Mapping Refugee Camps with AI: A Benchmark Dataset and Baseline Models for Humanitarian Applications

Amrita Gupta[1]*, Anthony Ortiz[1], Simone Fobi Nsutezo[1], Duncan Kebut[2], Seema Iyer[3], Rahul Dodhia[1], Juan M. Lavista Ferres[1]

[1]Microsoft AI for Good Lab [2]Humanitarian OpenStreetMap Team [3]USA for UNHCR

Abstract

Over 6.6 million people worldwide live in refugee camps, most of which lack comprehensive, up-to-date maps. This hinders effective resource distribution, infrastructure planning, and disaster response in these environments. Automated mapping with aerial imagery offers a promising solution, capturing the detail needed for effective camp management, but it requires datasets that reflect the distinct characteristics of refugee camps. Existing building footprint datasets focus on urban or semi-urban areas leaving refugee camps–characterized by irregular layouts, diverse building sizes, and varied materials–underrepresented and poorly served by current models. This study introduces the KAKUMAAERIAL *dataset, an open-source resource for humanitarian mapping. It pairs high-resolution aerial imagery from the Kakuma-Kalobeyei refugee camps in Kenya with annotations for buildings, solar panels, roof materials, and sanitation facilities. The dataset serves as a resource for benchmarking models on tasks crucial to humanitarian aid. Baseline machine learning models achieved strong performance on key tasks: building and solar panel segmentation (IoU of 0.848 and 0.813, respectively), roof material classification (accuracy of 85.6%), and toilet identification (accuracy of 97.8%). By applying these models to broader areas within the camps, the study provides actionable insights into camp infrastructure, including energy access and sanitation availability. This research demonstrates how geospatial technologies and machine learning can enable humanitarian organizations to improve operational efficiency while improving the living conditions and dignity of displaced populations.*

1. Introduction

Digital mapping technologies have made geospatial data more accessible and actionable than ever before. However, many parts of the world—often home to the most vulnerable populations—remain incomplete, outdated, or entirely absent in existing maps [9–11]. Refugee settlements are a striking example of these gaps [24]. As of mid-2024, the United Nations High Commissioner for Refugees (UNHCR) estimated that 122.6M people were forcibly displaced worldwide due to conflict, persecution, disasters, and other humanitarian crises [23], with many seeking refuge in either managed or self-settled camps. These camps are established rapidly and reactively to provide immediate support, often in remote or undeveloped locations lacking infrastructure and detailed maps to begin with. As a result, these camps are often unplanned and highly dynamic, making it difficult to keep accurate, up-to-date information on where people are and what resources are available. Without this information, government agencies, humanitarian groups, and relief teams cannot fully coordinate efforts to support underserved or rapidly changing areas, increasing the risk of resource misallocation and missed opportunities to improve living conditions for refugees.

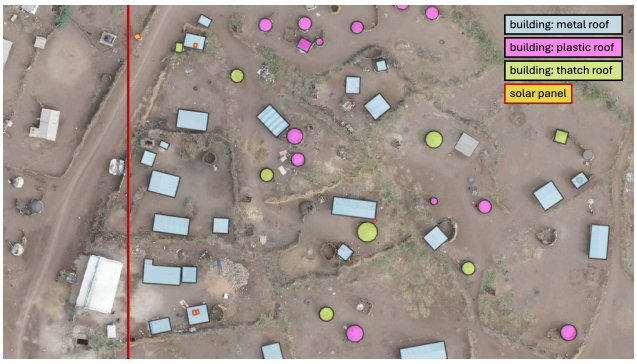


Figure 1. Raw aerial imagery (left) augmented with AI-generated overlays (right) showing delineated buildings, solar panels, and predicted attributes (e.g., roof material, sanitation facilities). Such overlays transform raw imagery into actionable information for infrastructure planning and humanitarian aid.

*Corresponding author: amritagupta@microsoft.com

To provide the spatial data needed for effective humanitarian action, traditional mapping in refugee camps relies on on-the-ground surveys, where field teams use Geographic Information Systems (GIS) tools to record buildings, roads, and essential facilities like water points and clinics. While this approach delivers accurate information, its labor-intensive and time-consuming nature has spurred interest in using remote sensing imagery to accelerate mapping. Satellite data has been widely used in humanitarian mapping, such as through mapathons organized by Missing Maps and HOT, where remote volunteers delineate buildings and roads visible in satellite imagery to fill in missing maps [9]. Drone and aerial imagery are increasingly used [1, 2, 15] due to their high resolution, capturing details essential for distinguishing closely spaced structures and smaller features like electricity infrastructure and sanitation facilities. However, visual inspection and manual delineation of significant volumes of imagery represent a time-consuming and expensive labeling workload. This has spurred growing interest in using machine learning to automatically predict critical features from imagery, offering a scalable solution to support humanitarian mapping.

There is a considerable body of prior work on building footprint extraction from satellite and aerial imagery, supported by numerous open datasets and machine learning challenges. Global footprint datasets, such as Google's Open Buildings [17], Microsoft Buildings [21], and OpenStreetMaps, provide valuable data with worldwide coverage, but lack the accompanying imagery needed for direct model training. In contrast, several high-resolution datasets do provide paired imagery and labeled building footprints and serve as popular benchmarks for model evaluation ([13], [12], [25], [5]). However, these are typically limited to only a few specific urban environments, limiting their generalizability to non-urban or informal settlements.

Mapping refugee camps poses unique challenges compared to other settlements, both in terms of structure and the types of information needed to support humanitarian efforts. Buildings in camps are often small, densely packed, and may be constructed from locally available materials, making them harder to differentiate from their surroundings [15]. Additionally, camp layouts and building characteristics vary widely across regions, which complicates model generalizability [27]. A few studies have focused on mapping building footprints specifically within refugee camps using satellite [4, 6] and drone [2] imagery. Beyond building footprints, roof material [14, 18] and building type [6] are also important details for assessing shelter durability and camp development. Other studies have leveraged remote sensing data to provide further insights on camp properties like spatial layout [27] and electricity access [26]. While these are valuable contributions, none of the datasets used in prior works on refugee camp mapping are made openly available in a machine learning-ready format. This lack of standardized, openly accessible domain-specific datasets creates a significant barrier to reproducing results, benchmarking new methods, and conducting comparative studies on these tasks.

To address this need, our work makes the following key contributions:

- **High-resolution geospatial dataset for humanitarian mapping:** We provide an easily accessible machine learning-ready dataset of aerial imagery aligned with enriched OpenStreetMap labels, to lower barriers for researchers and practitioners to engage with mapping refugee camps. The KAKUMAAERIAL dataset, introduced in this paper, is publicly available on Zenodo [7].

- **Standardized dataset splits and baseline benchmarks:** We define standardized training, validation, and test splits to facilitate consistent model training and evaluation on a benchmark dataset. Additionally, we provide baseline model performance for key tasks, namely building and solar panel segmentation, roof material classification, and toilet identification, showcasing how well simple methods perform and establishing a reference point for future improvements.

- **Large-scale predictions over the Kakuma-Kalobeyei settlement area:** We provide predictions for buildings, roof material, solar panel presence, and toilets across an additional set of drone-imaged scenes outside the core benchmark dataset. These predictions offer a practical demonstration of model utility for scaling analysis across broader areas and supporting humanitarian organizations in planning and resource allocation.

2. The KAKUMAAERIAL Dataset

2.1. Aerial Imagery

The aerial imagery presented in this paper is derived from a broader humanitarian drone mapping effort conducted in November 2022 by the Humanitarian OpenStreetMap Team (HOT), the United Nations High Commissioner for Refugees (UNHCR) Kenya, and Kenya Red Cross Society (KRCS). During this effort, unmanned aerial vehicle missions using a WingtraOne drone equipped with a Sony RX1R II camera were flown over roughly 76 square kilometers of the Kakuma-Kalobeyei area. This resulted in 95 high-resolution orthophoto scenes, with resolutions ranging from 5 to 15 cm per pixel. These scenes, processed by HOT and made publicly available on the OpenAerialMap platform, provide detailed visual data of the region. For our core benchmark dataset, we focus on 11 of these scenes

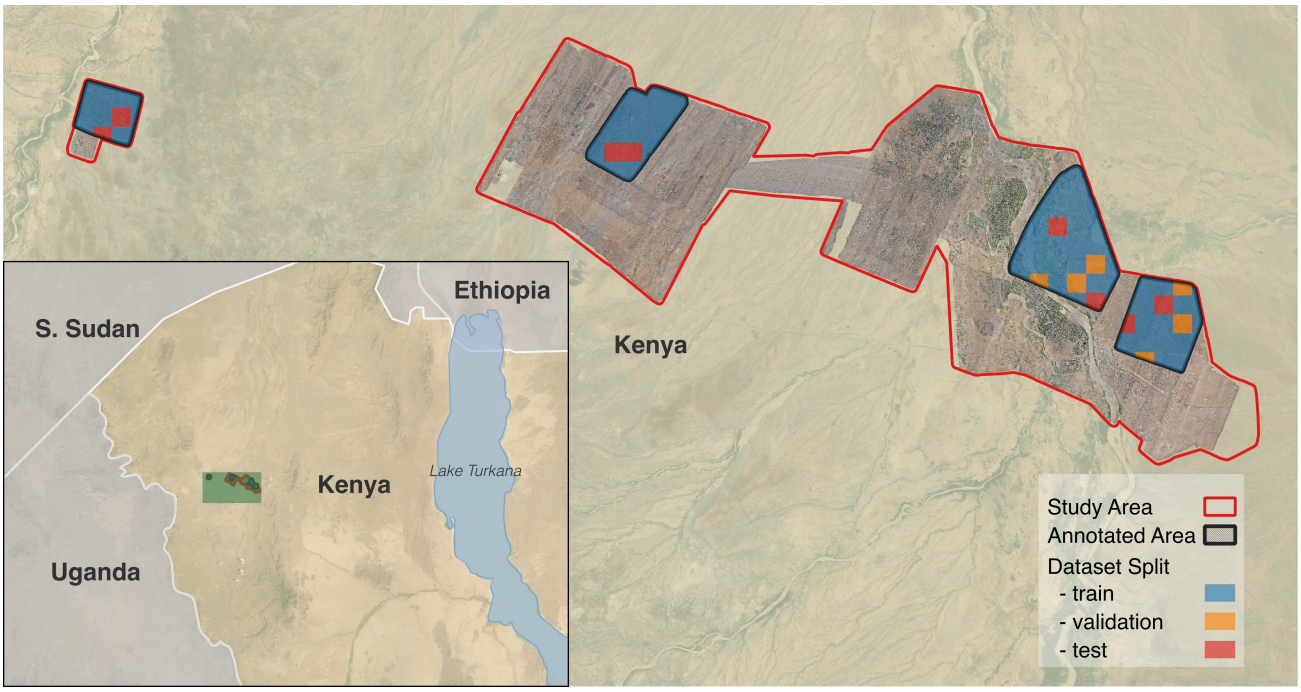


Figure 2. Study area in Kakuma and Kalobeyei refugee camps, Kenya. Red outlines indicate regions where aerial imagery was collected. Black crosshatched areas represent annotated regions belonging to the core benchmark dataset, divided into tiles for training (blue), validation (orange), and test (red) splits. The inset map shows the study area's location within the country of Kenya and its proximity to neighboring countries.

(listed in Table 4 in Appendix Section A.1) over parts of Kalobeyei Town, Kalobeyei Village 2, Kakuma Camp 1, and Kakuma Town that were specifically targeted for a participatory mapping effort with camp residents.

While the imagery provided was both georeferenced and orthorectified, some discontinuities were observed, particularly at scene boundaries. To mitigate this, we applied additional georeferencing adjustments in QGIS, using both Google Satellite basemaps and visible features in the imagery itself to improve alignment. By selecting control points visible across overlapping scenes or between scenes and basemaps, we adjusted the georeferencing to reduce visual misalignments between scenes. Following this, we clipped overlapping areas of the scenes to minimize discontinuities at the object level before mosaicking the scenes using GDAL. These post-processing steps resulted in a more visually coherent imagery dataset.

2.2. Annotations

The 11 aerial imagery scenes that form the KAKUMAAERIAL benchmark dataset were annotated through a participatory mapping effort led by the Humanitarian OpenStreetMap Team (HOT). Participatory mapping [22] directly involves local communities in the mapping and tagging process to ensure the geospatial data reflects their lived realities and needs. In this project, camp residents were engaged to interpret drone imagery, using shape, color, and patterns to identify and annotate critical infrastructure features, including buildings, solar panels, and other facilities. Their detailed, context-specific insights validated and enriched the dataset, which was further enhanced by remote contributions from the broader HOT network and volunteers, resulting in a comprehensive and accurate mapping effort tailored to the needs of the community. These annotations are accessible in OpenStreetMap (OSM). Annotations for the additional 84 OpenAerialMap scenes were derived using AI-assisted predictions refined and manually verified by the HOT network, OSM Kenya, and other volunteers. These annotations are also now publicly accessible in OSM.

To access the complete set of annotations for our work, we queried OSM data as of October 1, 2024 using the Overpass API via the `osmnx` Python package. The OSM query focused on retrieving tags relevant to camp buildings, sanitation facilities, and solar panels (details can be found in Appendix Section A.2). The number of solar panels, buildings, and toilet polygon features within the core benchmark dataset can be found in Table 1. Roof material information can be found in tags for a subset of the buildings. While the OSM features were already well-aligned with the imagery, minor manual adjustments were made to further clean the dataset and improve its reliability for model performance

Feature	Total	Train	Validation	Test
solar	1009	790	80	141
building	12440	9611	849	2106
metal sheet	8139	6587	480	1163
thatch	1572	999	91	496
plastic	1519	1119	180	230
concrete	14	11	0	3
wood	2	1	0	1
mud	1	1	0	0
unlabeled	1193	893	98	211
toilets	1642	1341	57	249

Table 1. Counts of solar panel, building, and toilet instances in the core dataset, with totals and distributions across training, validation, and test splits. Instances on split boundaries may appear in multiple splits, causing split count sums to exceed totals.

evaluations.

2.3. Dataset Splits

Standardized dataset splits were generated by dividing each contiguous imagery region into 500m by 500m tiles to ensure spatial coherence within training, validation, and test regions. From these tiles, 8 tiles (~10%) were randomly assigned to the test set, and 6 tiles to the validation set, with the remaining tiles used for training. The spatially-disjoint tile-based approach reduces the risk of spatial overlap between samples drawn from different splits while aiming to introduce variation in the test and validation sets through the random assignment. Details about the dataset splits are provided on Zenodo [7] to support reproducibility and ease of use as a machine learning benchmark.

2.4. Dataset Use Cases for Mapping in Refugee Camp Contexts

Although many other features are of interest in managing refugee settlements, we focus on the following use cases of the Turkana Camp Mapping Dataset:

Segmentation Buildings and Solar Panels: The initial step involves the creation of maps detailing the spatial extent of buildings from high-resolution drone imagery, which serves as the foundation for subsequent analysis. Identifying buildings themselves informs spatial planning and population estimation, while identifying solar panels helps map energy access for the population.

Classification of Building Roof Material: Once buildings have been identified, our objective extends to classifying the roof materials of the segmented buildings, with implications for the resilience of the structures to environmental conditions.

Identification of Building Sub-types: A third aspect of our problem focuses on the identification of sub-types of built structures, specifically toilets in our work here, with implications for public health and hygiene. Nevertheless, other subtypes can be considered, such as water towers and water tanks.

3. Baseline Experiments

3.1. Building and Solar Panel Segmentation

Given an aerial imagery scene X and corresponding pixel-wise mask Y, our objective is to optimize the parameters of a semantic segmentation model $f(X;\theta) = \hat{Y}$. The value of the mask pixel Y_{ij} represents whether the corresponding imagery X_{ij} is part of a building, solar panel, or neither. The model seeks to minimize the discrepancy between the predicted mask $\hat{Y}$ and the true mask Y.

In the mapping of refugee camps, precise segmentation of buildings and solar panels is critical for effective camp management and resource allocation. Particularly in such environments, buildings can be closely packed or constructed from materials that blend into the surrounding landscape, complicating the segmentation task. To address these specific challenges, we experimented with three distinct segmentation mask variants Y.

Default Mask: This mask classifies each pixel into one of three categories: background, building, or solar panel. It serves as the foundation for our segmentation models, providing a baseline for performance comparison.

Buffered Mask: The spatial arrangement of buildings can vary significantly within refugee camps, with some areas featuring closely spaced structures. This presents a challenge for segmentation models in accurately delineating the edges of adjacent buildings. To address this, we utilize a 'Buffered Mask,' which incorporates a 0.25-meter buffer extending outward from the building edges, thereby creating an additional boundary class. By explicitly marking the boundaries as a separate class, the model is encouraged to pay more attention to these regions during training. This focus is expected to improve the model's ability to distinguish between closely situated buildings by enhancing edge detection and reducing boundary blurring.

Eroded-Buffered Mask: This mask builds upon the concept of the 'Buffered Mask' by not only extending a 0.25-meter buffer outward from each building's edge but also eroding each building inward by an equivalent distance. This dual adjustment creates a comprehensive boundary zone that is 0.5 meters thick, encapsulating both the external and internal edges of the buildings.

To prepare the dataset for baseline model training, we sampled 256x256 pixel image chips from the complete train, validation, and test sets, balancing selection among three categories: chips centered on solar panels, buildings without solar panels, and background regions (including hard negative mining patches including bridges, roads, vehicles, boreholes, bricks, sand, and other construction materials, which were misclassified as buildings in initial model experiments without hard negative background chips). This sampling strategy addresses the underrepresentation of solar panels, which occupy a small fraction of pixels, by increasing their frequency in the training data. Additionally, we ensure that sampled chips do not overlap, maximizing the unique data seen by the models and promoting better generalization across diverse areas within the dataset.

To evaluate our segmentation models, each pixel in the aerial imagery is assigned to the class with the highest output probability, and the resulting classifications are polygonized into vector representations for buildings and solar panels. Performance is assessed using pixelwise IoU and object-level precision and recall, with a 0.5 IoU threshold applied to determine correct object detections.

We conducted an initial comparison of model architectures and encoder backbones. Specifically, we evaluated the performance of U-Net [16] and DeepLabV3+ [3] architectures, integrating them with ResNet-50 [8], ResNeXt-50 32x4d [28], and EfficientNet B3 [20] backbones, using model implementations available through the `torchgeo` package [19]. This analysis aims to identify the most effective combination of model architecture and encoder network for accurately segmenting buildings and solar panels from high-resolution aerial images. During this phase, we exclusively trained models using the default mask using the Jaccard loss to streamline the comparison of model architectures. Results, shown in Table 2, revealed the U-Net with EfficientNet B3 to be the most successful combination among those we compared.

After the architecture search, we extended our experiments to assess the impact of different mask variants and loss functions on model performance. We trained models using each of the three mask designs described above, with separate models trained on Jaccard, focal, and cross-entropy loss functions to evaluate their effectiveness with each mask type. Results are summarized in Table 3 and illustrated in Figure 3. The buffered and eroded-buffered mask variants improved segmentation accuracy for closely packed or adjacent buildings. However, all models showed a tendency to over-segment solar panels, as these often appear as small, closely spaced objects. This behavior contributed to the high pixelwise IoU scores but lower object-level performance metrics for solar panels.

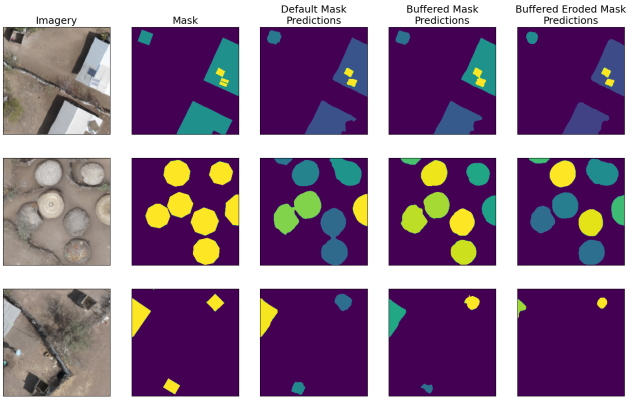


Figure 3. Rows show 256x256 pixel patches from the test set, including imagery, ground truth masks, and predicted segmentation masks from U-Net models with EfficientNet-B3 backbones trained with Jaccard loss on the three mask variants. For the predicted masks, buildings are colored by instance rather than class.

3.2. Roof Material Classification

Given an aerial imagery scene X and corresponding set of building masks $\mathbf{M} = \{M_1, M_2, \ldots, M_n\}$, where each mask M_i corresponds to a building identified within X, our objective is to assign a roof material class label c_i from a set of possible classes $C = \{c_1, c_2, \ldots, c_K\}$. This is achieved by a classification function $g(X, M_i; \phi) = \hat{c}$ parametrized by ϕ.

In refugee camps, identifying roof materials can provide insights into shelter durability and guide resource allocation decisions. The dataset includes three primary roof material classes—metal sheet, plastic, and thatch—which are prevalent in the Kakuma-Kalobeyei region. Severely underrepresented roof material classes (namely "concrete", "wood", and "mud", see Table 1) are grouped into a single "other" class, allowing the baseline model to focus on the most common roof types without introducing noise from infrequent cases.

For model training, we sample 256x256 pixel chips centered on buildings with roof material attribute data, distributing samples evenly across the three main classes to obtain approximately 1400 samples per class. Similar to the approach taken by [18], each chip includes four bands: RGB bands from the aerial imagery and a fourth band representing the building instance mask, allowing the model to focus specifically on the building in question. Chips are assigned to training, validation, and test sets based on the tiles they fall into. A modified ResNet-18 architecture is employed, adapted to accept four-channel inputs. The model is trained using cross-entropy loss, with data augmentation applied to enhance generalization across the classes.

Classification results are shown in Figure 4. The model achieved an overall classification accuracy of 85.59%. The

Model	Backbone	Pixel IoU		Object Precision		Object Recall	
		Buildings	Solar	Buildings	Solar	Buildings	Solar
U-Net	ResNet-50	0.777	0.735	0.622	0.422	0.622	0.700
	ResNeXt-50	0.812	0.772	0.677	0.500	**0.832**	0.700
	EfficientNet B3	**0.848**	**0.813**	**0.830**	**0.615**	0.822	**0.764**
DeepLabV3+	ResNet-50	0.772	0.744	0.664	0.318	0.692	0.671
	ResNeXt-50	0.815	0.756	0.703	0.325	0.772	0.643
	EfficientNet B3	0.775	0.167	0.705	0.013	0.742	0.607

Table 2. Performance metrics for different semantic segmentation model architecture-encoder combinations.

Mask	Loss	Pixel IoU		Object Precision		Object Recall	
		Buildings	Solar	Buildings	Solar	Buildings	Solar
Default	Jaccard	0.848	**0.813**	0.830	0.615	0.822	**0.764**
	Cross Entropy	**0.851**	0.711	0.802	0.255	**0.878**	0.536
	Focal	0.830	0.790	0.743	**0.777**	0.549	0.600
Buffered	Jaccard	**0.851**	0.774	**0.874**	0.396	0.846	0.721
	Cross Entropy	0.822	0.780	0.801	0.404	0.780	0.571
	Focal	0.812	0.735	0.656	0.364	0.781	0.586
Eroded-Buffered	Jaccard	0.830	0.787	0.870	0.500	0.824	0.700
	Cross Entropy	0.778	0.507	0.759	0.169	0.867	0.586
	Focal	0.822	0.757	0.704	0.360	0.773	0.621

Table 3. Performance metrics for different mask types combined with different training loss functions.

primary classification errors occur between the plastic and thatch roof material classes, likely because many buildings incorporate both materials, while only one label is assigned. This overlap creates challenges for both annotators, potentially introducing label noise, and for the model, which struggles to differentiate between the two classes under these conditions. Additionally, there may be relatively few labeled samples featuring rectangular buildings with plastic or thatch roofing, potentially reducing the model's ability to learn these cases.

3.3. Identification of Sanitation Facilities

Given an aerial imagery scene X and corresponding set of building masks $\mathbf{M} = \{M_1, M_2, \ldots, M_n\}$, where each mask M_i corresponds to a building identified within X, our objective is to assign a binary class label c_i encoding whether the building is a toilet. This is achieved by a classification function $g(X, M_i; \phi) = \hat{c}$ parametrized by ϕ.

In the Kakuma-Kalobeyei camps, household toilets are typically unlined pit latrines with a corrugated iron sheet enclosure and no roof—a distinctive feature in aerial imagery. Identifying these facilities is essential for assessing sanitation access and improving public health outcomes. To train a model for this task, we create a pre-chipped dataset of labeled toilet and non-toilet buildings. Specifically, we sample 256x256 pixel chips centered either on buildings tagged as toilets (in the 'amenity' OSM tag) or other buildings, to create a balanced dataset of 1200 samples per class. Again, chips are assigned to training, validation, or test set based on which split the tile they are sampled from belongs to. Then, we train a ResNet-18 model modified to take RGB channels as well as a mask channel as input and to predict two classes as output. We achieved a binary classification accuracy of 97.8% on the 500 test chips (237 toilet, 263 non-toilet).

4. Kakuma-Kalobeyei Mapping Results

To extend the analysis beyond the core benchmark dataset, we ran inference on additional scenes to generate predictions for segmentation masks, building types, and roof materials across a broader area. This effort aims to provide actionable insights for humanitarian applications, such as understanding the distribution and types of structures within the camp, and to demonstrate the scalability of the trained models for mapping refugee camps at a larger scale.

For building and solar panel segmentation, we selected the best-performing U-Net model, which was used to generate pixel-wise prediction masks at the scene level. The building mask for each scene was vectorized to serve as input for subsequent models. The vectorized building masks

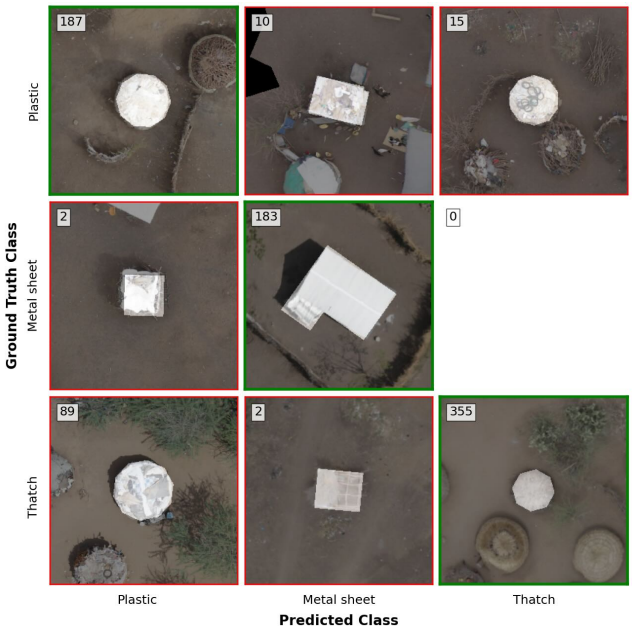


Figure 4. Confusion matrix for roof material classification across the three primary classes. Each subplot (i, j) displays a sample with ground truth label in row i and predicted label in column j. The count for each classification outcome is shown in the upper left corner of each subplot. Note that no sample is displayed at row 2, column 3 because there were no test instances with metal sheet roofing misclassified as thatch roofing.

were first processed by the toilet identification model, followed by the roof material classification model applied to buildings not identified as toilets. The final output consists of vectorized building polygons annotated with predicted attributes, including the presence of solar panels, roof material, and whether the building is a toilet.

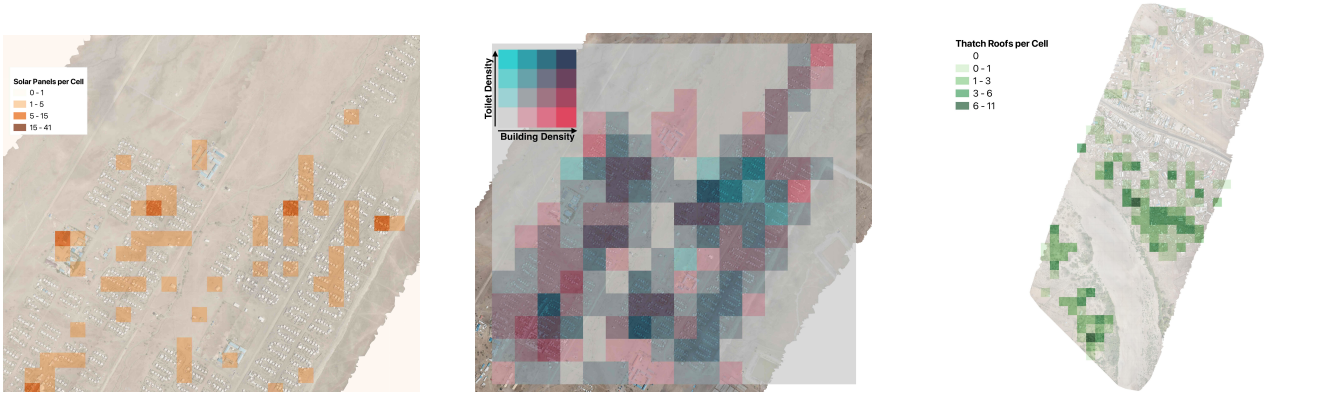
Figure 5 visualizes the spatial distribution of predicted attributes across sections of the extended dataset area. For instance, mapping the spatial density of solar panels highlights built-up areas lacking rooftop solar installations (Figure 5a), which could help aid organizations prioritize these areas for solar energy deployment to improve access to sustainable power. Similarly, combining predictions for buildings and sanitation facilities (Figure 5b) identifies areas with high building density but limited access to toilets, enabling targeted interventions to address sanitation needs. Mapping the distribution of thatch-roofed buildings (Figure 5c reveals high densities of these along the banks of the Tarach River, highlighting areas where residents may be vulnerable to flooding and where efforts to improve housing resilience could be prioritized.

5. Discussion and Conclusion

This study takes an important step toward improving the mapping of refugee camps, which are underrepresented in existing datasets and pose unique challenges due to their unplanned nature and location in remote areas. By developing a high-resolution, open-source dataset complete with machine learning splits, we provide a resource that facilitates broader engagement from the computer vision research community and supports direct comparison of methods and models tailored to refugee camp settings. The second contribution of this study—baseline models for identifying buildings, solar panels, roof materials, and sanitation facilities—provide a foundation for improving approaches on the benchmark dataset. Our results show that even simple segmentation models, such as the U-Net with an EfficientNet-B3 backbone, can effectively map buildings and solar panels with high precision and recall, serving as a strong starting point for further improvement. Additionally, classification models based on the ResNet-18 architecture proved capable of identifying roof materials and sanitation facilities, offering valuable insights into infrastructure resilience and public health conditions. These outputs, applied at scale, provide actionable data for humanitarian organizations, supporting efforts to improve living conditions in refugee camps.

Despite these contributions, there are limitations to this study. The dataset is limited to the Kakuma-Kalobeyei camps, and while the models demonstrate strong performance in this context, generalizability to other camps or informal settlements requires further investigation. Additionally, the reliance on high-quality aerial imagery raises concerns about scalability to regions where such data is unavailable or prohibitively expensive. Moreover, challenges in annotating features like mixed roof materials and sanitation facilities suggest a need for more sophisticated labeling techniques and the integration of domain knowledge.

In this study, commonly-used segmentation and classification models were used as baselines, but several areas remain for further exploration. Architectural improvements like transformers could better capture complex spatial relationships, while multi-task learning across related tasks (e.g., building segmentation and roof material classification) could improve performance by leveraging shared representations. For enhanced segmentation accuracy, loss functions that improve edge detection or post-processing techniques may improve delineation of closely situated buildings. Data quality and efficiency also can be avenues for further work: learning from partial or noisy labels could improve adaptability to real-world datasets, and automated correction of label misalignment could enhance robustness where imagery and label registration is poor. To address class imbalance, methods beyond pre-sampling, such as synthetic data generation, could better represent rare



(a) Number of detected solar panels per 50m x 50m cell.

(b) Number of detected buildings (blues) and toilets (reds) per 100m x 100m cell.

(c) Number of detected thatch roofs per 50m x 50m cell.

Figure 5. Predicted solar panels (a), toilets and buildings (b), and thatch roofed buildings (c) as examples of spatial insights gleaned from AI-generated mapping outputs.

classes. Finally, exploring data volume and quality effects on model performance could support data-efficient training, with self-training and low-shot learning reducing labeling demands.

In conclusion, this study demonstrates the potential of geospatial technologies and machine learning in addressing critical challenges in humanitarian contexts. By bridging critical data gaps, the methods and dataset presented here can empower organizations to make informed decisions, enhance operational efficiency, and ultimately improve the quality of life, safety, and dignity of displaced populations worldwide. Model outputs can streamline mapping workflows for organizations like HOT by integrating AI-generated predictions into tools like RapidID, JOSM, and the HOT Tasking Manager, where community mappers can validate and refine them with local knowledge. This research also supports broader goals, such as municipalizing the Kakuma-Kalobeyei settlement by connecting it to the town of Kakuma's electrical, water, and lighting grids—a process reliant on accurate, comprehensive mapping. Ultimately, this work highlights the importance of continued innovation and collaboration in developing data-driven solutions that improve the quality of life, safety, and dignity of displaced populations worldwide.

References

[1] Nahian Ahmed, Riasad Bin Mahbub, and Rashedur M Rahman. Learning to extract buildings from ultra-high-resolution drone images and noisy labels. *International Journal of Remote Sensing*, 41(21):8216–8237, 2020. 2

[2] Christopher Yan-Chak Chan, Matthias Weigand, Emran Alchikh Alnajar, and Hannes Taubenböck. Investigating the capability of uav imagery for ai-assisted mapping of refugee camps in east africa. *Proceedings of the Academic Track at State of the Map 2022*, pages 45–48, 2022. 2

[3] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018. 5

[4] Getachew Workineh Gella, Lorenz Wendt, Stefan Lang, Dirk Tiede, Barbara Hofer, Yunya Gao, and Andreas Braun. Mapping of dwellings in idp/refugee settlements from very high-resolution satellite imagery using a mask region-based convolutional neural network. *Remote Sensing*, 14(3):689, 2022. 2

[5] GFDRR Labs. Open Cities AI Challenge Dataset, 2020. 2

[6] Omid Ghorbanzadeh, Alessandro Crivellari, Dirk Tiede, Pedram Ghamisi, and Stefan Lang. Mapping dwellings in idp/refugee settlements using deep learning. *Remote Sensing*, 14(24):6382, 2022. 2

[7] Amrita Gupta, Anthony Ortiz, Simone Fobi, Duncan Kebut, Seema Iyer, Rahul Dodhia, and Juan M. Lavista Ferres. KakumaAerial Dataset (1.0.0) [Data set]. https://doi.org/10.5281/zenodo.14607339, Jan 2025. 2, 4

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 5

[9] Benjamin Herfort, Sven Lautenbach, João Porto de Albuquerque, Jennings Anderson, and Alexander Zipf. The evolution of humanitarian mapping within the openstreetmap community. *Scientific reports*, 11(1):3037, 2021. 1, 2

[10] Amanda Hoffman-Hall, Tatiana V Loboda, Joanne V Hall, Mark L Carroll, and Dong Chen. Mapping remote rural settlements at 30 m spatial resolution using geospatial data-fusion. *Remote Sensing of Environment*, 233:111386, 2019. 1

[11] Humanitarian OpenStreetMap Team. The Audacious Project. https://www.hotosm.org/audacious, 2020. 1

[12] Shunping Ji, Shiqing Wei, and Meng Lu. Fully convolutional networks for multisource building extraction from an open

aerial and satellite imagery data set. *IEEE Transactions on geoscience and remote sensing*, 57(1):574–586, 2018. 2

[13] Emmanuel Maggiori, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez. Can semantic labeling methods generalize to any city? the inria aerial image labeling benchmark. In *2017 IEEE International geoscience and remote sensing symposium (IGARSS)*, pages 3226–3229. IEEE, 2017. 2

[14] Md Nasir, Tina Sederholm, Anshu Sharma, Sundeep Reddy Mallu, Sumedh Ranjan Ghatage, Rahul Dodhia, and Juan Lavista Ferres. Dwelling type classification for disaster risk assessment using satellite imagery. *arXiv preprint arXiv:2211.11636*, 2022. 2

[15] John A Quinn, Marguerite M Nyhan, Celia Navarro, Davide Coluccia, Lars Bromley, and Miguel Luengo-Oroz. Humanitarian applications of machine learning with remote-sensing data: review and case study in refugee settlement mapping. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128):20170363, 2018. 2

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 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*, pages 234–241. Springer, 2015. 5

[17] Wojciech Sirko, Sergii Kashubin, Marvin Ritter, Abigail Annkah, Yasser Salah Eddine Bouchareb, Yann Dauphin, Daniel Keysers, Maxim Neumann, Moustapha Cisse, and John Quinn. Continental-scale building detection from high resolution satellite imagery. *arXiv preprint arXiv:2107.12283*, 2021. 2

[18] Roman Solovyev. Roof material classification from aerial imagery. *Optical Memory and Neural Networks*, 29(3):198–208, 2020. 2, 5

[19] Adam J Stewart, Caleb Robinson, Isaac A Corley, Anthony Ortiz, Juan M Lavista Ferres, and Arindam Banerjee. Torchgeo: deep learning with geospatial data. In *Proceedings of the 30th international conference on advances in geographic information systems*, pages 1–12, 2022. 5

[20] Mingxing Tan and Quoc Le. Efficientnet: Rethinking model scaling for convolutional neural networks. In *International conference on machine learning*, pages 6105–6114. PMLR, 2019. 5

[21] Bing Maps Team. Bing maps global building footprints released, Jun 2023. 2

[22] "Humanitarian OpenStreetMap Team". "participatory mapping toolkit: A guide for refugee contexts". https://www.hotosm.org/downloads/Toolkit-for-Participatory-Mapping.pdf, 2018. [Online; accessed 29-Jun-2023]. 3

[23] United Nations High Commissioner for Refugees. Global trends report 2023. https://www.unhcr.org/global-trends-report-2023, June 2024. [Online; accessed 01-July-2024]. 1

[24] Jamon Van Den Hoek and Hannah K Friedrich. Satellite-based human settlement datasets inadequately detect refugee settlements: a critical assessment at thirty refugee settlements in uganda. *Remote Sensing*, 13(18):3574, 2021. 1

[25] Adam Van Etten, Dave Lindenbaum, and Todd M Bacastow. Spacenet: A remote sensing dataset and challenge series. *arXiv preprint arXiv:1807.01232*, 2018. 2

[26] Shiyu Wang and Xi Li. Use of nighttime light images in evaluating refugee settlement. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024. 2

[27] Matthias Weigand, Simon Worbis, Marta Sapena, and Hannes Taubenböck. A structural catalogue of the settlement morphology in refugee and idp camps. *International Journal of Geographical Information Science*, 37(6):1338–1364, 2023. 2

[28] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017. 5