   | 83.1% ( $\pm 0.2\%$ )                           | AIS   | 155   | 73                    | 228   | 68.0% ( $\pm 0.4\%$ )                           |
| Not AIS                               | 85              | 862     | 947   | 91.0% ( $\pm 0.1\%$ )                           | Not AIS                                     | 24    | 432                   | 456   | 94.2% ( $\pm 0.1\%$ )                           |
| Total                                 | 435             | 933     | 1368  |   | Total                                       | 179   | 505                   | 684   |   |
| Producer's accuracy                   | 80.5%           | 92.4%   |       | F1 Score 81.8%                                  | Producer's accuracy                         | 86.6% | 85.5% ( $\pm 0.1\%$ ) |       | F1 Score 76.2%                                  |
|                                       | ( $\pm 0.2\%$ ) |         |       | Overall Accuracy 88.6%                          |   |       |                       |       | Overall Accuracy 85.8%                          |
|                                       |                 |         |       | Adjusted Overall Accuracy 93.0% ( $\pm 1.8\%$ ) |   |       |                       |       | Adjusted Overall Accuracy 90.4% ( $\pm 2.1\%$ ) |

GISA was more accurate, we compared differences in individual Sentinel-2 scene-based thresholds and globally-optimized (as defined by analysis across all six sites) thresholds as shown in Table III. The accuracy metrics are shown individually for each scene in Fig. 8. For GMIS-trained output the most accurate global threshold had overall accuracy of 82.7% with adjusted overall accuracy of 86.5% ( $\pm 2.3\%$ ) and F1 score of 77.6%. Whereas the GISA-trained most accurate global threshold had higher overall accuracy (86.9%) and adjusted overall accuracy (91.4% ( $\pm 1.7\%$ )) and F1 score (80.2%). When using the most accurate threshold combination determined from individual scene points instead of jointly with accuracy points combined from all six sites, the overall accuracy was 88.8% with adjusted overall accuracy of 93.8% ( $\pm 1.4\%$ ) and an F1 score of 83.4%. The confusion matrices indicate that each algorithm in this study slightly overestimated AIS extent with more commission errors than omission errors.

| Globally-optimized Threshold (Sen-2) |          |       |           |        | Scene-Optimized Thresholds (Sen-2) |          |       |           |        |
|--------------------------------------|----------|-------|-----------|--------|------------------------------------|----------|-------|-----------|--------|
| US-OG                                | 0.855    | 0.825 | 0.703     | 1.000  | US-OG                              | 0.868    | 0.831 | 0.733     | 0.961  |
| US-MA                                | 0.934    | 0.929 | 0.961     | 0.899  | US-MA                              | 0.947    | 0.943 | 0.971     | 0.917  |
| UK                                   | 0.880    | 0.787 | 0.781     | 0.794  | UK                                 | 0.880    | 0.787 | 0.781     | 0.794  |
| QA                                   | 0.794    | 0.725 | 0.689     | 0.765  | QA                                 | 0.851    | 0.821 | 0.709     | 0.975  |
| IN                                   | 0.858    | 0.778 | 0.824     | 0.737  | IN                                 | 0.880    | 0.821 | 0.827     | 0.816  |
| EG                                   | 0.895    | 0.600 | 0.486     | 0.783  | EG                                 | 0.904    | 0.633 | 0.514     | 0.826  |
|                                      | Accuracy | F1    | Precision | Recall |                                    | Accuracy | F1    | Precision | Recall |
| Globally-optimized Threshold (L8)    |          |       |           |        | Deep Learning Model (Sen-2)        |          |       |           |        |
| US-OG                                | 0.868    | 0.808 | 0.724     | 0.913  | US-OG                              | 0.890    | 0.812 | 0.964     | 0.701  |
| US-MA                                | 0.912    | 0.910 | 0.910     | 0.910  | US-MA                              | 0.846    | 0.854 | 0.797     | 0.919  |
| UK                                   | 0.850    | 0.892 | 0.859     | 0.927  | UK                                 | 0.894    | 0.824 | 0.767     | 0.889  |
| QA                                   | 0.772    | 0.600 | 0.780     | 0.487  | QA                                 | 0.877    | 0.791 | 0.981     | 0.662  |
| IN                                   | 0.814    | 0.708 | 0.750     | 0.671  | IN                                 | 0.832    | 0.768 | 0.716     | 0.829  |
| EG                                   | 0.754    | 0.364 | 0.262     | 0.593  | EG                                 | 0.974    | 0.850 | 1.000     | 0.739  |
|                                      | Accuracy | F1    | Precision | Recall |                                    | Accuracy | F1    | Precision | Recall |

Fig. 8. Accuracy metrics for each study site for the globally optimized and scene-optimized thresholds using GISA (Sen-2), and the globally optimized thresholds using GMIS (L8), and deep learning model from Esri.

The Esri model produced from a CNN, manual labels, and a full time series had an overall accuracy of 88.6% ( $\pm 1.4\%$ ) and an F1 score of 81.8%, both of which were the highest global model overall. As compared with the GISA-based global model, this dataset had higher rates of false positive AIS, but lower rates of false negative AIS. At the individual scene level as visualized in Fig. 8, the Esri model had far superior accuracy in EG with an overall accuracy of 97.1% and F1 score of 85.0% as compared with GISA global-based output with 89.5% overall accuracy and 60.0% F1 score. The other arid scene, QA, also had higher accuracy with the Esri model. Interestingly, the US-MA scene had relatively low accuracy in the Esri LC model (Overall accuracy: 84.6%, F1 score: 85.4%) as compared with the GISA global-based model (overall accuracy: 93.4%, F1 score: 92.9%), which was attributed largely to low precision in the Esri model for this scene.

### E. Climate Region Adaptive Thresholds

We evaluated the utility of climate-based NDVI thresholds when generating GISA-trained AIS output and whether these types of thresholds yielded changes to overall accuracy and F1-scores. The confusion matrices can be seen in Table III. The three temperate climate scenes were grouped together (U.K., US-MA, US-OG), and the three tropics and subtropics scenes were a second group (EG, IN, QA). The highest accuracy for the temperate group was 88.6% (STD  $\pm 3.9\%$ ) with F1-score of 84.2% (STD  $\pm 7.5\%$ ) with NDVI maximum thresholds for dense AIS of 0.15 (winter), 0.15 (nonwinter), and NDVI thresholds for light AIS of 0.20 (winter minimum), 0.25 (nonwinter minimum), 0.65 (winter maximum), 0.45 (nonwinter maximum). For the tropics and subtropics group, the highest overall accuracy was 85.8% (STD  $\pm 1.4\%$ ) with F1-score of 71.7% (STD  $\pm 14.3\%$ ) and NDVI maximum thresholds for dense AIS of 0.15 (winter), 0.25 (nonwinter), and NDVI thresholds for light density AIS of



0.25 (winter minimum), 0.15 (nonwinter minimum), 0.45 (winter maximum), and 0.65 (nonwinter maximum). Furthermore, arid and nonarid regions were evaluated using climate-based NDVI thresholds. The QA and EG scenes were the arid group, and the remaining four scenes were the nonarid group. The highest accuracy for the arid group was 86.0% (STD  $\pm$ 1.9%) with F1-score of 68.7% (STD  $\pm$ 18.9%) with NDVI maximum thresholds for dense AIS of 0.15 (winter), 0.25 (nonwinter), and NDVI thresholds for light density AIS of 0.25 (winter minimum), 0.15 (nonwinter minimum), 0.45 (winter maximum), and 0.65 (nonwinter maximum). Whereas for the nonarid group the highest accuracy was 87.9% (STD  $\pm$ 3.5%) with F1-score of 82.8% (STD  $\pm$ 6.3%) and NDVI maximum thresholds for dense AIS of 0.25 (winter), 0.15 (nonwinter), and NDVI thresholds for light density AIS of 0.15 (winter minimum), 0.20 (nonwinter minimum), 0.65 (winter maximum), and 0.55 (nonwinter maximum).

#### F. Outputs AIS Maps

The AIS extent maps for the global GISA-trained output and GMIS-trained output from the stepwise threshold output are shown in Fig. 9. At this scale, the figure visualizes differences between the two outputs over areas of notable variation. For the two arid scenes (EG and QA), the GISA-trained output generally detected more AIS extent, including more false positives than GMIS in QA, but detected fewer errors in EG especially along the shoreline of Lake Nasser in the southern portion of the scene. This difference even accounts for the large false positive AIS area from GISA-based output over the water body in EG as shown in Fig. 9. For the IN scene, GISA-trained output appeared to better extract AIS in mixed pixels with tree cover and other subtropical vegetation, but with additional commission errors along water, agriculture, and wetland areas in the southwest portion of the image. For the US-OG site, the GISA-trained output resulted in more AIS commission errors on sand dunes in the western part of the image, however, the GMIS-trained output resulted in more AIS commission errors in agricultural areas, especially north of the city of Eugene as shown in Fig. 9. The QA scene illustrates large areas of Doha's AIS missed by GMIS-based output, but captured by the GISA-based output. The GISA-based output also has the trend of increased AIS false positives along water body boundaries, whereas the GMIS-based output was stronger in this regard.

#### G. Beach Bare Ground and NDVI Standard Deviation

The utility of the beach bare ground mask and the annual composite of NDVI standard deviation in the classification workflow was evaluated by comparing the overall accuracy variation on the 90th percentile and higher ranked threshold combinations on the GISA-trained output by computing them with and without these inputs. The results are visualized in the boxplots shown in Fig. 10. In 5/6 scenes the median overall accuracy was higher when using both additional inputs, the exception being the QA scene. This was likely due to many of the suburban AIS areas that are adjacent to agricultural areas with both small water bodies and temporally varied NDVI. However,

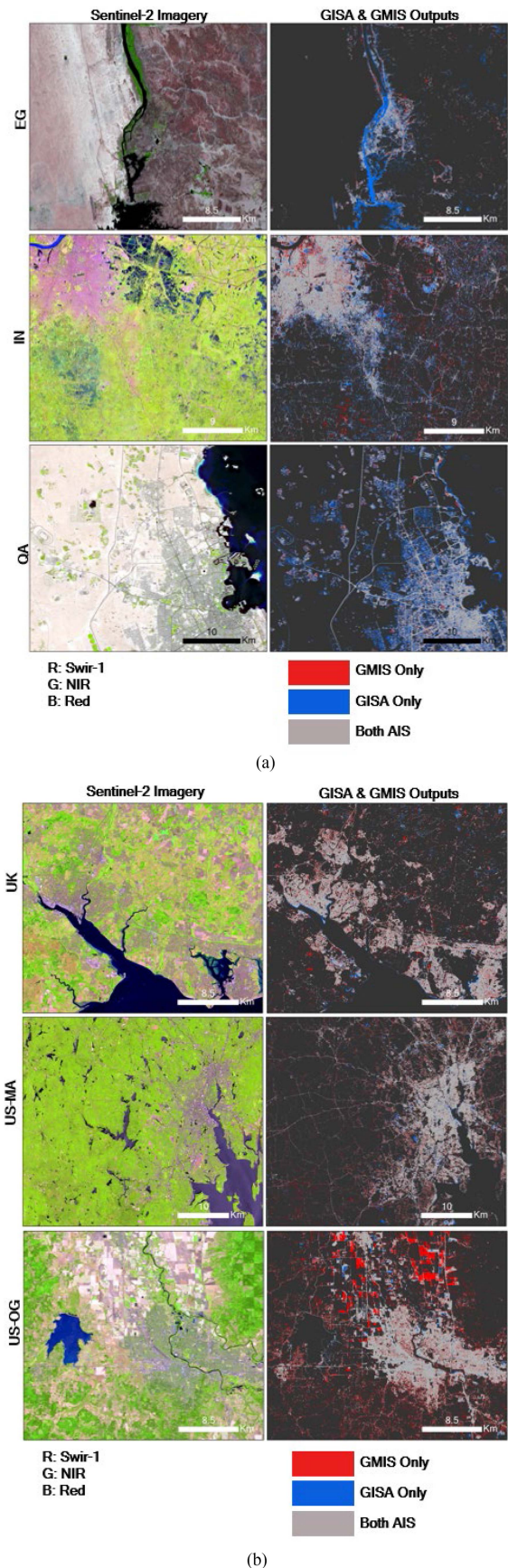


Fig. 9. (a), (b) GISA-based global output vs GMIS-based global output zoomed to areas of disagreement.



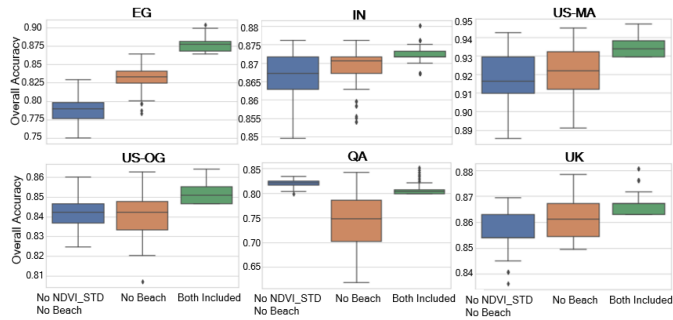


Fig. 10. Evaluation of the 90th percentile of most accurate threshold combinations from GISA-based output for each site highlighting variation in overall accuracy with and without the novel beach bare ground sampling input and annual NDVI standard deviation composite input.

the highest overall accuracy in QA was still attained using both inputs, as shown by the outlier value in the boxplot. The most substantial effect on the overall accuracy was observed in the EG scene, where both the beach bare ground sampling input and NDVI\_STD resulted in substantially improved accuracy. This was likely due to a large area of shoreline in this scene, as well as a substantial area of sparsely vegetated areas which would resemble AIS in one or two dates of imagery but are differentiated with the time series composite. The remaining scenes also yielded improvements from these additional inputs, but to a lesser degree.

#### H. Spatial Qualitative Intercomparison of All AIS Models

The workflow outputs comparing the major inputs and models evaluated are shown in Fig. 11 and zoomed in to focus on areas of substantial variation within each scene. The first column of the figure shows the Sentinel-2 imagery, the middle column visualizes differences between the global models (GISA and GMIS based outputs) and the Esri model. The right column visualizes differences between the GISA global model, scene model, and GISA global model without beach bare ground sampling and the annual NDVI standard deviation composite. The results produced some interesting observations. Within all six scenes, the GISA and GMIS based output capture roads and highways that are not located within settlement areas and are missed by the Esri model. Small areas of false positive AIS by GISA and GMIS outputs can be observed in the US-OG and U.K. scenes in wetlands and croplands, with GMIS outputs appearing to have more. Large areas of false positive AIS can be observed over water and hills in EG. Whereas, in US-MA, the Esri model appears to have a buffer around AIS areas leading to large false positive rate. Interestingly, the errors by the Esri LC model tend to be close to AIS areas and could often be considered as an urban land use, but typically cover non-AIS land covers such as grasses, trees, and other low vegetation. For the IN scene, the Esri LC model appears to capture false positive AIS in a similar manner as the US-MA scene, but to a lesser degree. For the QA scene, the Esri model is visualize capturing low-density AIS that was missed by the GMIS and GISA based outputs. The right-side column visualizing differences between GISA

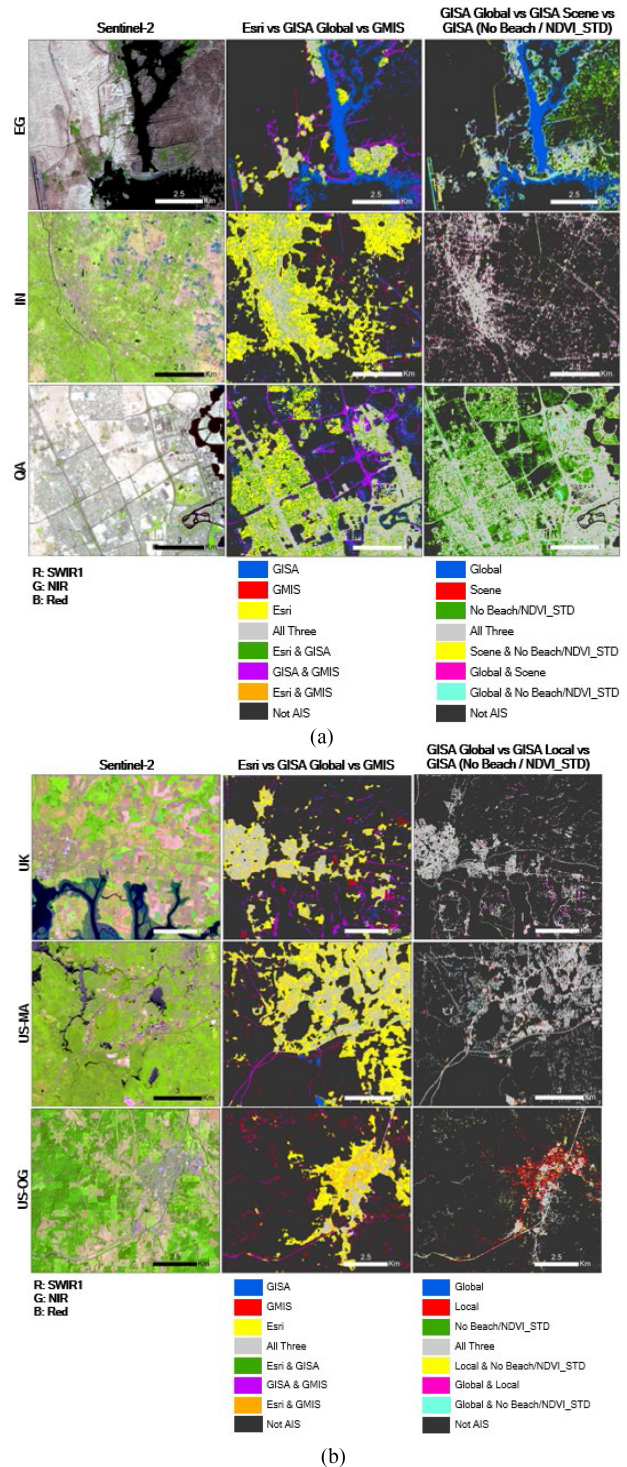


Fig. 11. (a), (b) Areas of significant variation between the three main global and local models (GISA2016, GMIS2010, and Esri).

global, GISA scene, and GISA global model without NDVI standard deviation or beach sampling shows some interesting patterns. Both arid scenes appear to be the most impacted by the exclusion of beach bare ground and NDVI standard deviation. In EG the exclusion of these results in AIS false positives in the sparsely vegetated hills, and in QA it leads to AIS false

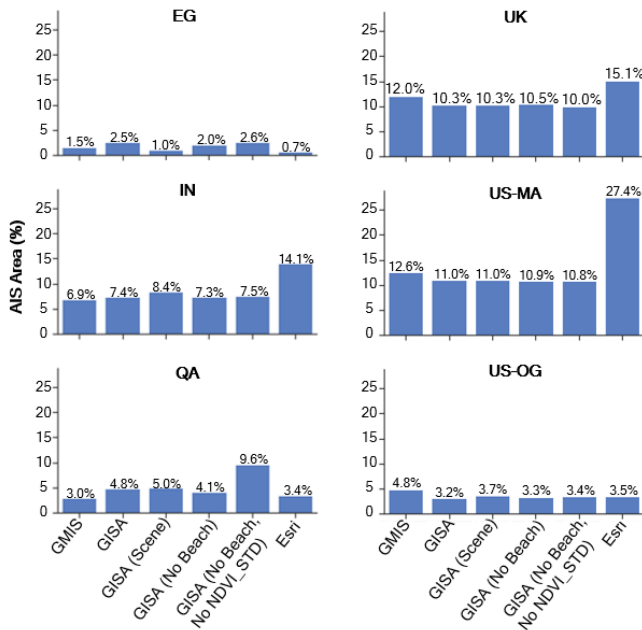


Fig. 12. Percentage of area mapped as AIS for each scene compared among the different models.

positives seen in bare ground adjacent to AIS as well as in areas adjacent to water. However, in the QA scene, the exclusion also led to increased and more accurate detection of AIS as seen in the green pixels. Minimal variation was observed in the US-MA and IN scenes. The US-OG scene visualizes a small town where the GISA scene model detects significantly more AIS than the GISA global model. In the U.K. scene, a few extra tracts of shoreline are falsely detected as AIS than in the GISA models. The use of the NDVI standard deviation layer and beach layer had the least effect on the three temperate sites (U.K., US-MA, US-OG).

### I. Quantitative Intercomparison of All AIS Models

While the maps from Fig. 11 provide examples of variation between the different models, they do not quantify the results across the entire scene. The AIS areal variation between the different layer is shown in Fig. 12. In general, the Esri model detected more AIS than the GISA global model in the U.K., US-MA, and U.K. site, but it detected substantially less AIS in EG and QA, the two arid sites. Here we can see that mapped areas are very similar between the different models for the US-OG scene, but we see a massive difference in the US-MA scene between Esri and the GISA and GMIS based outputs.

The top of Fig. 13 shows the pixel agreement between the GISA-based global output, GMIS-based global output, and Esri models for each scene. One interesting trend to note is the US-MA scene has approximately 54% of AIS pixels detected only by the Esri model (many false positives), and around 30% of AIS pixels detected by all three models. The QA and IN scenes appear to have the lowest percentage of AIS pixels where all three models agree. The US-OG scene has the highest percentage of GMIS-model only AIS pixels.

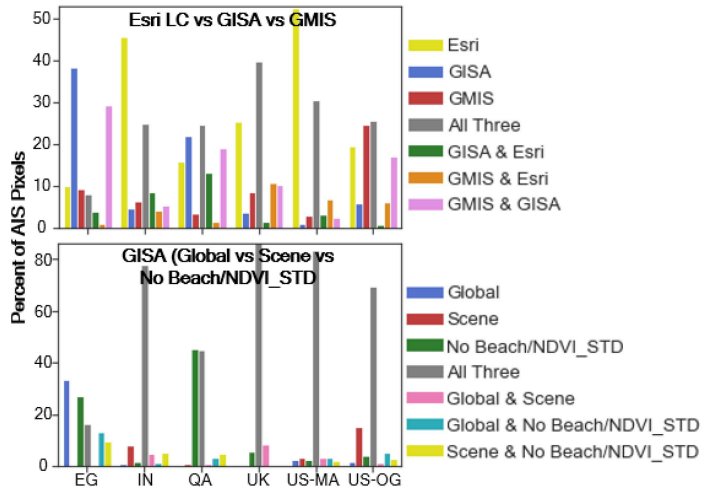


Fig. 13. Pixel agreement for AIS pixels between the three main global models (top facet) and the three variations on the GISA2016 model (bottom facet).

The bottom of Fig. 13 shows the pixel agreement variation between the GISA-based global, GISA-based scene, and GISA-based global without NDVI standard deviation and beach bare ground models. For the US-OG, US-MA, U.K., and IN scenes more than 65% of AIS pixels are in agreement between the models. For the QA scene, more than 40% of AIS pixels are from the GISA model without the NDVI standard deviation.

## V. DISCUSSION

Although the results produced similar accuracy levels as previous studies, our approach offers a way to generate AIS extent maps on-demand for current years without manual training data labeling. We created an automated training data and image classification approach and evaluated it using different existing AIS layers as input (GMIS and GISA) produced at different spatial resolutions and time periods, as well as the use of region and climate based optimal NDVI threshold combinations when generating training data. The evaluation was carried out across six broad, international geographic regions using a robust and balanced accuracy assessment scheme encompassing the entirety of the Sentinel-2 granule. In comparison to other studies mapping AIS using Sentinel-2, we had similar accuracy. One study using unsupervised classification with spectral indexes on one Sentinel-2 scene centered on Wuhan, China had 92.6% overall accuracy [75], whereas our study's most accurate single scene was US-MA with 93.4% (global threshold) and 94.7% (scene threshold). One study evaluated automatic training data selection for AIS using an NTL dataset and Sentinel-2 imagery in Paris, Harbin, and Beijing with a high overall accuracy above 97.0% [76].

Another study in Xiamen City using decision tree classification with optimized thresholds on Sentinel-2 and Landsat imagery found an overall accuracy of 94.5% [77]. One study conducted on a time series of Sentinel-2 imagery in Jinan City, China using an encoder-decoder deep learning framework achieved an F1 score of 0.87 which was higher than 4/6 of



the scenes evaluated in our study [78]. Several other studies evaluated scene-optimized spectral index thresholds with strong accuracy ( $>90\%$ ) on individual subscenes of Sentinel-2 imagery [7], [79]. One recent study used fusion of Sentinel-1 SAR and Sentinel-2 time series with manual training data and a random forest achieving an accuracy of 94.8% over the Yellow River Delta [80]. Another used Sentinel-2 imagery with CNNs and automated training data selection across 12 international cities to map human settlement extent had overall accuracies ranging from 81.8% to 94.8% [38]. This result was similar to the accuracies found in our study; however, their F1-scores were higher than those in this study. Other recent studies have leveraged time series of SAR and multispectral imagery for impervious surface extent mapping with accuracies above 94%, often with various deep learning models [53], [81], [82], [84]. One of the mentioned studies found SAR typically enhances overall accuracy by about 2%. Ultimately, the studies reported in the literature have a range of accuracy with many of the recent studies yielding slightly higher overall accuracies or F1-scores than our study. However, this increased accuracy often requires a time series of SAR and/or optical imagery, manual training data, or more intensive deep learning models. Whereas, this study relies only on two dates of imagery, a random forest model, and automatically generated training data.

Interestingly, our study's results were similar in accuracy to the Esri Land Cover dataset which was generated using much more intensive methods using a full time series of imagery and manually labeled data fed into a CNN. However, the Esri model had significantly higher accuracy in the two arid scenes (Qatar and Egypt), but significantly lower accuracy in the United States Massachusetts scene. The remaining scenes had similar accuracy levels. This difference is interesting and may be indicative of the advantages of contextual, deep learning models in regions where AIS spectral signature is similar to nontarget land cover (e.g., rocky terrain and bare ground which are prevalent in the arid scenes).

Because our accuracy assessment included equalized random sampling with 228 points per scene, it is possible that this method captures more errors than a standard simple random or stratified random sampling approach found in some of the previous studies. Some of the previous studies focus on one or two test sites or areas with majority AIS pixels which reduces the chances to capture false positives in areas such as cropland, wetland, etc. This variation in evaluated scenes and areas, compounded by variation in accuracy assessment methods, makes it difficult to equitably compare metrics. Comprehensive reviews of results and accuracies of other AIS studies are available [18], [84], [85], [86].

This study's results indicated that using climate region-based NDVI thresholds often yielded substantial improvements to the F1-score and modest increases to overall accuracy. With half of the scenes divided into the temperate climate NDVI threshold model, this yielded an overall accuracy of 88.5% and F1-score of 85.1%, and with the remaining half into the tropics group, this yielded an overall accuracy of 85.8% and F1-score of 76.2%. The globally optimized threshold model, meanwhile, was a compromise between the regions and had an overall accuracy

of 86.9% and F1-score of 80.2%. Although this is better than the tropics model, the average is raised due to the temperate scenes, and performance of the individual tropics scenes is lower. Interestingly, there was only about a 2% reduction in overall accuracy and 3% reduction in F1-score when using the globally-optimized model compared with the individual scene-optimized models. The global thresholds perform well in all sites except QA, where accuracy is reduced by about 6%. This study also evaluated using two different AIS layers generated from different data sources, time periods, and resolutions with GMIS from 2010 Landsat imagery at 30 m resolution and GISA from 2016 Sentinel-1/2 imagery at 10 m resolution. The overall accuracies were unsurprisingly higher for GISA-trained output due to its higher resolution and newer date and achieved an overall accuracy that was about 4% higher than GMIS-trained output, and 3% higher F1-score. However, the results clearly indicate that GMIS can still be accurately used even though it is from imagery more than ten years outdated.

There are several important limitations to mention from this study. While this study used robust accuracy assessments as guided by the literature [73], we were only able to evaluate six different international sites. While diverse, it did not include any high latitude locations or extremely mountainous terrain. While there are several globally available AIS layers, the two analyzed layers were useful for comparison spanning two sensors and different map dates (2016 versus 2010) and additional layers would likely not add much to the findings. Furthermore, the GMIS and GISA datasets which were modified and used as training data, contain their own data errors (e.g., false positive AIS) which can be carried over into the output produced from this study. The GMIS dataset was produced from 30 m Landsat imagery which has different pixel alignment and size than the 20 m Sentinel-2 used in this study. While effort was taken to resample and align, this discrepancy could have led to some errors. While this study used stepwise thresholds on NDVI and texture layers to reduce this, it is likely these errors played a role in this study's accuracy. While this study did evaluate the utility of using outdated AIS layers to capture current AIS extent, it did not analyze AIS change over time for a given location, or how mapping multiple years of change may result in varied accuracies. These topics would be beyond the scope of the article but would be of interest to explore in future research.

## VI. CONCLUSION

This study developed methods to extract AIS using Sentinel-2 20 m imagery with automatically generated training data for subsequent image classification using a random forest classifier. The AIS mapping framework used logical decision functions based on Euclidean distances and spectral index values to create binary masks of nontarget land cover such as water, wetlands, and bare ground. Target AIS binary masks were created using existing global masks generated from years prior and with stepwise optimized thresholds (e.g., NDVI thresholds). Specifically, the study evaluated the use of GMIS (2010) and GISA (2016). First, intermediate AIS extent maps were generated across the six Sentinel-2 granule-sized international test sites based on the

GISA and GMIS-based output and selected initial thresholds. Then, 1368 equalized random accuracy points were generated to evaluate 115 thresholds applied to the GMIS dataset such as minimum Euclidean distance of bare ground data from the AIS layer, minimum distance of AIS from water, and percentage pixel imperviousness, etc., to reduce commission errors in the training sets. Subsequently, the GISA stepwise thresholds were evaluated with 729 different NDVI minimum and maximum combinations for low- and high-density AIS classes. The thresholds were optimized based on different climate regions. Specifically, scene-based (thresholds optimized individually per scene) and global thresholds (thresholds optimized based on all six scenes combined) were evaluated first and found that global-based thresholds still yield overall accuracy that was only about 2% lower, suggesting a global model can accurately map AIS. For the climate regions, the temperate region group found promising overall accuracies of 88.5% ( $\pm 2.2\%$ ) and F1-score of 85.1%, while the tropics region group found overall accuracies of 85.8% ( $\pm 2.1\%$ ) and F1-score of 76.2%. Ultimately, the results suggested climate region-based thresholds yield higher accuracies than a global model. The study found that both GMIS and GISA datasets can be used to generate current-year AIS extent maps. However, the GISA-based output had higher accuracy at 86.9% ( $\pm 1.7\%$ ) than GMIS-based output at 82.7% ( $\pm 2.3\%$ ) using globally optimized thresholds for each. Differences between the GISA-based output's optimal NDVI thresholds were observed across different regions. The dense AIS NDVI winter maximum of 0.15 and light AIS NDVI winter maximum of 0.45 were both key for high accuracy in the arid and tropical sites. Whereas, for temperate and nonarid regions the light density AIS NDVI winter maximum of 0.65 showed up in the majority of the 90th percentile of most accurate stepwise threshold outputs. The analysis of the beach bare ground sampling layer and annual composite of NDVI standard deviation indicated modest improvements or negligible change at all but the EG site, suggesting these ancillary inputs could be excluded.

Lastly, comparison against a more data intensive, manually labeled CNN, found that overall accuracy and F1-score were not too different overall (86.9% GISA vs 88.6% Esri), except in the two arid sites (EG, QA) where Esri's CNN had significantly superior performance, and one US site where this study had significantly superior performance.

This automated training data generation framework can produce AIS extent layers relevant to recent time periods. This method is ideal for end-users who require a reproducible framework to quickly generate AIS extent maps with low data intensity, minimal manual input, and robust performance over a variety of international landscapes. Future work could explore the evaluation of this workflow in mountainous and snowy terrain where additional challenges would need to be overcome.

#### ACKNOWLEDGMENT

Permission to publish was granted by the ERDC Public Affairs Office. Any opinions expressed in this paper are those of the authors and are not to be construed as official positions of the funding agency.

#### REFERENCES

- [1] E. T. Stonecker, D. B. Jennings, and D. Garofalo, "Remote sensing of impervious surfaces: A review," *Remote Sens. Rev.*, vol. 20, no. 3, pp. 227–255, 2001.
- [2] A. Larkin et al., "Relationships between changes in urban characteristics and air quality in East Asia from 2000 to 2010," *Environ. Sci. Technol.*, vol. 50, no. 17, pp. 9142–9149, 2016.
- [3] S. V. Chithra et al., "Impacts of impervious surfaces on the environment," *Int. J. Eng. Sci. Invention*, vol. 4, no. 5, pp. 27–31, 2015.
- [4] K. Liu, X. Li, and S. Wang, "Characterizing the spatiotemporal response of runoff to impervious surface dynamics across three highly urbanized cities in southern China from 2000 to 2017," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 100, 2021, Art. no. 102331.
- [5] K. Vadrevu et al., "Spatial variations in vegetation fires and emissions in South and Southeast Asia during COVID-19 and pre-pandemic," *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 18233.
- [6] S. Bouzekri, A. A. Lasbet, and A. Lachehab, "A new spectral index for extraction of built-up area using Landsat-8 data," *J. Indian Soc. Remote Sens.*, vol. 43, no. 4, pp. 867–873, 2015.
- [7] J. Chen et al., "A comparative study of impervious surface extraction using Sentinel-2 imagery," *Eur. J. Remote Sens.*, vol. 53, no. 1, pp. 274–292, 2020.
- [8] Y. Zha, J. Gao, and S. Ni, "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery," *Int. J. Remote Sens.*, vol. 24, no. 3, pp. 583–594, 2003.
- [9] R. Bouhennache et al., "A new spectral index for the extraction of built-up land features from Landsat 8 satellite imagery," *Geocarto Int.*, vol. 34, no. 14, pp. 1531–1551, 2019.
- [10] K. Ezimand, A. A. Kakroodi, and M. Kiavarz, "The development of spectral indices for detecting built-up land areas and their relationship with land-surface temperature," *Int. J. Remote Sens.*, vol. 39, no. 23, pp. 8428–8449, 2018.
- [11] Z. Sun et al., "A modified normalized difference impervious surface index (MNDISI) for automatic urban mapping from Landsat imagery," *Remote Sens.*, vol. 9, no. 9, 2017, Art. no. 942.
- [12] H. Fang, Y. Wei, and Q. Dai, "A novel remote sensing index for extracting impervious surface distribution from Landsat 8 OLI imagery," *Appl. Sci.*, vol. 9, no. 13, 2019, Art. no. 2631.
- [13] H. Xu, "A new index for delineating built-up land features in satellite imagery," *Int. J. Remote Sens.*, vol. 29, no. 14, pp. 4269–4276, 2008.
- [14] A. Sekertekin, S. Abdikan, and A. M. Marangoz, "The acquisition of impervious surface area from LANDSAT 8 satellite sensor data using urban indices: A comparative analysis," *Environ. Monit. Assessment*, vol. 190, no. 7, pp. 1–13, 2018.
- [15] Z. Wang et al., "Application of a normalized difference impervious index (NDII) to extract urban impervious surface features based on Landsat TM images," *Int. J. Remote Sens.*, vol. 36, no. 4, pp. 1055–1069, 2015.
- [16] N. Wondrade et al., "GIS based mapping of land cover changes utilizing multi-temporal remotely sensed image data in Lake Hawassa Watershed, Ethiopia," *Environ. Monit. Assessment*, vol. 186, no. 3, pp. 1765–1780, 2014.
- [17] Y. Qin et al., "Quantifying annual changes in built-up area in complex urban-rural landscapes from analyses of PALSAR and Landsat images," *ISPRS J. Photogramm. Remote Sens.*, vol. 124, pp. 89–105, 2017.
- [18] Z. Shao et al., "Emerging issues in mapping urban impervious surfaces using high-resolution remote sensing images," *Remote Sens.*, vol. 15, no. 10, 2023, Art. no. 2562.
- [19] Q. Weng, "Remote sensing of impervious surfaces in the urban areas: Requirements, methods, and trends," *Remote Sens. Environ.*, vol. 117, pp. 34–49, 2012.
- [20] C. Li et al., "The first all-season sample set for mapping global land cover with Landsat-8 data," *Sci. Bull.*, vol. 62, no. 7, pp. 508–515, 2017.
- [21] X. Li and Y. Zhou, "Urban mapping using DMSP/OLS stable night-time light: A review," *Int. J. Remote Sens.*, vol. 38, pp. 6030–6046, 2017.
- [22] X. Li, P. Gong, and L. Liang, "A 30-year (1984–2013) record of annual urban dynamics of Beijing City derived from Landsat data," *Remote Sens. Environ.*, vol. 166, pp. 78–90, 2015.
- [23] P. Gong, X. Li, and W. Zhang, "40-Year (1978–2017) human settlement changes in China reflected by impervious surfaces from satellite remote sensing," *Sci. Bull.*, vol. 64, no. 11, pp. 756–763, 2019.
- [24] C. Small, F. Pozzi, and C. D. Elvidge, "Spatial analysis of global urban extent from DMSP-OLS night lights," *Remote Sens. Environ.*, vol. 96, nos. 3/4, pp. 277–291, 2005.

- [25] R. Chowdhury and S. Maithani, "Monitoring growth of built-up areas in Indo-Gangetic plain using multi-sensor remote sensing data," *J. Indian Soc. Remote Sens.*, vol. 38, no. 2, pp. 291–300, 2010.
- [26] X. Ma et al., "A new fusion approach for extracting urban built-up areas from multisource remotely sensed data," *Remote Sens.*, vol. 11, no. 21, 2019, Art. no. 2516.
- [27] N. Levin et al., "Remote sensing of night lights: A review and an outlook for the future," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111443.
- [28] C. Liu et al., "Automated extraction of built-up areas by fusing VIIRS nighttime lights and Landsat-8 data," *Remote Sens.*, vol. 11, no. 13, 2019, Art. no. 1571.
- [29] Q. Zheng, Q. Weng, and K. Wang, "Characterizing urban land changes of 30 global megacities using nighttime light time series stacks," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 10–23, 2021.
- [30] 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.
- [31] K. Shi et al., "Evaluation of NPP-VIIRS night-time light composite data for extracting built-up urban areas," *Remote Sens. Lett.*, vol. 5, no. 4, pp. 358–366, 2014.
- [32] M. Pesaresi et al., "Toward global automatic built-up area recognition using optical VHR imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 4, no. 4, pp. 923–934, Dec. 2011.
- [33] S. I. Deliry, Z. Y. Avdan, and U. Avdan, "Extracting urban impervious surfaces from Sentinel-2 and Landsat-8 satellite data for urban planning and environmental management," *Environ. Sci. Pollut. Res.*, vol. 28, no. 6, pp. 6572–6586, 2021.
- [34] R. Xu, J. Liu, and J. Xu, "Extraction of high-precision urban impervious surfaces from Sentinel-2 multispectral imagery via modified linear spectral mixture analysis," *Sensors*, vol. 18, no. 9, 2018, Art. no. 2873.
- [35] F. Wu et al., "Built-up area mapping in China from GF-3 SAR imagery based on the framework of deep learning," *Remote Sens. Environ.*, vol. 262, 2021, Art. no. 112515.
- [36] Y. Tan, S. Xiong, and Y. Li, "Automatic extraction of built-up areas from panchromatic and multispectral remote sensing images using double-stream deep convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 3988–4004, Nov. 2018.
- [37] W. Li et al., "Deep learning-based classification methods for remote sensing images in urban built-up areas," *IEEE Access*, vol. 7, pp. 36274–36284, 2019.
- [38] C. Qiu et al., "A framework for large-scale mapping of human settlement extent from Sentinel-2 images via fully convolutional neural networks," *ISPRS J. Photogramm. Remote Sens.*, vol. 163, pp. 152–170, 2020.
- [39] W. Liu, K. Quijano, and M. M. Crawford, "YOLOv5-Tassel: Detecting tassels in RGB UAV imagery with improved YOLOv5 based on transfer learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, 2022, pp. 8085–8094.
- [40] T. Gao, Q. Niu, J. Zhang, T. Chen, S. Mei, and A. Jubair, "Global to local: A scale-aware network for remote sensing object detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5615614.
- [41] E. F. Justiniano et al., "Proposal for an index of roads and structures for the mapping of non-vegetated urban surfaces using OSM and Sentinel-2 data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 109, 2022, Art. no. 102791.
- [42] Y. Forget et al., "Supervised classification of built-up areas in sub-Saharan African cities using Landsat imagery and OpenStreetMap," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1145.
- [43] Z. Ouyang, P. Fan, and J. Chen, "Urban built-up areas in transitional economies of Southeast Asia: Spatial extent and dynamics," *Remote Sens.*, vol. 8, no. 10, 2016, Art. no. 819.
- [44] X. Ma et al., "Optimized sample selection in SVM classification by combining with DMSP-OLS, Landsat NDVI and GlobeLand30 products for extracting urban built-up areas," *Remote Sens.*, vol. 3, 2017, Art. no. 236.
- [45] F. Schug et al., "Mapping urban-rural gradients of settlements and vegetation at national scale using Sentinel-2 spectral-temporal metrics and regression-based unmixing with synthetic training data," *Remote Sens. Environ.*, vol. 246, 2020, Art. no. 111810.
- [46] G. Lisini et al., "SAR-based urban extents extraction: From ENVISAT to Sentinel-1," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 8, pp. 2683–2691, Aug. 2018.
- [47] Y. Ban, A. Jacob, and P. Gamba, "Spaceborne SAR data for global urban mapping at 30 m resolution using a robust urban extractor," *ISPRS J. Photogramm. Remote Sens.*, vol. 103, pp. 28–37, 2015.
- [48] X. Feng, P. Li, and T. Cheng, "Detection of urban built-up area change from Sentinel-2 images using multiband temporal texture and one-class random forest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 6974–6986, 2021.
- [49] H. Cao et al., "Operational built-up areas extraction for cities in China using Sentinel-1 SAR data," *Remote Sens.*, vol. 10, no. 6, 2018, Art. no. 874.
- [50] M. Chini et al., "Towards a 20 m global building map from Sentinel-1 SAR data," *Remote Sens.*, vol. 10, no. 11, 2018, Art. no. 1833.
- [51] C. Corbane et al., "Enhanced automatic detection of human settlements using Sentinel-1 interferometric coherence," *Int. J. Remote Sens.*, vol. 39, no. 3, pp. 842–853, 2018.
- [52] K. Koppel et al., "Sentinel-1 for urban area monitoring—Analysing local-area statistics and interferometric coherence methods for buildings' detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2015, pp. 1175–1178.
- [53] B. Hu et al., "Improving urban land cover classification with combined use of Sentinel-2 and Sentinel-1 imagery," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 8, 2021, Art. no. 533.
- [54] S. Hafner, Y. Ban, and A. Nascetti, "Unsupervised domain adaptation for global urban extraction using Sentinel-1 SAR and Sentinel-2 MSI data," *Remote Sens. Environ.*, vol. 280, 2022, Art. no. 113192.
- [55] A. Garzelli and C. Zoppetti, "Large-scale monitoring of new built-up areas from joint use of Sentinel-1/2 images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 4388–4391.
- [56] N. Petrushevsky, M. Manzoni, and A. Monti-Guarnieri, "Fast urban land cover mapping exploiting Sentinel-1 and Sentinel-2 data," *Remote Sens.*, vol. 14, no. 1, 2021, Art. no. 36.
- [57] Z. Sun et al., "High-resolution urban land mapping in China from Sentinel 1A/2 imagery based on Google Earth Engine," *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 752.
- [58] X. Zhang et al., "Development of a global 30 m impervious surface map using multisource and multitemporal remote sensing datasets with the Google Earth Engine platform," *Earth System Sci. Data*, vol. 12, no. 3, pp. 1625–1648, 2020.
- [59] X. Huang et al., "Toward accurate mapping of 30-m time-series global impervious surface area (GISA)," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 109, 2022, Art. no. 102787.
- [60] X. Huang et al., "Mapping 10 m global impervious surface area (GISA-10m) using multi-source geospatial data," *Earth System Sci. Data*, vol. 14, no. 8, pp. 3649–3672, 2022.
- [61] E. B. De Colstoun et al., *Documentation For the Glob. Man-Made Impervious Surf. (GMIS) Dataset from Landsat*, NASA Socioeconomic Data Appl. Center (SEDAC), Palisades, NY, USA, 2017.
- [62] A. Schneider and C. M. Mertes, "Expansion and growth in Chinese cities, 1978–2010," *Environ. Res. Lett.*, vol. 9, no. 2, 2014, Art. no. 024008.
- [63] F. Gascon et al., "Copernicus Sentinel-2 mission: Products, algorithms and Cal/Val," in *Earth observing Syst. XIX*, vol. 9218, 2014, pp. 455–463.
- [64] A. A. Jarihani et al., "Satellite-derived Digital Elevation Model (DEM) selection, preparation and correction for hydrodynamic modelling in large, low-gradient and data-sparse catchments," *J. Hydrol.*, vol. 524, pp. 489–506, 2015.
- [65] K. Karra et al., "Global land use/land cover with Sentinel 2 and deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2021, pp. 4704–4707.
- [66] N. Mzid et al., "An analysis of bare soil occurrence in arable croplands for remote sensing topsoil applications," *Remote Sens.*, vol. 13, no. 3, 2021, Art. no. 474.
- [67] D. Žižala, R. Minařík, and T. Zádorová, "Soil organic carbon mapping using multispectral remote sensing data: Prediction ability of data with different spatial and spectral resolutions," *Remote Sens.*, vol. 11, no. 24, 2019, Art. no. 2947.
- [68] J. Al-Doski et al., "NDVI differencing and post-classification to detect vegetation changes in Halabja City, Iraq," *IOSR J. Appl. Geol. Geophys.*, vol. 1, no. 2, pp. 01–10, 2013.
- [69] K. Lasko et al., "Automated training data generation from spectral indexes for mapping surface water extent with Sentinel-2 satellite imagery at 10 m and 20 m resolutions," *Remote Sens.*, vol. 13, no. 22, 2021, Art. no. 4531.
- [70] K. Lasko, "Gap filling cloudy Sentinel-2 NDVI and NDWI pixels with multi-frequency denoised C-band and L-band synthetic aperture radar (SAR), texture, and shallow learning techniques," *Remote Sens.*, vol. 14, no. 17, 2022, Art. no. 4221.
- [71] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
- [72] F. Pedregosa et al., "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.



- [73] P. Olofsson et al., "Good practices for estimating area and assessing accuracy of land change," *Remote Sens. Environ.*, vol. 148, pp. 42–57, 2014.
- [74] P. Olofsson, "Updates to good practices for estimating area and assessing accuracy of land cover and land cover change products," in *Proc. IEEE Int. Geosci Remote Sens. Symp.*, 2021, pp. 1982–1985.
- [75] C. Li et al., "A comparative analysis of index-based methods for impervious surface mapping using multiseasonal Sentinel-2 satellite data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 3682–3694, 2021.
- [76] P. Tang et al., "A novel sample selection method for impervious surface area mapping using JL1-3B nighttime light and Sentinel-2 imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3931–3941, 2020.
- [77] L. Hua et al., "A new technique for impervious surface mapping and its spatio-temporal changes from Landsat and Sentinel-2 images," *Sustainability*, vol. 15, no. 10, 2023, Art. no. 7947.
- [78] J. Liu et al., "Monitoring impervious surface area dynamics in urban areas using Sentinel-2 data and improved Deeplabv3+ model: A case study of Jinan City, China," *Remote Sens.*, vol. 15, no. 8, 2023, Art. no. 1976.
- [79] S. Su et al., "An impervious surface spectral index on multispectral imagery using visible and near-infrared bands," *Remote Sens.*, vol. 14, no. 14, 2022, Art. no. 3391.
- [80] J. Liu et al., "Large-scale impervious surface area mapping and pattern evolution of the Yellow River Delta using Sentinel-1/2 on the GEE," *Remote Sens.*, vol. 15, no. 1, 2023, Art. no. 136.
- [81] Y. Lin et al., "Incorporating synthetic aperture radar and optical images to investigate the annual dynamics of anthropogenic impervious surface at large scale," *Remote Sens. Environ.*, vol. 242, 2020, Art. no. 111757.
- [82] B. Shrestha, S. Ahmad, and H. Stephen, "Fusion of Sentinel-1 and Sentinel-2 data in mapping the impervious surfaces at city scale," *Environ. Monit. Assess.*, vol. 193, no. 9, pp. 1–21, 2021.
- [83] W. Wu et al., "CroFuseNet: A semantic segmentation network for urban impervious surface extraction based on cross fusion of optical and SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 16, pp. 2573–2588, 2023.
- [84] Y. Wang and M. Li, "Urban impervious surface detection from remote sensing images: A review of the methods and challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 3, pp. 64–93, Sep. 2019.
- [85] S. W. Myint et al., "Per-pixel vs. object-based classification of urban land cover extraction using high spatial resolution imagery," *Remote Sens. Environ.*, vol. 115, no. 5, pp. 1145–1161, 2011.
- [86] R. Kaur and P. Pandey, "A review on spectral indices for built-up area extraction using remote sensing technology," *Arabian J. Geosci.*, vol. 15, no. 5, 2022, Art. no. 391.



**Kristofer Lasko** received the B.S. degree in geographical sciences, the M.S. degree in geospatial information science, and the Ph.D. degree in geographical sciences with an emphasis on remote sensing supervised by Dr. Krishna Vadrevu (NASA) and Dr. Chris Justice (University of Maryland) all from the University of Maryland, College Park, MD, USA, in 2012, 2014, and 2018, respectively.

Since 2018, he has been working as a Research Geographer with the Geospatial Research Laboratory, Engineer Research and Development Center, Alexandria, VA, USA.

Prior to this position, he was a Graduate Research Assistant and Geospatial Scientist with the University of Maryland, as well as an intern with NASA Goddard Space Flight Center and NASA Jet Propulsion Laboratory. He has served as a Lecturer at the graduate and undergraduate levels in GIS, remote sensing, and spatial modeling courses. His current research interests include the automation of land cover mapping from moderate and fine resolution remotely sensed imagery, biomass burning fires and associated emissions, and SAR-multispectral imagery fusion.



**Francis Daniel O'Neill** received the M.S. degree in geographical sciences from the University of Maryland, College Park, MD, USA, in 2019.

Since 2021, he has been employed as a Research Geographer with the Geospatial Research Laboratory, Engineer Research and Development Center, which is housed under the U.S. Army Corps of Engineers. He previously worked as a Geospatial Scientist with the University of Maryland, supporting the NASA Harvest Consortium. His research interests include machine learning, geospatial data processing, and

imagery analysis.