Mapping Africa's Settlements: High Resolution Urban and Rural Map by Deep Learning and Satellite Imagery

Mohammad Kakooei[1,5*], **James Bailie**[2,5], **Albin Söderberg**[1,5,†], **Albin Becevic**[1,5,†], **and Adel Daoud**[1,3,4,5]

[1]Department of Computer Science and Engineering, Chalmers University of Technology, Gothenburg, Sweden
[2]Department of Statistics, Harvard University, Massachusetts, USA
[3]Institute for Analytical Sociology, Linköping University, Sweden
[4]Center for Advanced Study in the Behavioral Sciences, Stanford University, United States
[5]The AI and Global Development Lab (www.global-lab.ai)
[*]corresponding author: Mohammad Kakooei (kakooei@chalmers.se)
[†]these authors contributed equally to this work

ABSTRACT

Accurate Land Use and Land Cover (LULC) maps are essential for understanding the drivers of sustainable development, in terms of its complex interrelationships between human activities and natural resources. However, existing LULC maps often lack precise urban and rural classifications, particularly in diverse regions like Africa. This study presents a novel construction of a high-resolution rural-urban map using deep learning techniques and satellite imagery. We developed a deep learning model based on the DeepLabV3 architecture, which was trained on satellite imagery from Landsat-8 and the ESRI LULC dataset, augmented with human settlement data from the GHS-SMOD. The model utilizes semantic segmentation to classify land into detailed categories, including urban and rural areas, at a 10-meter resolution. Our findings demonstrate that incorporating LULC along with urban and rural classifications significantly enhances the model's ability to accurately distinguish between urban, rural, and non-human settlement areas. Therefore, our maps can support more informed decision-making for policymakers, researchers, and stakeholders. We release a continent wide urban-rural map, covering the period 2016 and 2022.

Background & Summary

Land Use and Land Cover (LULC) maps are essential for understanding the interconnections between human activities and the natural environment[1–4]. Policymakers rely on precise geoinformation regarding these interactions to enhance sustainable development and planning strategies[5–9]. Typically, land cover mapping involves categorizing the landscape into various classes. For instance, cropland maps can specify different crop types[10], and human settlement (HS) maps can be subdivided into more detailed subclasses. Similarly, urban-rural maps provide a more focused view on human settlements and population distributions. The Global Human Settlement Layer (GHSL) project by the European Commission[11] has produced a comprehensive dataset covering the period from 1975 to 2030 (in 5-year intervals), which includes both historical data and future projections. This dataset, derived from satellite imagery and census data, offering maps at a 1 km resolution. It provides various metrics for settled areas (urban and rural) and experimental features like multiple-class land cover classification for more detailed analysis.

Two main approaches, including shallow models and deep learning models, are commonly used for LULC tasks. Shallow models refer to Machine Learning (ML) algorithms such as Random Forest (RF), Support Vector Machines (SVM), and K-nearest Neighbors (KNN)[12]. These models utilize not only the raw spectral reflectance values from Earth Observation (EO) images but also extract spectral indices that represent different features within a satellite image[13]. Historically, before the advent of advanced sensor technologies and increased computational power, shallow models were predominantly used for generating land cover maps[4]. However, shallow learning approaches face limitations concerning scale, context, and their ability to capture complex patterns in data compared to deep learning models.[14].

The availability of temporal satellite imagery on a global scale, coupled with advancements in computational power and ML techniques, has led to a shift from shallow learning to deep learning in remote sensing[14–20]. The most prevalent deep learning approach for land cover classification employs encoder-decoder network architectures, such as the U-Net architecture[21], which has gained prominence in semantic segmentation tasks[22]. Various backbones have been developed and applied to land cover classification with promising results[23–28]. For instance, Garg et al.[25] demonstrated that the DeepLabV3+ network architecture

significantly outperformed traditional shallow models across all benchmarks, even when using a small dataset.

one finding of this article is that urban and rural classes exhibit relationships with their surrounding environment that is beyond traditional rule-based definitions. For instance, a small human settlement area surrounded by cropland or vegetation is more likely to be classified as rural rather than urban. This context-aware classification approach enhances map resolution compared to rule-based definitions. This insight guided our use of deep learning models to integrate LULC and HS data, aiming to produce high-resolution rural and urban maps. A notable application of such high-resolution maps is the generation of poverty and wealth maps using satellite data, where precise urban/rural labels are crucial for socio-economic studies.

We generated a 3-class map, the High-Resolution Urban-Rural (HUR) map, covering the African continent at a 10 m spatial resolution, utilizing deep learning models and satellite imagery. This map classifies areas into rural, urban, and non-human settlement (NonHS) categories, with data spanning the period from 2016 to 2022. Figure 1 provides a schematic overview of this methodology. To create the target variables, we combined the ESRI LULC map with the JRC SMOD map, as shown in the Target data section. For the independent variables, we used yearly median images from Landsat-8 and nighttime light data, as illustrated in the Input data section. We trained a deep model and evaluated the resulting map with multiple strategies and metrics. In the final product, areas other than urban and rural are classified as Non-HS, forming a 3-class map with rural, urban, and Non-HS categories. To evaluate the map, we used a separate dataset with rural and urban labels from the Demographic and Health Survey (DHS) to compare the HUR map and the JRC SMOD map. The deep learning-based classification demonstrated superior performance in relation to the DHS labels, supporting our hypothesis that neighborhood information is a valuable feature for this type of classification. Finally, to reduce the impact of artificial boundaries between tiles, the map was generated using a sliding window approach.

Methods

Data Preparation. This study relies on remote sensing satellite imagery as its primary data source, accessed through Google Earth Engine (GEE)[29], a platform that facilitates the collection, storage, and analysis of large volumes of satellite data[30]. Although satellite imagery is highly effective for EO and land cover analysis[31], creating a robust satellite image dataset presents challenges, particularly regarding data quality and quantity. EO satellites collect data through regular and systematic orbits, but achieving a consistent number of high-quality observations over time is difficult due to issues such as cloud cover and shadows, which can degrade the imagery quality.

Landsat Data. The primary satellite imagery utilized in this study comes from Landsat-8. It provides multi-spectral data spanning from 2013 to the present. This makes it well-suited for our objective of generating an urban-rural map covering the years 2016 to 2022. Furthermore, as the Landsat mission program, initiated by NASA in 1974, has consistently deployed satellites to monitor and collect Earth surface data, allowing us to extend the temporal range of the dataset in the future. Landsat-8's core bands offer a resolution of 30×30 meters, capturing detailed information about the Earth's surface. In addition to these core bands, Landsat-8 includes a panchromatic band with 15×15 meters resolution and thermal bands with resolutions ranging from 60 to 120 meters[32]. While Landsat-8 offers 15-meter resolution in the panchromatic band, this study primarily relies on the 30-meter and coarser bands. This approach supports potential future expansions in dataset temporality aligned with the Landsat mission program, given that older Landsat satellites lacked a panchromatic band.

Nightlight Data. The Visible Infrared Imaging Radiometer Suite (VIIRS) provides satellite imagery that captures nightlight intensity and has been available since 2012 with a resolution of 463 meters per pixel[33].The Defense Meteorological Satellite Program (DMSP) also provides nighttime light data, covering the period from 1992 to 2013 at a coarser resolution of 927 meters. VIIRS offers a superior spatial resolution compared to DMSP[34], making it the only suitable choice for generating an urban-rural map for the years 2016 to 2022. Previous research has effectively utilized VIIRS data in combination with multi-spectral imagery to distinguish built-up areas (BuA) from other land cover types, such as forests, deserts, and vegetation, and to differentiate between rural and urban classes in Africa[17,35]. We anticipate that VIIRS nighttime light data will enhance the model's capacity to differentiate urban and rural areas

Land Cover Products. To generate land cover maps using a supervised deep learning approach, labeled target data is essential. Utilizing pre-existing land cover maps is generally preferable to manually labeling a smaller custom dataset, which is a challenging task. Table 1 provides details on existing land cover datasets, including their resolution and temporal coverage. This information was used to evaluate which datasets would be most suitable for serving as target labels in the mapping process.

The selection of target data for this study is based on three primary features, including temporality, spatial coverage, and spatial resolution. Firstly, temporality: datasets with a long and consecutive temporal span allow for temporal evaluation, providing insights into the dataset's reliability over time. Secondly, spatial coverage: the dataset should ideally cover the entirety of Africa to capture the continent's diverse land surface features. Lastly, spatial resolution: higher spatial resolution datasets enable the generation of more detailed maps, which is crucial for precise land cover classification.

ESA WorldCover and ESRI LULC stand out among the datasets due to their high spatial resolution in 10m, which provides a detailed level of mapping. Of the two, ESRI LULC also offers superior temporal coverage. From a temporal perspective,

NASA's GlanCE-v001[3] provides extensive temporal coverage, its lower spatial resolution makes it unsuitable for this study's requirements. Similarly, GlobeLand30[36] has a good temporal span, but insufficient 30m spatial resolution that falls short compared to the higher resolution needed. Consequently, this work proceeded with ESRI LULC due to its superior temporal coverage, higher estimated class accuracy, and better alignment with the deep learning approach, as it was produced using a deep model.

The ESRI Land Use Land Cover (LULC) 2017-2022 dataset was selected as part of the target data to train the deep model. Developed by Esri, a global leader in geographic information system (GIS) software, in partnership with Microsoft, this dataset provides a 10-meter resolution global land cover time-series map covering the years 2017-2022, known as the ESRI LULC 2017-2022 Time Series. With an estimated accuracy of 85%, this dataset was generated using a deep learning model for semantic segmentation, utilizing Sentinel-2 satellite data (10-60 meter resolution) as input. The dataset includes nine classes: Water, Trees, Flooded Vegetation, Crops, Built-up Area, Bare Ground, Snow/Ice, Clouds, and Rangeland. Performance assessments for these classes are provided in Table 2.

GHS-SMOD 2020. GHS-SMOD is a component of the Global Human Settlement Layer (GHSL) project[11]. It provides a human settlement map in a coarse resolution (1×1 km) raster format, classifying grid cells by their degree of urbanization. This dataset is generated by combining various GHSL products, which include data on built-up surfaces and population estimates, and applying a clustering scheme detailed in Table 3. Despite its coarse resolution, which limits its suitability for applications requiring medium to high-resolution human settlement data, GHS-SMOD can still serve as a valuable auxiliary dataset when merged with higher resolution labels. The dataset comprises eight classes: (1) Water Grid Cell, (2) Very Low Density Rural Grid Cell, (3) Low Density Rural Grid Cell, (4) Rural Cluster Grid Cell, (5) Suburban or Peri-urban Grid Cell, (6) Semi-Dense Urban Cluster Grid Cell, (7) Dense Urban Cluster Grid Cell, and (8) Urban Centre Grid Cell.

DHS dataset. The Demographic and Health Survey (DHS) dataset is a global survey conducted in multiple countries. In addition to survey responses collected directly from participants, the dataset includes labels indicating whether locations are classified as rural or urban, along with the corresponding displaced geographic coordinates. This displacement is part of the privacy protection method applied to DHS data before public release. In this study, we use the DHS urban and rural labels to assess the agreement between DHS labels and GHS JRC labels, as well as between DHS labels and our generated HUR map.

Retrieving Data Through Google Earth Engine. Various bands from Landsat-8 form the core of the input data for this study. Additionally, nightlight data from VIIRS is included to enhance the accuracy of human settlement classification. GHS-SMOD, a human settlement data product, is used as an auxiliary dataset for label augmentation. The ESRI LULC land cover data product is appended as the final band, serving as the target data for the semantic segmentation task. The complete setup of bands is detailed in Table 4. Notably, bands 1-7 comprise the model input, while bands 8 and 9 serve as the target labels during model training.

Image data covering the entire surface of the continent of Africa was gathered, to form the training set. This to ensure a distribution of land cover classes that is as representative as possible for the continent as a whole. It also ensures that the model sees a sufficiently large set of landcover examples during training. The alternative method would have been to gather a smaller scale dataset using stratified sampling, which would however have introduced a number of issues regarding how to sample the images evenly, from where to sample them and how many images would be enough. In order to manage the organization of data, image tiles were retrieved and stored country-wise. The data retrieval process would follow the same procedure for every country. First, a number of points (as many as possible) were distributed across the country, spaced with a distance of 10 km, forming a grid. An example of this can be seen in Figure 2(a). Next, an image tile would be sampled from each point, in a 10x10 km bounding square. Put together, the image tiles cover the surface of the country, as observed in Figure 2(b). Apart from the land cover labels (ESRI LULC) which represent an entire year, the satellite data itself is accessed as a collection of images; a time series. A given location is revisited many times a year, meaning that the value of a pixel will vary between observations. An image composite method therefore had to be applied, to form a yearly image representation. For this purpose the best option was to take the median of all observed pixel values, as the mean can be affected by outliers. To ensure that the captured satellite imagery was of sufficient quality, a mask was used to filter away pixels covered by cloud or shadow. Ultimately, the dimension of each retrieved tile was 10×10 km or 1000×1000 px (1px=10m). Data is sampled at 10 m resolution, which is the native resolution of the target data. Images from all other bands (their native resolution shown in Table 4 are resampled to 10 m resolution with nearest neighbour interpolation.

Classification Task. Let Y_g be the land cover class of a geo-location g discretized at a size of $10m \times 10m$ in the image X_g captured at that geo-point. The estimated land cover class for the same pixel, produced by the machine learning model M, is denoted as $\hat{Y}_g$. Thus, $\hat{Y}_g = M(X_g)$ represents the land cover estimate generated by the model given the satellite imagery at point g. The representation of a land surface as a raster of pixels is a necessary simplification for this task, but it is not without its limitations. A pixel is rigid in the sense that its value is homogeneous, while the underlying data may exhibit heterogeneity. This implies that the land area corresponding to a single pixel might encompass multiple land cover classes, yet the pixel is assigned only one class. Consequently, the rasterized representation may oversimplify the actual land cover complexity within

each pixel.

Target Data. Semantic segmentation is a classification task that requires pixel-wise labeled image data. In this study, models were trained on target data from ESRI LULC, augmented with human settlement labels from GHS-SMOD. ESRI LULC provided high-resolution (10m) labeled land cover data, serving as the core of the land cover target labels. GHS-SMOD, despite its much coarser resolution (1km), contributed large-scale differentiation between rural and urban settlement areas. It is important to note that both ESRI LULC and GHS-SMOD are products generated by deep convolutional neural networks, meaning they are predictions of the ground truth rather than actual ground truth data.

Model input. As described earlier, data is saved in tiles of dimension $10 \text{ km} \times 10 \text{ km}$ (1000×1000 pixels). This differs from the input dimensions used by the model. The base model used in this study, DeepLabV3, is a fully-convolutional network, meaning it only performs convolutional operations and is therefore independent of image size. A smaller image size is generally favored in terms of performance, both computationally and in training outcomes. In many deep learning tasks involving images, the typical dimension is 224×224 pixels. In this study, an input size of 250×250 pixels was deemed more suitable, as one tile of size 1000×1000 pixels can be evenly split into 16 sub-tiles of this size. Figure 2(c) visualizes the data throughput process.

Deep model. The most common architectures for land cover classification and various other segmentation tasks are U-net, Segnet, and DeepLab architectures. Each architecture has its own strengths, and the choice of which one to use depends on the data and specific requirements of the task at hand. U-net was designed to perform well on smaller datasets, featuring a simpler architecture with skip connections to capture small fine details[21]. Segnet is a memory-efficient network due to its utilization of pooling indices and performs well on tasks with well-defined object boundaries. However, it lacks skip connections, which can limit its ability to capture finer details[24]. On the other hand, DeepLabv3 excels at capturing context information at multiple scales and achieves state-of-the-art performance on several benchmarks. Nonetheless, it can be computationally more demanding compared to the previous two architectures and requires a large dataset for effective training due to its complexity[37].

The Landsat satellite imagery dataset contains diverse landscapes, including objects at multiple scales such as forests, water bodies, urban areas, and agricultural regions. Additionally, it includes irregularly shaped objects like rivers, coastlines, and vegetation boundaries. Given the considerations of the dataset's characteristics and the availability of a large Landsat imagery dataset, the chosen architecture for this study is DeepLabv3. This decision leverages DeepLabv3's ability to capture context at multiple scales, which is crucial for accurately classifying the varied and complex features present in the dataset.

Model architecture and Training The model configuration employs DeepLabV3 as the classification core and bases its predictions on input data from each year. This setup allows the model to produce land cover estimates on a year-by-year basis. Specifically, the model is trained on satellite data from the year 2020, with labels sourced from the ESRI LULC (2020) dataset and GHS-SMOD (2020). The training setup is depicted in Figure 1. The architecture follows a standard DeepLabV3 model, with ResNet-50 serving as its backbone. This configuration leverages the strengths of DeepLabV3 in capturing complex features at multiple scales and the robust feature extraction capabilities of ResNet-50.

Preprocessing the data is a crucial step in the training process, as it involves normalizing input data, handling missing data, and remapping labels. This step ensures that data is consistently and correctly formatted for input into the DeepLabV3 model during training. First, the satellite image bands are normalized to values between zero and one. Next, pixels with missing data are addressed. During the data retrieval step, a pixel quality mask was applied to filter out pixels affected by clouds or shadows, marking all pixels without valid observations with the value negative infinity to indicate missing data. Leaving these pixels as is would disrupt the model throughput, while removing each tile containing at least one missing pixel would result in unnecessary data loss. To address this issue, these pixels are assigned the value zero. The final step of preprocessing involves remapping the target labels, which is conducted in three substeps: re-indexing the labels to a value between 0 and $N-1$ (where N is the number of classes), removing unwanted classes, and augmenting class labels by merging different label datasets. Class removal, i.e., marking a class to be ignored, is applied to two classes: Snow/Ice and Clouds. Snow/Ice is a very limited land cover class on the African continent and would therefore risk needlessly interfering with other class predictions. Clouds represent an unknown land cover class due to noisy data and are thus excluded, as this setup utilizes a different satellite data collection.

Label augmentation is a crucial and intricate part of preprocessing. ESRI LULC includes a 'Built Area' class that aggregates all man-made structures such as buildings, roads, and other artificial features. While this general classification suffices for many land cover applications, it lacks the granularity needed for detailed human settlement analysis. To address this, a more refined classification scheme is necessary. For this purpose, a preprocessing method was devised to split the 'Built Area' class into more informative sub-classes. This method utilizes the auxiliary dataset GHS-SMOD (2020), which classifies land surface by degree of urbanization. By overlaying the ESRI LULC labels with those from GHS-SMOD, all pixels classified as 'Built Area' are relabeled according to the corresponding GHS-SMOD class at that location. GHS-SMOD provides seven specific settlement classes, but these are not always visually distinct, necessitating a merge to balance detectability and informativeness.

The chosen configuration aggregates GHS-SMOD classes into two broad categories: 'urban' and 'rural.' This approach consolidates the more detailed subclasses of both rural and urban into their respective superclasses. A visualization of this label augmentation applied to a sample tile is shown in Figure 3. The final set of target labels after preprocessing is summarized in

Table 5.

The loss function used to train the model is Cross-Entropy-Loss, commonly used for classification tasks, which measures the differences between two probability distributions. It is applied to the model output on a pixel-level. For each pixel in the input image, a vector of logits is given as output by the model, which after application of softmax can be viewed as a probability distribution over all classes. The definition of the Cross-Entropy-Loss-function, as applied to each pixel is shown in equation 1, where y is a one-hot encoded vector of the target label, $\hat{y}$ the vector containing the class probabilities, $K = 8$ being the number of classes, k used as a vector index pointing towards the value for class k and k^* denoting the true class of the pixel. The loss value is averaged for the pixels in the minibatch.

$$L(\hat{y}, y) = -\sum_{k=1}^{K} y^{(k)} log(\hat{y}^{(k)}) \tag{1}$$

Handling Class Imbalance: Weighted Loss. For classification tasks, achieving a balanced class distribution in the training set is crucial to avoid biases toward more prevalent classes. However, achieving this balance is often challenging, especially in cases where some classes are significantly underrepresented compared to others. This is particularly evident in land cover classification, where larger classes can be more than ten times the size of smaller ones. Figure 4 illustrates the class distribution of ESRI LULC (2020) for the African continent, highlighting the imbalance. To address class imbalance, various strategies can be employed, with weighted loss being one of the most effective. Weighted loss modifies the loss function to assign greater weight to underrepresented classes, thereby increasing the model's focus on these classes and penalizing misclassifications more heavily. This approach helps in improving the model's performance on minority classes.

Several weighting strategies in equation 2 were explored: complement probability ($w_k = 1 - p_k$), negative log of the probability ($w_k = -log(p_k)$), and inverse probability ($w_k = 1/p_k$). Among these, inverse probability weighting was chosen due to its superior performance in capturing rural settlements. This choice was based on visual inspections and evaluation score comparisons, which indicated that inverse probability weighting provided the best balance between detecting rare classes and overall classification accuracy.

$$L(\hat{y}, y) = -\sum_{k=1}^{K} w_k y^{(k)} log(\hat{y}^{(k)}) \tag{2}$$

K-fold-cross-validation. An effective evaluation strategy is essential for assessing how well a model generalizes to unseen data. K-fold cross-validation is a robust method for this purpose, as it provides insight into the model's performance across different subsets of the data. In this approach, the dataset is divided into k folds, and multiple models are trained on various fold configurations. Each fold serves as the test set exactly once, while the remaining folds are used for training and validation. For evaluating the performance of the deep model in this study, a 5-fold cross-validation strategy was employed. This involved training five distinct models, each on a different fold configuration. Specifically, each configuration consists of three folds used for training, one fold for validation, and one fold for testing, denoted as $\{(Train, Validation, Test)\}$. The full set of fold configurations used for model training was: $\{(123, 4, 5), (234, 5, 1), (345, 1, 2), (451, 2, 3), (512, 3, 4)\}$ In this setup, each model was trained on a unique combination of folds, with the corresponding validation and test folds ensuring a comprehensive evaluation across different subsets of the data.

Spatial autocorrelation is a fundamental aspect of satellite imagery[38], where neighboring pixels often exhibit higher similarity compared to those that are farther apart. This spatial dependence can influence the performance of machine learning models trained on such data. When a model is trained on one image tile and evaluated on an adjacent tile, the proximity of the tiles may result in an overestimation of the model's generalization capabilities, as the data may not be truly independent. This challenge underscores the importance of careful consideration in designing the train-test split to avoid skewed performance metrics. Standard random fold splitting, which typically ignores spatial dependencies, can lead to artificially inflated performance results because adjacent or nearby tiles may share similar features. Rolf et al.[39] explored the impact of spatial separation on model performance by using a checkerboard grid with varying spacing. Their findings suggested that tighter grid layouts, which maintain spatial separation, often yielded more accurate assessments of model performance compared to more loosely spaced grids, indicating a potential overestimation in performance when spatial dependencies are not considered. To address this issue, the study adopts a country-wise partitioning approach for the data split. By assigning entire countries to each fold, this method effectively enforces spatial boundaries and minimizes the risk of data leakage between training and test sets. Additionally, country-wise partitioning allows for evaluation at the country level, providing a more realistic measure of the model's performance across diverse geographical regions.

Data was divided into five folds, with each fold comprising a subset of countries. To achieve balanced folds, countries were first sorted by area and then assigned to the folds in a cyclic manner. This approach ensured that each fold contained a roughly equal number of countries and was of similar size. The resulting folds and their corresponding countries are as follows:

Fold 1: Algeria, Niger, Mauritania, Mozambique, Central African Republic, Zimbabwe, Guinea, Malawi, Togo

Fold 2: Democratic Republic of the Congo, Angola, Egypt, Zambia, Madagascar, Congo, Ghana, Eritrea, Guinea-Bissau,

Fold 3: Sudan, Mali, United Republic of Tanzania, Morocco, Botswana, Côte d'Ivoire, Uganda, Benin, Lesotho

Fold 4: Libya, South Africa, Nigeria, South Sudan, Kenya, Burkina Faso, Senegal, Liberia, Equatorial Guinea

Fold 5: Chad, Ethiopia, Namibia, Somalia, Cameroon, Gabon, Tunisia, Sierra Leone, Burundi

A visualization of the split is displayed in Figure 5, with each fold highlighted by a separate color.

Output generation: smooth tiling. After the models have been trained and evaluated, it is time for deployment. They can now be used to generate land cover maps of Africa. The simplest procedure would be to feed the model, as input, a grid of non-overlapping tiles covering the desired area and join them together in one larger image. This method does however lead to unwanted edge effects in the final image. As the models have been trained on images of dimension 250×250 px, using the same input dimensions when running inference is the natural choice. The deep model can be applied to images of arbitrary size, as it is a fully-convolutional-network, but should work optimally for the dimensions trained on. A method for smoother image tiling was implemented[40], to reduce the aforementioned edge effects. The edge effects are a result of the outer edge of an image getting less surrounding context, leading to neighboring tile edges not always matching. To reduce this problem, only the center of the prediction (which has the most context) is kept, while the outer edge of the prediction is thrown away as shown in Figure 6(a). This is displayed in Figure A larger number of (partially overlapping) image tiles are then needed to stitch together the full output tile. With neighboring tiles now receiving more context along the edges, it appears smoother and more accurate in Figure 6(b).

Data Records

The dataset is freely available for download on the AI and Global Developmenet Lab website (global-lab.ai) Figshare (https://doi.org/10.7910/DVN/YFRECD)[41]. It provides a high-resolution rural-urban map of the African continent, projected in the WGS84 coordinate system. The dataset covers the entire continent using multiple tiles, each with its respective Coordinate Reference System (CRS). It offers a spatial resolution of 10 meters and is organized into 86 Geotiff tiles. The file naming convention follows the format: *RuralUrban_aa_COG.tif*, where *aa* represents the tile number. Additionally, a PNG file named *AfricaGeometry.png* is included to illustrate the tile layout across the continent.

An interactive Google Earth Engine (GEE) application, (https://kakooeimohammd.users.earthengine.app/view/rural-urban-africa) is prepared to facilitate data access, where users can click on individual tiles to retrieve download links for the corresponding data. Each GeoTIFF file includes seven bands (b1 to b7), representing data from the years 2016 to 2022.

Technical Validation

The objective of this work is to develop a machine learning model capable of generating 10-meter resolution land cover maps that classify areas into LULC including urban and rural classes among others based on satellite imagery. The validation process involves assessing the performance of the trained deep learning model in accurately generating LULC classes and HUR map. Finally, the HUR map, which includes classifications for rural, urban, and non-human settlement areas, is evaluated using the DHS dataset to demonstrate its reliability.

Evaluating the model's performance. To achieve LULC map, a deep learning model was trained using a 5-fold cross-validation approach. The model's performance is assessed based on the average results across all five folds. Each of the five models, trained in the cross-validation process, was evaluated on its respective test fold, and the classification results were summarized in confusion matrices to assess model performance. The raw counts from the five individual confusion matrices were combined to form a fold-average confusion matrix, which represents the overall performance across the entire African continent. In addition to this, a separate confusion matrix was generated for each individual country, using the model that was trained on the data excluding that particular country (i.e., the model corresponding to the test fold in which the country was included). This approach allowed for a detailed analysis of model performance at a country-specific level. The resulting data was used to create boxplots of the performance metrics, illustrating their distribution and variation across different countries.

As semantic segmentation is fundamentally a classification task, standard classification metrics are employed. These metrics are derived from the predicted class, which is determined by the highest predictive probability. The key variables used in these metrics are True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN), which represent the counts of correctly identified positive cases, incorrectly identified positive cases, correctly identified negative cases, and incorrectly identified negative cases, respectively.

Accuracy. Accuracy is the simplest classification metric, measuring the percentage of correct predictions made by the model. It is calculated using the formula 3.

$$accuracy = \frac{number\ of\ correctly\ classified\ pixels}{total\ number\ of\ pixels} = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

The average fold accuracy achieved is 0.849. This overall pixel accuracy provides a broad measure of performance, indicating that the model reliably produces accurate land cover maps. Given its high value, the accuracy suggests that the model performs well in classifying pixels correctly, thus ensuring the generated maps are a reliable representation of the land cover.

Accuracy, while a useful benchmark for overall model performance, does not always provide a complete picture in the context of semantic segmentation. Metrics such as Precision, Recall, F1 Score, and Intersection over Union (IoU) are crucial for providing a more detailed evaluation of how well the model performs across all classes.

Confusion Matrix: Recall and Precision. A confusion matrix offers a detailed view of per-class performance in semantic segmentation tasks. It is a 2D matrix where each cell represents the count of pixels from one class that are classified as another class. By examining the rows and columns, you can gain insights into how well the model distinguishes between different classes. The diagonal of the confusion matrix represents correct classifications, while off-diagonal elements indicate misclassifications. A higher sum of the diagonal elements relative to the total sum indicates better overall model performance, reflecting the model's effectiveness in accurately identifying and distinguishing between classes.

Recall. Recall measures the percentage of correctly classified pixels for a specific class, reflecting the proportion of actual pixels belonging to 'class A' that are accurately identified as 'class A'. This metric can be obtained from the confusion matrix by normalizing the diagonal values for each class by the sum of their respective rows. Alternatively, recall can be computed using the formula 4:

$$recall = \frac{TP}{TP + FN} \tag{4}$$

Table 6 illustrates a well-defined diagonal and generally low levels of confusion, indicating high recall. Most classes achieve satisfactory recall performance of at least 0.75, with several classes surpassing this target by over 0.15. The sole exception is the class 'rural,' which only reaches a recall of 0.65. Notably, there is significant confusion between 'rural' and 'urban,' with 'rural' recall being lower at 0.65 compared to 'urban's' 0.79. However, this confusion is mostly contained within the settlement classes, meaning they are rarely confused with non-settlement classes.

Recall is a particularly informative metric in this context, as it reveals how accurately the model classifies each respective class by showing the percentage of pixels correctly classified. This is especially important for the classes 'rural' and 'urban,' which are crucial for the sub-goal of human settlement mapping. In the fold average recall matrix, a clearly defined diagonal indicates a low amount of confusion between classes, which is a positive sign of the model's performance.

Among the non-settlement classes, all achieve a recall score above 75%, except for 'crops' and 'rangeland,' which display comparatively lower recall scores. The decrease in recall performance for these two classes coincides with an increase in the recall performance of 'rural' (and to some extent 'urban'). This is due to the overlap between sparse settlement areas, rangeland, and crops, combined with the weighting in favor of rural and urban classes. This trade-off is expected and necessary to achieve higher recall scores for the rural class in particular.

There is a notable level of confusion between the two settlement classes, rural and urban. Several factors likely contribute to this effect, including shortcomings of the preprocessing scheme, imperfect label quality, and visual overlap between these classes. The upside is that this confusion is mostly contained within the settlement classes. However, while the recall score for the urban class exceeds the target accuracy of 75%, the recall score for the rural class falls short at 65%.

Precision. Precision measures the percentage of pixels classified as a specific class that truly belong to that class, reflecting the proportion of predicted pixels for 'class A' that are correctly identified as 'class A'. This metric indicates the reliability of the predictions for that class. Precision can be derived from the confusion matrix by normalizing the diagonal values of each column by the sum of the respective columns. Alternatively, precision can be calculated using the formula 5:

$$precision = \frac{TP}{TP + FP} \tag{5}$$

The metrics of recall and precision are complementary and together provide a comprehensive understanding of classification performance. Ideally, both metrics should indicate strong performance. However, depending on the specific use case and data characteristics, emphasis might be placed on one metric over the other. Relying on a single metric can lead to a skewed

evaluation of performance, especially in the presence of class imbalance. For example, a high recall is particularly valuable when the goal is to identify as many true positives as possible, ensuring that as few relevant cases as possible are missed.

Precision varies between classes, with the model performing best on water, tree, bare ground, and rangeland. Table 7 reveals a less clearly defined diagonal compared to the recall matrix, with flooded vegetation and rural exhibiting the poorest performance. For example, flooded vegetation is often confused with rangeland, and rural shows similar confusion. Additionally, crops are frequently mistaken for rangeland. Most of the confusion involves rangeland being misclassified as flooded vegetation, crops, or rural.

The precision matrix reveals a less distinct diagonal compared to the recall matrix. While most classes achieve a precision above 0.5, some classes show poorer performance. Rural, in particular, exhibits the lowest precision. Normally, such results would be concerning, but given the data properties and the applied weighting scheme, these outcomes align with expectations. The precision matrix indicates that much of the land predicted as rural by the model is labeled as rangeland and crops. This is reasonable because the model has learned to cluster all elements of sparse rural areas, including other classes, to classify settlements effectively.

Another class with notably low precision is flooded vegetation. This is partly due to the aggressive weighting and the class being one of the smallest. Additionally, the nature of the flooded vegetation class—primarily a variant of tree and rangeland with permanent or seasonal flooding—contributes to the low precision. This class is also estimated to have the lowest precision in the ESRI LULC dataset (see Table 2), which likely exacerbates the low precision observed in our model. Crops, too, show considerable confusion with rangeland. This confusion is logical given the visual similarity between the two classes and the seasonal nature of crops, which makes them harder to identify accurately.

IoU and F1. Intersection over Union (IoU) and the F1-score integrate both recall and precision into a single, objective metric, providing a comprehensive measure of model performance. The IoU score ranges from zero to one, with values closer to one indicating better performance. An IoU score of at least 0.5 is typically regarded as satisfactory[42]. Similarly, the F1-score, which also ranges from zero to one, reflects the harmonic mean of precision and recall, with higher values denoting superior model performance. The IoU and F1-score are positively correlated, meaning that improvements in one metric generally correspond to improvements in the other. Both metrics can be calculated for each class individually or as an average across all classes. These metrics provide a balanced evaluation by combining the insights from both precision and recall, making them robust indicators of model performance. The formulas for calculating IoU and F1-score are available in formulas 6 and 7 respectively.

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{6}$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

IoU and F1-score provide a more condensed view of performance, combining aspects of both recall and precision. Because both recall and precision affect the IoU and F1 scores, high performance in both metrics is necessary to achieve a high value. Table 8 shows the scores for all classes, with the majority reaching an acceptable level of 0.5 or above. However, the classes flooded vegetation, crops, and rural stand out for having the lowest scores among the classes. This can be connected to their poor performance in precision.

Country-based evaluation. Figure 7(a) further illustrates the recall performance variation between countries, independent of country size. 'Water' remains the most stable class, while other classes exhibit more varied score distributions. 'Bare ground' shows the most notable variation, likely due to its differing percentage across countries, especially in the desert areas of Northern Africa. Errors in predicting 'bare ground' in countries with minimal bare ground significantly affect the metric. 'Crops' also show considerable variation, with about a fourth of its values in the lower spectrum, likely due to differences in crop percentage and arrangement across countries.

Figure 7(b) illustrates the variation in precision performance across countries. The three classes with the highest precision performance—'water,' 'tree,' and 'rangeland'—exhibit stable variation, while 'bare ground' shows slightly more variation. 'Flooded vegetation' and 'rural' consistently demonstrate the lowest precision, with relatively low variance across countries. This consistency suggests that these classes are particularly challenging to classify accurately.

Figure 7(c,d) shows the distribution of scores among countries in greater detail, reflecting the IoU, and F1-score performance quite well. Notably, the class 'bare ground,' which otherwise has a high IoU, and F1-score, shows significant variation. This can likely be attributed to the larger countries, which have a greater influence on the total metric, predominantly achieving higher scores for this class.

The IoU and F1-scores provide a comprehensive overview of recall and precision by combining both metrics. A value of at least 0.5 is generally considered successful. Both class-wise metrics and their means were gathered. The worst-performing classes according to IoU and F1 are 'flooded vegetation,' 'crops,' and 'rural,' all of which exhibited notable reductions in precision. Other classes achieved good results, often well above 0.5.

Urban/Rural Evaluation using DHS. One downstream task of the HUR map involves leveraging it as prior information, addressing the geographical displacement in the DHS dataset. More generally, a data user may want to use the map to determine whether a neighborhood, cluster, or village would be classified in a DHS survey as an urban or rural location. Therefore, it is desirable that our map's urban-rural classification align with that of the DHS. However, according to DHS guidelines, the DHS do not themselves classify regions of a surveyed country as urban or rural. Instead, they rely on the classifications provided by the country's government or statistics office. These definitions are country-specific and may vary from one country to another.

Given this, we utilize recent DHS survey points to label nearby human settlements as either rural or urban without the complexity of classifying these points ourselves. This approach avoids the challenge of categorizing locations according to the urban-rural definitions used by the DHS, which is beyond the scope of our work. The goal is to enable the data user to determine whether a neighborhood, cluster, or village would be classified in a DHS survey as an urban or rural location. In constructing this validation dataset, our goal was to ensure the distribution of validation locations aligns with potential use cases of the map. The validation dataset should also contain examples from all three classes: urban, rural and non-human settlement. For reasons that will become clear we generated validation data separately for non-human settlements and for urban and rural areas, combining these into a single dataset for evaluation.

One potential use case of the map is in assessing the expansion of cities and towns over time. However, it is not feasible to uniformly sample from the boundaries of cities and towns due to the lack of access to precise boundary data, and even where such data exists, defining town boundaries can be complex. However, we can approximate this distribution using DHS-perturbed points that do not fall within human settlement areas according to JRC HS map. These points are approximately uniformly distributed near city/town boundaries because DHS cluster locations are representative of towns, cities, and villages, and because the privacy perturbation is applied uniformly randomly in any direction. Therefore, we will use these points as validation data for non-human settlements class.

For urban and rural validation data, we take the perturbed DHS points collected after 2013 and impute their plausible unperturbed locations. Specifically, for each DHS cluster $k = 1, \ldots, K$, we compute 20 imputed locations $g_{k1}, \ldots, g_{k20}$ as draws from a Bayesian posterior distribution. (This posterior is constructed from a uniform prior on the human-settlement pixels of the JRC HS map and a likelihood given by the DHS perturbation procedure.) In constructing our validation data, we rely on the JRC HS map to inform our Bayesian prior in imputing the locations of DHS points. However, it is important to acknowledge that the JRC HS map is not ground truth– it is an estimate.

The map classifies each of these locations as urban, rural, or non-human settlement. Let $f(g_k)$ denote the map's classification for location g_k. The 20 imputations are aggregated into a single category using majority voting, denoted $\bar{f}(g_{k\cdot})$, representing the most frequent category among $f(g_{k1}), \ldots, f(g_{k20})$ (with ties split randomly). The urban/rural validation dataset consists of pairs $\{(\bar{f}(g_{k\cdot}), y_k)\}_{k=1}^{K}$, where y_k is the DHS's urban/rural classification of cluster k. In this context, $\bar{f}(g_{k\cdot})$ is our HUR predicted classification, and y_k is the true classification.

This validation dataset is approximately uniformly distributed across urban and rural locations in Africa because the unperturbed DHS cluster locations are uniformly distributed across these areas. The evaluation is shown in Figure 8. Our HUR map provides yearly map covers the period from 2016 to 2022, while the GHS SMOD provides maps at five-year intervals from 1975 to 2030. Thus, comparing our HUR 2016 map to the 2015 SMOD map, and our HUR 2020 map to SMOD 2020, is a fair comparison. However, as illustrated in Figure 8, there is no significant performance difference between these two time points.

At the continental scale, the overall accuracy of our map is 65%, representing a substantial improvement over the GHS SMOD map, which has an accuracy of 57%. The Kappa coefficient for our map is 47%, also higher than the 38% for the GHS SMOD map.

In the country-wise comparison, our map generally outperforms the GHS SMOD map in most countries. However, in countries where the accuracy of the GHS SMOD map is particularly low—such as Burundi, Egypt, Malawi, Rwanda, and Uganda—our map performs significantly better. This underscores the finding that using a deep learning model to generate rural/urban maps at a continental scale is a generalizable method that produces consistent results across different countries.

As a benchmark, we compared our map to the GHS SMOD map. However, the GHS map uses different definitions for urban, rural, and non-human settlement areas than those in our validation dataset. Therefore, we are working with three distinct sets of definitions: (1) those used to create the DHS-based validation data, (2) those used in our HUR map, and (3) those used in the GHS SMOD map. The improved accuracy of our HUR map over the GHS SMOD map may partly result from a closer alignment between definitions 1 and 2 as compared to between 2 and 3. Nevertheless, we have demonstrated that our map is a better choice for a data user interested in determining urban-rural locations under definition 1 – such as in the context of predicting DHS urban-rural classifications. Moreover, since our training set is also derived from the GHS SMOD map, there

is good reason to believe definitions 2 and 3 are closely aligned. This highlights that the deep learning models we trained to generate the map are more closely aligned with country-specific definitions of rural and urban areas than the GHS SMOD map.

Code availability

The results were produced using Python (v3.10.12) on the National Academic Infrastructure for Supercomputing in Sweden (NAISS) and Google Earth Engine (GEE). More information about the scripts are available on GitHub in the repository of the AI and Global Development Lab: https://github.com/AIandGlobalDevelopmentLab/Africa-Rural-Urban-Map.

References

1. Karra, K. *et al.* Global land use/land cover with sentinel 2 and deep learning. In *2021 IEEE international geoscience and remote sensing symposium IGARSS*, 4704–4707 (IEEE, 2021).

2. Zanaga, D. *et al.* Esa worldcover 10 m 2021 v200. - (2022).

3. Arevalo, P. *et al.* Global land cover mapping and estimation yearly 30 m v001. *Distributed by NASA EOSDIS Land Process. DAAC* (2022).

4. Mirmazloumi, S. M. *et al.* Elulc-10, a 10 m european land use and land cover map using sentinel and landsat data in google earth engine. *Remote. Sens.* **14**, 3041 (2022).

5. Jerzak, C. T., Johansson, F. & Daoud, A. Image-based treatment effect heterogeneity. *arXiv preprint arXiv:2206.06417* (2022).

6. Ratledge, N., Cadamuro, G., de la Cuesta, B., Stigler, M. & Burke, M. Using machine learning to assess the livelihood impact of electricity access. *Nature* **611**, 491–495 (2022).

7. Daoud, A. Unifying studies of scarcity, abundance, and sufficiency. *Ecol. Econ.* **147**, 208–217 (2018).

8. Daoud, A., Halleröd, B. & Guha-Sapir, D. What is the association between absolute child poverty, poor governance, and natural disasters? a global comparison of some of the realities of climate change. *PLoS one* **11**, e0153296 (2016).

9. Balgi, S., Pena, J. M. & Daoud, A. Personalized public policy analysis in social sciences using causal-graphical normalizing flows. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, 11810–11818 (2022).

10. Amani, M. *et al.* Application of google earth engine cloud computing platform, sentinel imagery, and neural networks for crop mapping in canada. *Remote. Sens.* **12**, 3561 (2020).

11. Pesaresi, M. *et al.* *Operating procedure for the production of the Global Human Settlement Layer from Landsat data of the epochs 1975, 1990, 2000, and 2014* (Publications Office of the European Union Luxembourg, 2016).

12. Sheykhmousa, M. *et al.* Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **13**, 6308–6325 (2020).

13. Zheng, Y., Tang, L. & Wang, H. An improved approach for monitoring urban built-up areas by combining npp-viirs nighttime light, ndvi, ndwi, and ndbi. *J. Clean. Prod.* **328**, 129488 (2021).

14. Zhu, X. X. *et al.* Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE geoscience remote sensing magazine* **5**, 8–36 (2017).

15. Diakogiannis, F. I., Waldner, F., Caccetta, P. & Wu, C. Resunet-a: A deep learning framework for semantic segmentation of remotely sensed data. *ISPRS J. Photogramm. Remote. Sens.* **162**, 94–114 (2020).

16. Kakooei, M. & Baleghi, Y. Spatial-temporal analysis of urban environmental variables using building height features. *Urban Clim.* **52**, 101736 (2023).

17. Pettersson, M. B., Kakooei, M., Ortheden, J., Johansson, F. D. & Daoud, A. Time series of satellite imagery improve deep learning estimates of neighborhood-level poverty in africa. In *IJCAI*, 6165–6173 (2023).

18. Daoud, A. *et al.* Using satellite images and deep learning to measure health and living standards in india. *Soc. Indic. Res.* **167**, 475–505 (2023).

19. Daoud, A. & Dubhashi, D. Statistical modeling: the three cultures. *arXiv preprint arXiv:2012.04570* (2020).

20. Kino, S. *et al.* A scoping review on the use of machine learning in research on social determinants of health: Trends and research prospects. *SSM-population Heal.* **15**, 100836 (2021).

21. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 234–241 (Springer, 2015).

22. Wang, H., Cao, P., Wang, J. & Zaiane, O. R. Uctransnet: rethinking the skip connections in u-net from a channel-wise perspective with transformer. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, 2441–2449 (2022).

23. Chantharaj, S. *et al.* Semantic segmentation on medium-resolution satellite images using deep convolutional networks with remote sensing derived indices. In *2018 15th International joint conference on computer science and software engineering (JCSSE)*, 1–6 (IEEE, 2018).

24. Badrinarayanan, V., Kendall, A. & Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis machine intelligence* **39**, 2481–2495 (2017).

25. Garg, R., Kumar, A., Bansal, N., Prateek, M. & Kumar, S. Semantic segmentation of polsar image data using advanced deep learning model. *Sci. Reports* **11**, 15365, 10.1038/s41598-021-94422-y (2021).

26. Boonpook, W. *et al.* Deep learning semantic segmentation for land use and land cover types using landsat 8 imagery. *ISPRS Int. J. Geo-Information* **12**, 14, 10.3390/ijgi12010014 (2023).

27. Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Gill, E. & Molinier, M. A new fully convolutional neural network for semantic segmentation of polarimetric sar imagery in complex land cover ecosystem. *ISPRS J. Photogramm. Remote. Sens.* **151**, 223–236, https://doi.org/10.1016/j.isprsjprs.2019.03.015 (2019).

28. Hamida, A. B. *et al.* Deep learning for semantic segmentation of remote sensing images with rich spectral content. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2569–2572 (IEEE, 2017).

29. Gorelick, N. *et al.* Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote. sensing Environ.* **202**, 18–27 (2017).

30. Amani, M. *et al.* Google earth engine cloud computing platform for remote sensing big data applications: A comprehensive review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* **13**, 5326–5350 (2020).

31. Kansakar, P. & Hossain, F. A review of applications of satellite earth observation data for global societal benefit and stewardship of planet earth. *Space Policy* **36**, 46–54 (2016).

32. Young, N. E. *et al.* A survival guide to landsat preprocessing. *Ecology* **98**, 920–932, https://doi.org/10.1002/ecy.1730 (2017). https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1002/ecy.1730.

33. Hall, R. C. *A history of the military polar orbiting meteorological satellite program* (Office of the Historian, National Reconnaissance Office, 2001).

34. Shi, K. *et al.* Evaluating the ability of npp-viirs nighttime light data to estimate the gross domestic product and the electric power consumption of china at multiple scales: A comparison with dmsp-ols data. *Remote. Sens.* **6**, 1705–1724 (2014).

35. Yeh, C. *et al.* Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nat. communications* **11**, 1–11 (2020).

36. Chen, J. *et al.* Global land cover mapping at 30 m resolution: A pok-based operational approach. *ISPRS J. Photogramm. Remote. Sens.* **103**, 7–27 (2015).

37. Chen, L.-C., Papandreou, G., Schroff, F. & Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587* (2017).

38. Karasiak, N., Dejoux, J.-F., Monteil, C. & Sheeren, D. Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. *Mach. Learn.* **111**, 2715–2740, 10.1007/s10994-021-05972-1 (2022).

39. Rolf, E. *et al.* A generalizable and accessible approach to machine learning with global satellite imagery. *Nat. Commun.* **12**, 4392, 10.1038/s41467-021-24638-z (2021).

40. Pfeuffer, A., Schulz, K. & Dietmayer, K. Semantic segmentation of video sequences with convolutional lstms (2019). 1905.01058.

41. Kakooei, M., Bailie, J., Söderberg, A., Becevic, A. & Daoud, A. Africa Rural-Urban Map, 10.7910/DVN/YFRECD (2024).

42. Dai, J., He, K. & Sun, J. Instance-aware semantic segmentation via multi-task network cascades. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3150–3158 (2016).

43. Karra, K. *et al.* Methodology & accuracy summary 10m global land use land cover maps. Tech. Rep., Impact Observatory (2022).

44. Applying the degree of urbanisation: a methodological manual to define cities, towns and rural areas for international comparisons (2021).

Acknowledgements

Author contributions statement

Conceptualization, M.K. and A.D.; Methodology, A.S., A.B., J.B. and M.K.; Investigation, A.S., A.B., M.K., and A.D.; Writing—original draft, M.K.; Writing—review and editing, M.K., J.B. and A.D.; Supervision, M.K., and A.D.; Funding acquisition, A.D. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Figures & Tables

Table 1. Landcover datasets

	Dataset	Resolution(m)	Temporality
1	ESA WorldCover	10 m	2020, 2021
2	ESRI Land Use/Land Cover	10 m	2018-2022
3	NASA GlanCE-v001	30 m	2001-2019
4	GlobeLand30	30 m	2000, 2010, 2020
5	From-GLC	30/10 m	2010&2015/2017
6	GlobCover	300 m	2009

Table 2. ESRI LULC Accuracy Assessment - Per-class Recall and Precision, recreated from[43]

	Precision[%]	Recall[%]
water	87,2	90,1
trees	82,3	85,4
flooded veg.	57,3	53,9
crops	90,2	72,1
built-area	79,6	94,5
bare ground	72,1	38,2
rangeland	57,8	70,5

Table 3. The classification scheme used for the GHS-SMOD product[44]. Which class a 1×1 km cell is given is based both on the population density of the cell itself and the total population size of all neighboring cells of the same density category. According to this scheme, orange cells are considered as urban, while green cells are considered as rural.
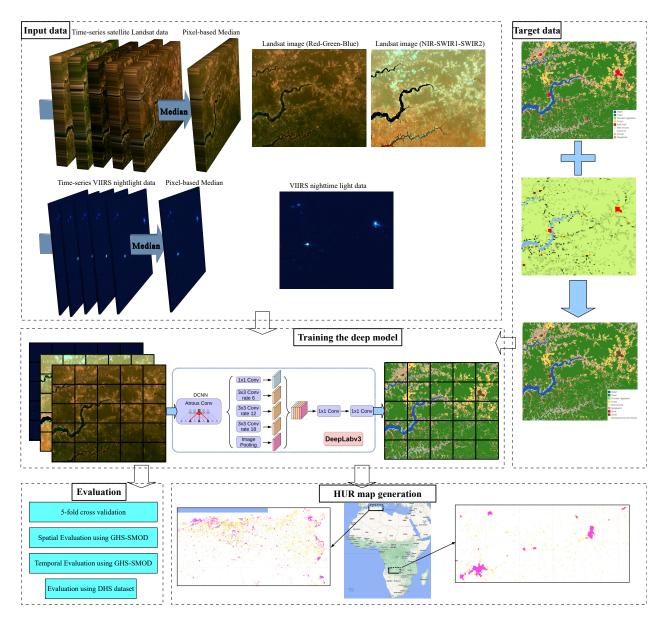
Population density of cells, inhabitants per km²		Population size thresholds of the cluster of cells (settlement size)			No population size criterion (not a settlement)
		≥ 50 000	5 000 − 49 999	500 − 4 999	
	≥ 1 500	Urban centers	Dense urban clusters		
	≥ 300		Semi-dense urban clusters	Rural clusters	Suburban or peri-urban grid cells
	≥ 50				Low-density rural grid cells
	< 50				Very low-density rural grid cells

Table 4. Data bands used by the model

	Band	Type	Resolution
1	Blue	Landsat	30 m
2	Green	Landsat	30 m
3	Red	Landsat	30 m
4	Near Infrared (NIR)	Landsat	30 m
5	Shortwave Infrared 1 (SWIR1)	Landsat	30 m
6	Shortwave Infrared 2 (SWIR2)	Landsat	30 m
7	Nightlights	VIIRS	464 m
8	GHS-SMOD 2020	Human Settlement Map	1000 m
9	ESRI LULC 2018-2020	Land Cover Map	10 m

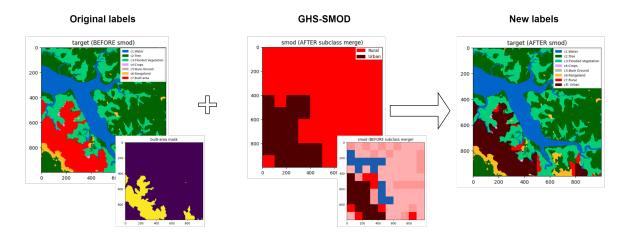**Table 5.** Target labels after preprocessing (-1 denotes an ignored class)

	Class	Label
1	Water	0
2	Trees	1
3	Flooded Vegetation	2
4	Crops	3
5	Bare Ground	4
6	Rangeland	5
7	Rural	6
8	Urban	7
9	Missing + Snow/Ice + Clouds	-1

Figure 1. A flowchart overview of the methodology, highlighting the different sections of the process. The "Input Data" section provides the satellite imagery and other data used to train the deep learning models. The "Target Data" section integrates the LULC map with rural and urban classifications to create the target dataset. The "Training the Deep Model" section utilizes both the input and target data to train the deep learning model. The trained model is then evaluated both at a country level and annually over time. The final output is a high-resolution map generated at a continental scale.

Figure 2. Data gathering from Sierra Leone; (a) the points from which images are sampled; (b) the image tiles covering the country. (c) A visualization of the format in which data is fed to the model.

Figure 3. Description: Label augmentation, as applied to one tile. Information from ESRILULC and GHS-SMOD is combined to enhance the 'Built-Area' class. **Left**: Shows original labels, with bright red being 'Built-Area'. Only the area marked by the 'Built-Area' mask is affected. **Center**: Shows the GHS-SMOD labels for this tile, with dark red being urban and bright red being rural. The small tile shows the original eight-class GHS-SMOD dataset, before merging to urban and rural. **Right**: The resulting labels of this tile after preprocessing.
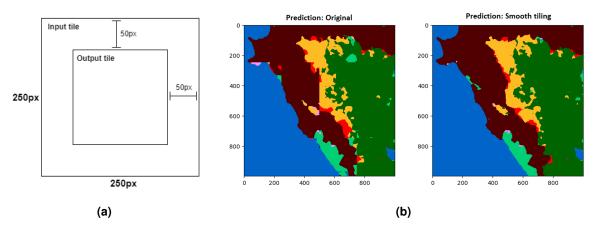

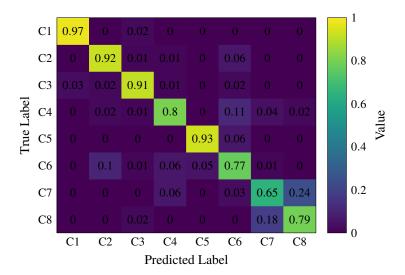**Figure 4.** The class distribution of ESRI LULC 2020 across Africa

Figure 5. The 5 folds in which countries are split, highlighted in different colors (Fold1:Blue, Fold2:Red, Fold3:Green, Fold4:Yellow, Fold5:Purple). Screenshot taken in Google Earth Engine

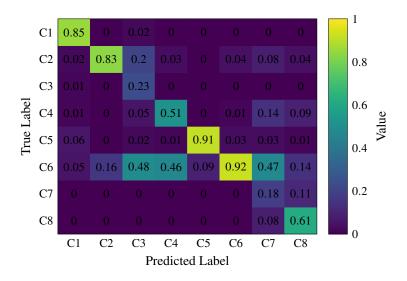(a)   (b)

Figure 6. (a) Shows the input-/output-dimensions of the tiling scheme. Inference is done over the entire image (large square) but only the center of prediction (smaller square) is kept. (b) A comparison of original prediction (left) and with tile smoothening (right). In the original image, clear edge artifacts are visible, while in the smoothened image it appears more cohesive.

Table 6. Fold average recall matrix. Confusion matrix divided by sum of columns. The classes are C1:Water, C2:Tree, C3:Flooded vegetation, C4:Crops, C5:Bare Grounds, C6:RangeLands, C7:Rural, C8:Urban

Table 7. Fold average precision matrix. Confusion matrix divided by sum of rows. The classes are C1:Water, C2:Tree, C3:Flooded vegetation, C4:Crops, C5:Bare Grounds, C6:RangeLands, C7:Rural, C8:Urban

Table 8. Fold average IoU and F1-scores

Class	Water	Tree	Flooded veg	Crops	Bare Ground	Rangeland	rural	Urban	Mean
IoU	0.831	0.777	0.223	0.449	0.855	0.720	0.163	0.520	0.567
F1-score	0.908	0.875	0.364	0.620	0.922	0.837	0.280	0.684	0.686

(a)

(b)

(c)
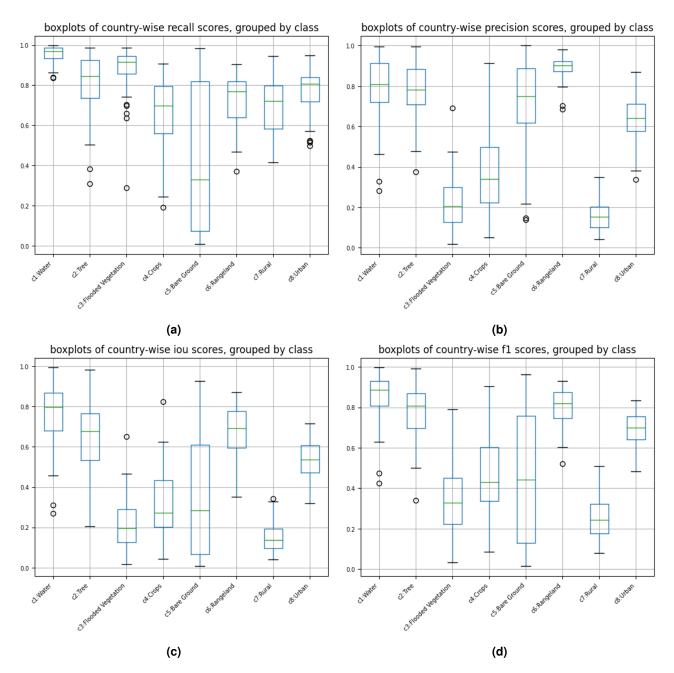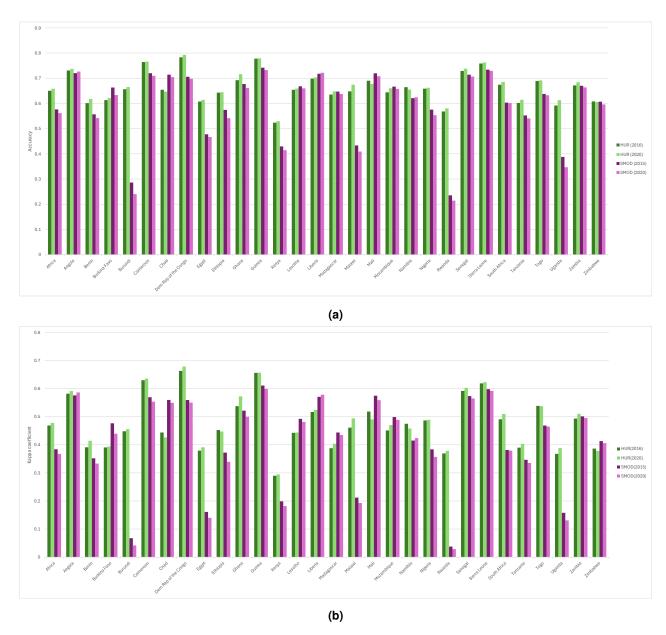

(d)

Figure 7. Country-wise performance boxplots in terms of (a) Recall, (b) Precision, (c) IoU, and (d) F1-score.

(a)

(b)

Figure 8. Country-wise comparison of our map and the GHS SMOD map based on the DHS survey evaluation: (a) Accuracy, and (b) Kappa coefficient.