# An Open Machine Learning Challenge to Map Urban Development and Resilience in Diverse African Cities from Aerial Imagery

Miller, E. (*DrivenData, Denver, CO, United States*);
Luo, D. (*Global Facility for Disaster Reduction and Recovery, World Bank, Washington, DC, United States*);
Deparday, V. (*Global Facility for Disaster Reduction and Recovery, World Bank, Washington, DC, United States*);
Giovando, C. (*Global Facility for Disaster Reduction and Recovery, World Bank, Washington, DC, United States*);
Gevaert, C. (*Global Facility for Disaster Reduction and Recovery, World Bank, Washington, DC, United States*);
Emanuele, R. (*Azavea, Philadelphia, PA, United States*);
Kassel, S. (*Azavea, Philadelphia, PA, United States*);
Needham, E. (*Azavea, Philadelphia, PA, United States*);
Lipstein, G. (*DrivenData, Denver, CO, United States*);
Bull, P. (*DrivenData, Denver, CO, United States*);
Slavitt, I. (*DrivenData, Denver, CO, United States*)

## Abstract

Machine learning (ML) for remote sensing at city scale has focused mostly to date on models and methods developed via labeled data of satellite imagery in cities of higher-income countries. Lack of data diversity limits their applicability in underrepresented settings like much of the developing world. We present a novel open labeled imagery dataset and ML challenge reflecting the diversity of fast-growing, densely populated cities across Africa. These locations are imaged by unmanned aerial vehicles (UAV) at very high resolution with disaster exposure and resilience-relevant labels derived from OpenStreetMap (OSM) data created by local participatory mapping through the World Bank's Open Cities Africa (OCA) and related programs.

The dataset's imagery covers >1,000 sq. km of diverse African cities like Monrovia, Liberia and Dar es Salaam, Tanzania at resolutions of 2-42 cm ground sample distance. Labels include >100,000 building footprints and >2,000 km of drainage mapped and classified. Challenge tasks tentatively include instance segmentation of building footprints and chip classification of drainage features. Monetary and non-monetary prizes for challenge contestants incentivize open-source ML performance on quantitative metrics like F1-score and diversity of participation across experience levels and geographies.

We report on unique complexities, approach, and learnings from open data creation by thousands of community mapping contributors, preparation of a challenge-ready dataset from public imagery and OSM data of inconsistent quality and completeness, challenge participation (est. thousands of data scientists globally) and performance results (interim if still ongoing) and post-challenge integration of top solutions into further mapping efforts.

This will be one of the largest open ML datasets of labeled UAV imagery on cities in the developing world. The challenge will accelerate and diversify participation in the development of better performing open-source ML to map exposure (buildings) and resilience (flood drainage) features from aerial imagery. Thoughtfully integrating these ML advances with participatory mapping efforts like OCA will improve our ability to gain timely, relevant insights on urban development and resilience in these heretofore underrepresented settings.

**Publication:**

American Geophysical Union, Fall Meeting 2019, abstract #IN52B-07

**Pub Date:**

December 2019

**Bibcode:**

2019AGUFMIN52B..07M

**Keywords:**

- 1640 Remote sensing;
- GLOBAL CHANGE;
- 1942 Machine learning;
- INFORMATICS;
- 4329 Sustainable development;
- NATURAL HAZARDS;
- 6620 Science policy;
- POLICY SCIENCES & PUBLIC ISSUES