MAPPING AND ANALYSING URBAN EXPANSION IN THE NAIROBI METROPOLITAN REGION (2017–2024) USING SATELLITE EMBEDDINGS AND K-MEANS CLUSTERING

Emmanuel Otieno 1,2 (Presenter), Piero Boccardo 2

(1) Department of Civil, Building and Environmental Engineering (DICEA), Sapienza Università di Roma, Italy

(2) SDG11lab, Interuniversity Department of Regional and Urban Studies and Planning (DIST), Politecnico di Torino , Italy

GeoAI Torino 2025

Abstract

Rapid urban expansion in Nairobi poses major challenges for sustainability and planning. This study maps and quantifies built-up areas across 26 subcounties of the Nairobi Metropolitan Region (2017–2024) using Google Earth Engine's AlphaEarth Satellite Embedding dataset and unsupervised K-means clustering with 7 clusters selected via the Elbow Method. Urban areas were identified automatically each year by maximising F1 score against reference data. Validation against ESA WorldCover 2021 yielded 80% user's accuracy and 69% F1, while Dynamic World gave 75% F1. Built-up area increased from 253.64 km² in 2017 to 391.29 km² in 2024 (54% cumulative growth, 6% AAGR). Fastest-growing subcounties: Kabete (226%), Matungulu (195%), Githunguri (194%), driven by peri-urban sprawl. Bootstrapping provided a 95% confidence interval of approximately 68–71% for F1. A supervised Random Forest baseline reached 82% F1 but required labelled data. The unsupervised embedding-based approach proves efficient and scalable for data-scarce regions, delivering actionable subcounty-level insights for urban policy and SDG11 monitoring.

- **Urban growth** in Nairobi strains infrastructure and environment
- **Study analyzes** 2017-2024 using satellite embeddings and K-means
- **Key objectives:** Map extents, compute metrics, validate, compare methods
- **Focus:** Slums & Peri-urban sprawl in 26 subcounties in Nairobi county and its neighbouring region.

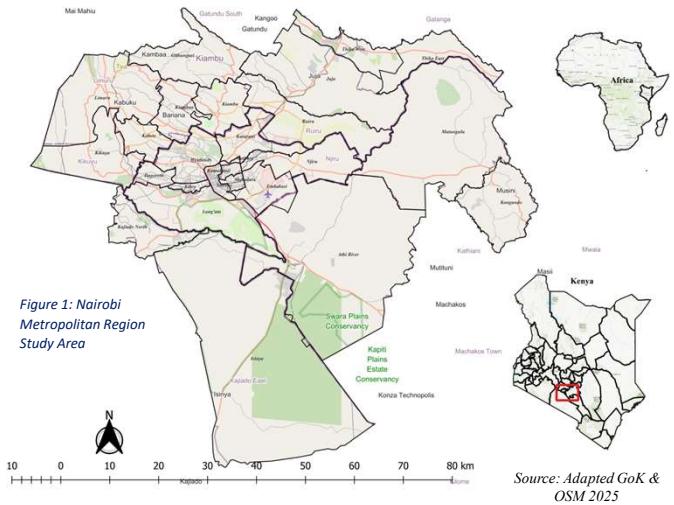
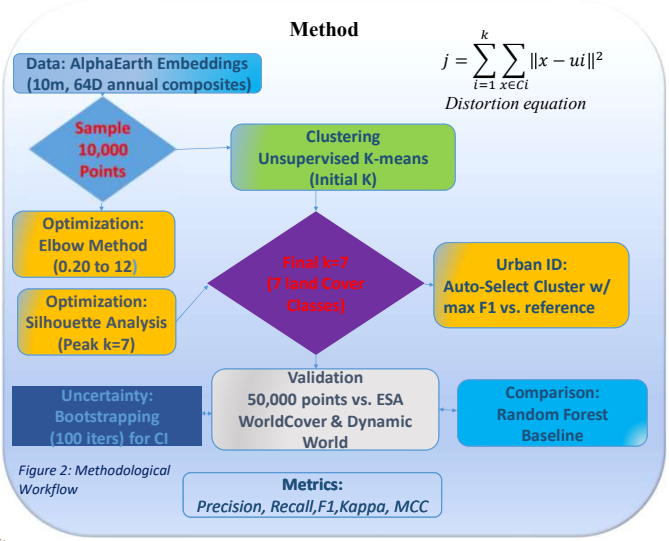


Figure 1: Nairobi Metropolitan Region Study Area

Source: Adapted GoK & OSM 2025

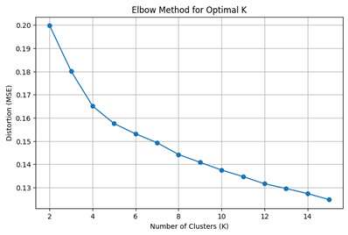
Method



$$j = \sum_{i=1}^{k} \sum_{x \in Ci} \|x - ui\|^2$$

Distortion equation

Figure 2: Methodological Workflow



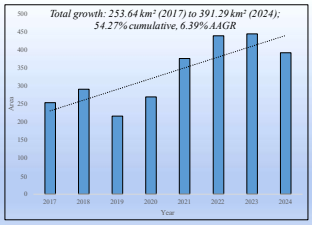
Validation: User's Accuracy 80.21%, F1 69.68% (ESA), 75% (Dynamic World)

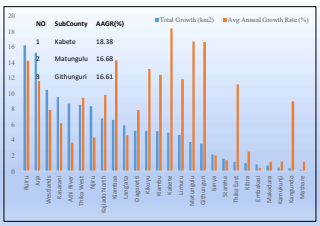
Metric	Value
Precision (User's Accuracy)	80.21%
Recall (Producer's Accuracy)	61.59%
F1 Score	69.68%
Overall Accuracy	94.84%
Cohen's Kappa	0.67
Matthews Correlation Coefficient	0.68

ARI stability 0.57

The 2024 built-up estimate of the study aligns with World Bank projections and reports from UN-Habitat/OECD on rapid informal peri-urban expansion

Results





Discussion

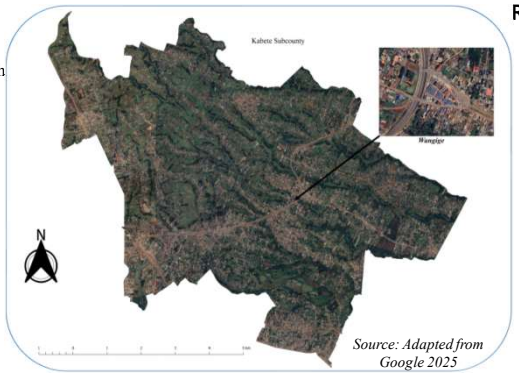
- Growth driven by peri-urban sprawl (low costs, accessibility)
- Unsupervised method efficient vs. supervised baselines
- Consistent with independent sources (Demographia, World Bank) on trends, though conservative estimates.
- Implications: Prioritize infrastructure in hotspots; Advances SDG11 monitoring

Challenges

- Data scarcity in African contexts limits high-resolution embeddings
- Peri-urban transitions cause classification uncertainty (higher in bootstrapping)
- Methodological fragility from manual tweaks (addressed via automation, but scalability issues remain)
- Discrepancies with broader sources due to built-up definitions

Way Forward

- Use of supervised classification to create a baseline
- Enhance with transfer learning for finer sub-urban classes
- Collaborate on ground-truth validation in peri-urban areas
- Integrate with other urban growth metrics
- Scale to other African cities for regional SDG11 tracking



Source: Adapted from Google 2025

Figure 3: Highest Built-Up Growth Rate Subcounty

References

Brown et al. (2025)
Cobbinah et al. (2015)
Demographia (2023)
KNBS (2019)
OECD/UN ECA/AfDB (2022)
Rousseeuw (1987)
Sang et al. (2022)
UN-Habitat (2024)
World Bank (2016)
Zanaga et al. (2022)

GitHub