
Towards Modeling Road Access Deprivation in Sub-Saharan Africa Based on a New Accessibility Metric and Road Quality

Sebastian Hafner¹ Qunshan Zhao^{1*} Bunmi Alugbin² Kehinde Baruwa²
Caleb Cheruiyot³ Sabitu Sa'adu Da'u⁴ Xingyi Du¹ Peter Elias²
Helen Elsey⁵ Ryan Engstrom⁶ Serkan Girgin⁷ Diego F. P. Grajales¹
Esther Judith³ Caroline Kabaria³ Monika Kuffer⁷ Oluwatoyin Odulana²
Francis C. Onyambu³ Adenike Shonowo¹ Dana R. Thomson⁸ Mingyu Zhu¹
João Porto de Albuquerque¹

¹University of Glasgow ²University of Lagos ³APHRC ⁴Bayero University

⁵University of York ⁶George Washington University ⁷University of Twente

⁸Columbia University

*Corresponding author: qunshan.zhao@glasgow.ac.uk

Abstract

Access to motorable roads is a critical dimension of urban infrastructure, particularly in rapidly urbanizing regions such as Sub-Saharan Africa. Yet, many urban communities, especially those in informal settlements, remain disconnected from road networks. This study presents a road access deprivation model that combines a new accessibility metric, capturing how well buildings are connected to the road network, with road surface type data as a proxy for road quality. These two components together enable the classification of urban areas into low, medium, or high deprivation levels. The model was applied to Nairobi (Kenya), Lagos (Nigeria), and Kano (Nigeria) using open geospatial datasets. Across all three cities, the majority of built-up areas fall into the low and medium road access deprivation levels, while highly deprived areas are comparatively limited. However, the share of highly deprived areas varies substantially, ranging from only 11.8 % in Nairobi to 27.7 % in Kano. Model evaluation against community-sourced validation data indicates good performance for identifying low deprivation areas ($F1 > 0.74$), moderate accuracy for medium deprivation in Nairobi and Lagos ($F1 > 0.52$, lower in Kano), and more variable results for high deprivation ($F1$ ranging from 0.26 in Kano to 0.69 in Nairobi). Furthermore, analysis of grid cells with multiple validations showed strong agreement among community members, with disagreements occurring mainly between adjacent deprivation levels. Finally, we discussed two types of sources for disagreement with community validations: (1) misalignment between the conceptual model and community perceptions, and (2) the operationalization of the conceptual model. In summary, our road access deprivation modeling approach demonstrates promise as a scalable, interpretable tool for identifying disconnected areas and informing urban planning in data-scarce contexts.

1 Introduction

Urban access to roads is a critical dimension of development and a key enabler for service delivery, economic participation, and spatial integration (UN-Habitat, 2014). In rapidly urbanizing regions such as Sub-Saharan Africa, road networks serve as a backbone for mobility, emergency response,

and infrastructure provision. However, large segments of urban populations, especially those residing in informal settlements, remain inadequately connected to motorable roads, reinforcing spatial inequality and limiting development opportunities (Bettencourt and Marchio, 2025). As a consequence, for example, fires are very recurrent without the option of firetrucks accessing densely built-up settlements.

A growing body of literature has emphasized the importance of well-planned street networks for urban poverty reduction. Angel et al. (2011) and UN-Habitat (2014) argue that coherent road layouts support inclusive growth by improving accessibility and service reach. More recently, the Million Neighborhoods project developed a topological analysis to quantify the relationship between roads and buildings (Brelsford et al., 2018, 2019). Brelsford et al. (2018) introduced block complexity, a measurement of the connectedness of a city block. A higher complexity value is associated with increased difficulty in reaching places within the block from the street network. They demonstrated that this topological analysis is effective in identifying slums based on the lack of spatial access and related services, using a township of Cape Town, South Africa, as an example. In Brelsford et al. (2019), they provide further evidence of the effectiveness of their topological analysis in revealing the absence of access to places within a slum. Beyond that, they proposed a method to reduce block complexity via topological optimization, generating an access network that makes each structure in a settlement accessible.

Enabled by the availability of open building footprint and street network data, topological analysis of street access can now be performed at large scales (Soman et al., 2020; Bettencourt and Marchio, 2025; Thomson and Blei, 2025). Soman et al. (2020) applied the concept of block complexity to the entire Global South using OpenStreetMap (OSM) data. More recently, Bettencourt and Marchio (2025) combined building footprints data from Ecopia Landbase Africa and OSM street network data to compute block complexity in Sub-Saharan Africa. They demonstrated that block complexity expresses not only spatial access deprivation but also entails many dimensions of low human development. For example, they showed that higher block complexity is systematically associated with lower female literacy, lower access to a water source on premises, or lower access to an improved sanitation facility. In contrast, Thomson and Blei (2025) developed the city segment layer, which identifies sub-city areas characterized by high building density and low road connectivity, serving as a proxy for neighborhood socioeconomic status. Unlike block complexity, the city segment layer measures road connectivity in terms of meters of road per building.

Beyond street access, studies have demonstrated that road surface type serves as a useful proxy for identifying informal or underserved urban areas (Yeboah et al., 2021; Kohli et al., 2012). The importance of road surface type was further highlighted in a recent study that found a positive correlation between average road pavedness and human development index values across different countries (Randhawa et al., 2025). In particular, significant gaps in paved road coverage were noted in specific areas of Africa (Randhawa et al., 2025). Road quality has also been linked to issues such as flood resilience and health access in deprived urban and rural settings (Burelle, 2021; Li et al., 2022).

Despite significant progress in quantifying street access, we identified two key limitations in understanding and modeling road access deprivation. First, while there are well-established metrics measuring street access to buildings such as block complexity, the quality of road infrastructure is not considered by these metrics. Second, there remains a need to ground and validate such deprivation concepts through engagement with local communities, ensuring their relevance and interpretability in local contexts. Achieving this requires concepts that can be easily communicated and intuitively understood by non-experts.

To enable community-based validation of road access deprivation, this study proposes a simplified, distance-agnostic accessibility metric that captures the number of buildings obstructing the path between a structure and its nearest motorable road. We further combine this accessibility measure with road surface type to define levels of road access deprivation. In doing so, we aim to more accurately identify areas where communities are functionally disconnected from adequate road infrastructure.

The main contributions of this study are as follows:

- We co-design a conceptual framework for road access deprivation tailored to urban contexts in Sub-Saharan Africa;
- We introduce a simple obstruction-based road accessibility metric;

- We generate gridded road access deprivation outputs for three cities: Nairobi (Kenya), Lagos (Nigeria), and Kano (Nigeria);
- We validate the model outputs using community-sourced data, providing insights into the spatial distribution and classification performance of the model.

Through this work, we aim to support efforts to map and mitigate infrastructure-related deprivation and to contribute a scalable, interpretable tool for urban planners, policymakers, and community stakeholders.

2 Study Areas and Grid

2.1 Study Areas

The study areas of this research are Nairobi, Kenya, Lagos, Nigeria, and Kano, Nigeria (Figure 1). These cities represent different types of urbanization and geographies. Nairobi represents an inland city with very densely built-up settlements. Its administrative area spans 695 km² with an estimated population of 4.4 million (Kenya National Bureau of Statistics, 2019). Approximately two-thirds of the city's urban population resides in deprived areas as of 2014 (Wamukoya et al., 2020). Nairobi is also home to Kibera, often regarded as the largest informal settlement in Africa. Lagos, on the other hand, represents a coastal city with lagoon settlements and sprawling unplanned areas. It is the largest urban agglomeration in Nigeria. Different authorities estimated the population at 12.4 million and over 21 million in 2016 (UN-Habitat, 2020). Kano represents a secondary city located in a conflict zone. It is also the largest city in northern Nigeria with an area of 499 km² in 2015 (Yahaya and Suleiman, 2017). Kano state is home to an estimated population of 15 million (Koko et al., 2023).

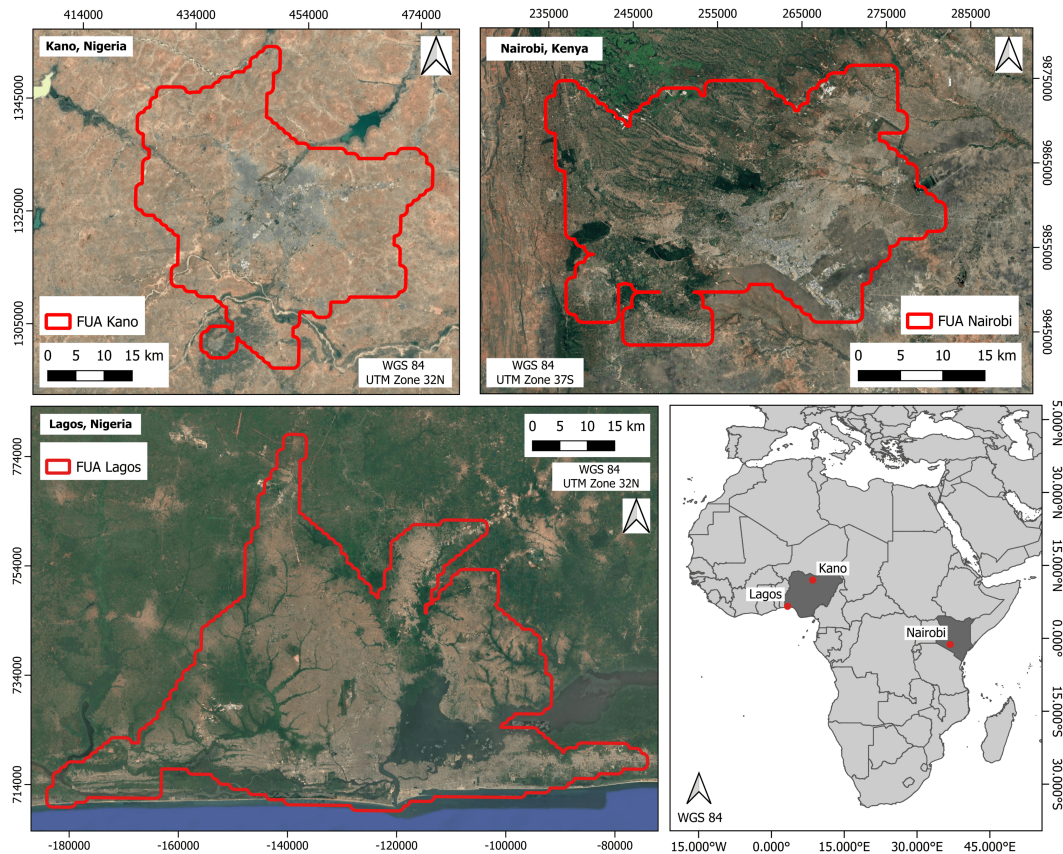


Figure 1: Overview of the study areas.

To ensure that the study areas reflect the actual extent of cities and settlements, we base our analysis on functional rather than administrative extents. Administrative boundaries are often limiting since

they are typically shaped by historical, political, or administrative decisions and may not necessarily reflect the actual patterns of human settlement or economic activity, especially in fast-growing African cities. For example, the administrative boundary of Nairobi has remained unchanged since 1963 and is now outdated given the substantial urban expansion of Nairobi (Abascal et al., 2022). Likewise, Lagos has expanded significantly beyond its city administrative boundary and is spreading to its neighboring administrative area (Oyalowo, 2022). Functional urban areas (FUAs) offer a more realistic alternative, as they are defined by actual urban sprawl and human activities, encompassing the core city and economically or socially integrated surrounding regions (Dijkstra et al., 2019). We apply a uniform buffer (1 km) to FUAs, ensuring peripheral urban zones are included.

2.2 Grid

Road access deprivation is modeled at a $100\text{ m} \times 100\text{ m}$ grid. The grid system is based on the Mollweide projection system, an equal-area projection displaying the globe as an ellipse with an axes proportion of 2:1 (Lapaine, 2011). Due to its suitability for global applications requiring accurate area representations, the grid system is also used by the global human settlement layer (GHSL) data suite (Pesaresi et al., 2024). Consequently, this grid also facilitates an analysis of gridded population statistics, sourced from the GHSL data suite, within morphological informal areas.

3 Methodology

3.1 Conceptual Model

The proposed model categorizes road access deprivation into three levels: low, medium, and high, based on road accessibility and road surface type. This level of granularity was chosen since stakeholder discussions indicated that binary distinctions between deprived and non-deprived areas fail to reflect the variation of deprivation. On the other hand, a higher level of granularity (such as a degree of deprivation Abascal et al., 2022), can be overwhelming for non-technical stakeholders.

In the following, we provide definitions for each road access deprivation level:

Definition 1 (Low road access deprivation). The road network is paved and in good condition. The majority of buildings in these areas are directly connected to motorable roads, meaning no buildings block their access to the nearest road. Areas with no buildings are also categorized as low.

Definition 2 (Medium road access deprivation). Roads are unpaved but motorable. Most buildings are directly connected to roads, meaning no buildings block their access to the road network.

Definition 3 (High road access deprivation). Motorable roads are sparse. Most buildings have only indirect access to the road network via paths, meaning more than one building is generally located between a building and its nearest road.

Figure 2 gives visual examples of the three road access deprivation levels using Google satellite imagery and illustrations.

3.2 Model Implementation

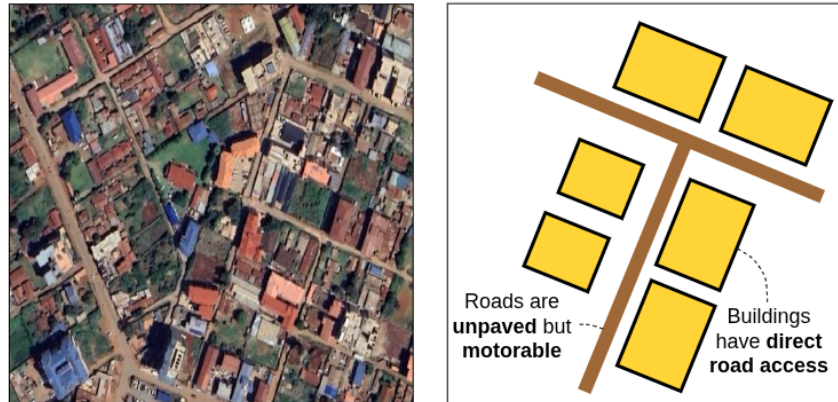
The road access deprivation modeling pipeline can be summarized in four main steps (Figure 3). Step 1 prepares the building footprints, road network, and road surface type data. Then, a new road accessibility is computed at the building level based on the number of buildings intersected between a building centroid and its nearest road. Furthermore, information about the surface type of the nearest road is added to each building. In Step 3, these building-level metrics are aggregated to the $100\text{ m} \times 100\text{ m}$ grid level using different mathematical operators. Finally, Step 4 applies a simple logic to classify cells into three levels of road access deprivation, low, medium, and high, based on the road accessibility metric and the predominant road surface type. The following subsections describe each step in detail. Code for the model is available from the official IDEAMAPS Data Ecosystem model repository on GitHub¹. Additionally, we created a Google Earth Engine App to visualize the input data, the metrics at the building and grid level, and the final road access deprivation levels².

¹<https://github.com/urbanbigdatacentre/ideamaps-models>

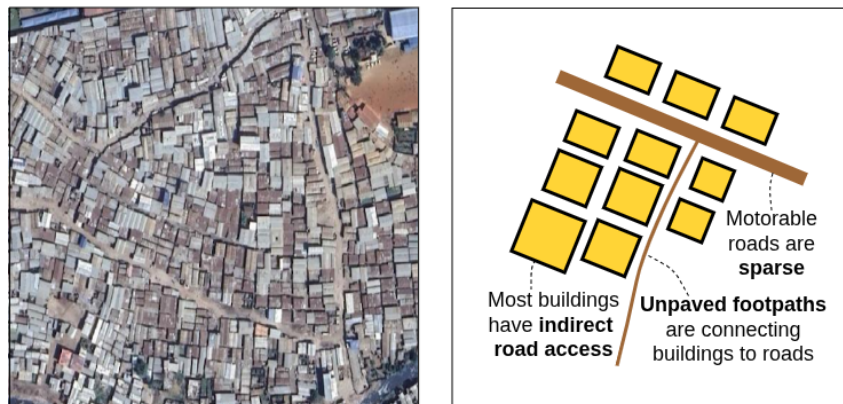
²<https://ee-hafnersailing.projects.earthengine.app/view/ideamaps-road-access-deprivation>



(a) Low road access deprivation



(b) Medium road access deprivation



(c) High road access deprivation

Figure 2: Conceptual road access deprivation model showing the definitions of the three deprivation levels: **low**, **medium**, and **high**. Satellite imagery: Google, Maxar Technologies.

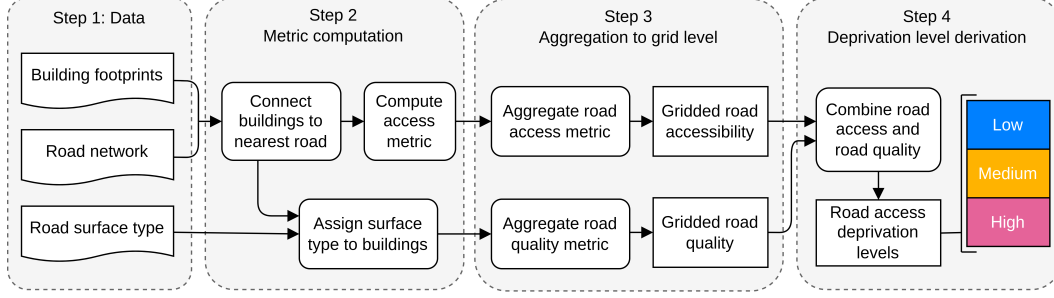


Figure 3: Overview of the road access deprivation modeling pipeline.

3.2.1 Data and Preprocessing

We consider three datasets to model road access deprivation: road network, road surface type, and building footprints.

Road network: The road networks were retrieved from OpenStreetMap (OSM) (Haklay and Weber, 2008). All accessible pathways and roads available in OSM are included in the road network. It should also be noted that while OSM is a crowdsourced database, its completeness for roads is generally good at a global level (above 80 %), which is also the case for several developing countries (Barrington-Leigh and Millard-Ball, 2017). Since we are modeling accessibility to motorable roads (see conceptual model in Section 3.1), only road classes that are considered accessible by a motorized vehicle were selected from the OSM data (see Section A.1 in Appendix).

Road surface type: Road surface type data (paved or unpaved) was obtained for Kenya from Zhou et al. (2024). These data were generated using a deep learning approach that determines road surface type for point locations along the OSM road network based on Google satellite imagery. The classified points were combined to determine the road surface type for a road segment. For Lagos and Kano, on the other hand, the road surface type data was obtained from Zhou et al. (2025), who produced these data for Nigeria and Cameroon using a similar approach to Zhou et al. (2024). However, instead of Google satellite imagery, the model was trained on 16 variables, including road network variables, socio-economic variables, and land cover variables. The accuracy assessments in both studies indicate that the resulting road surface type data are accurate (i.e., F1 scores > 0.88).

Buildings: The Open Buildings (V3) dataset provided by Google is used as building footprints data (Sirko et al., 2021). This dataset is used as it has greater coverage and completeness within urban areas compared to other openly accessible datasets such as Ecopia, OSM and Microsoft (Chamberlain et al., 2024). Although Overture³, an open-data initiative that integrates several building footprint sources, was also considered for its improved completeness, issues of consistency limit its suitability for cross-city comparisons. In contrast, Open Buildings focuses on classifying buildings across Africa and much of the Global South, making it particularly appropriate for analyzing both current and emerging urban areas (Sirko et al., 2021).

3.2.2 Metrics Computation

This section first introduces our new distance-agnostic accessibility metric. Specifically, we measure road accessibility based on the number of buildings intersecting lines connecting buildings to their nearest motorable roads.

We argue that this accessibility metric is preferable over distance-based metrics, as buildings in informal settlements are often not directly connected to a motorable road, yet are still near one in terms of geographic distance due to the high built-up density. Conversely, buildings in formal areas typically have direct access to motorable roads, but distances to roads can be relatively far due to larger plot sizes. Furthermore, this metric does not require footpath data, which is often incomplete or completely unavailable for informal settlements.

³[overturemaps.org](https://www.overturemaps.org)

To compute the accessibility metric, we determine the nearest point on the nearest motorable road for all buildings. We connect this point to the middle point (i.e., centroid) of the building (Figure 4).

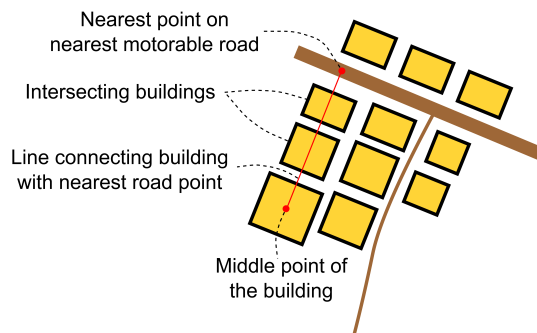


Figure 4: Illustration of the proposed distance-agnostic accessibility metric.

Based on the line connecting a building to its nearest motorable road point, we determine the number of buildings intersecting this line. The left and right illustrations of Figure 5 show the resulting accessibility values for the buildings in an informal area and a formal area, respectively. Values are expected to be higher in informal areas because many buildings are only connected to motorable roads via footpaths.

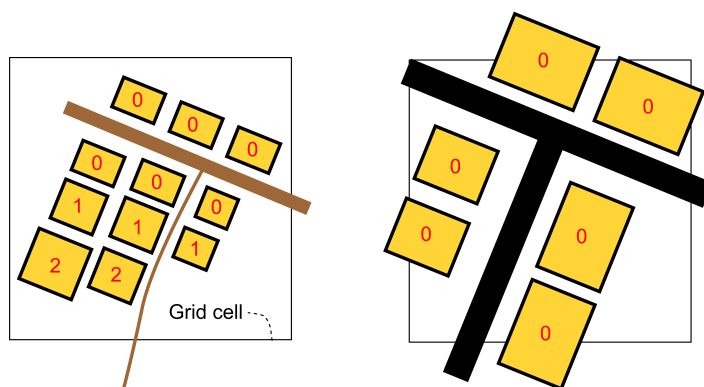


Figure 5: Examples of values for the proposed accessibility metric in (left) an informal area and (right) a formal area.

In addition to the computation of accessibility values at the building level, we assign a road surface type value (paved or unpaved) to each building corresponding to the surface type of the building's nearest road. We utilize surface type as a proxy for road quality and capture whether a buildings have access to high-quality roads (i.e., paved motorable roads) or low-quality roads (i.e., unpaved motorable roads).

3.2.3 Grid Aggregation

Since we model deprivation at a $100\text{ m} \times 100\text{ m}$ grid level, the building-level metrics are aggregated. For the accessibility metric, we average the values of the buildings located within a grid cell (see Figure 5 for example values that are averaged). The resulting grid-level metric represents the average number of buildings intersecting the line connecting a building to its nearest road.

For the aggregation of the road quality metric, we compute the mode of the road surface types at the building level. The aggregated metric represents the predominant road surface type that buildings within a grid cell have access to based on their nearest road.

3.2.4 Deprivation Level Assignment

The two metrics are combined to derive three road access deprivation levels - low, medium, and high. If the average number of intersecting buildings for a grid cell exceeds a threshold of 1, the grid cell is associated with high road access deprivation. Although direct access would require setting the threshold to 0, in practice, intersecting buildings frequently occur in formal areas due to data quality limitations, such as inaccurate building footprints data, where a single building is represented as multiple building polygons, or missing road segments. Otherwise, if the average number of intersecting buildings is below the threshold of 1, the cell is associated with low or medium road access deprivation. The differentiation between low and medium is based on the road quality metric. Grid cells with predominant paved roads are associated with low deprivation, whereas cells where unpaved roads are predominant are associated with a medium level.

3.3 Model evaluation

The model is evaluated based on community-sourced reference data, which is available for a limited number of grid cells. These validation data were collected using the IDEAMAPS Data Ecosystem Platform⁴. Participatory action research sessions with two local partner communities of the IDEAMAPS Data Ecosystem project in each city were held, where community members validated grid cells by assigning a low, medium, or high deprivation level using the platform interface. Additionally, city-wide validation campaigns were conducted to collect validation data at the city level.

In detail, validation data were collected in Nairobi with community members from Korogocho and Viwandani on 4 and 5 June 2025, respectively. The city-wide validation campaign in Nairobi was conducted on 30 July 2025 (validations on 1 August 2025 were also included). For Lagos, the city-wide validation campaign, including members from the local partner communities Okerube and Ajegunle-Ikorodu, was conducted on 11 and 12 June 2025. The inclusion of data up to 16 June 2025 accounts for substantial community validation that extended beyond the scheduled sessions. Validation data in Kano were collected for the partner communities on 22 May 2025, and the city-wide validation campaign was conducted from 27 June to 3 July 2025.

Table 1 lists the number of people and validations per study area for road access deprivation. In total, 131 people validated 15,361 cells across the three cities. However, different people validated some cells multiple times. For these multi-validation cells, we checked for consensus among the validations. Specifically, a consensus is reached if one level receives a majority of validation votes (i.e., at least one more vote than any other level). Grid cells without a clear majority (i.e., two or more levels receive the same highest number of votes) are excluded due to a lack of consensus. Cells with only one validation are considered to have consensus by default. This was done to ensure the reliability and consistency of the ground truth, as conflicting responses could indicate uncertainty or disagreement about the level of deprivation in those locations.

Table 1: Overview of validation data for road access deprivation.

	# People	# Validations	# Validated cells	# Validated cells with consensus			
				Total	Low	Medium	High
Nairobi	59	6,456	5,643	5,479	3,857	971	651
Lagos	52	5,586	4,478	4,363	3,418	555	390
Kano	20	3,319	3,295	3,280	1,536	1,104	640

Based on this reference data, model outputs are evaluated using the metrics overall accuracy (Acc) and F1 score, defined as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}, \quad (1)$$

$$F1 = \frac{TP}{TP + \frac{1}{2} \times (FP + FN)}, \quad (2)$$

⁴<http://www.ideamapsdataecosystem.org/>

where TP, TN, FP, and FN denote the number of true positives, true negatives, false positives, and false negatives pixels, respectively.

4 Results

4.1 Modeling Outputs

Figure 6 presents the modeled road access deprivation for Nairobi (Kenya), Lagos (Nigeria), and Kano (Nigeria), showing low, medium, and high deprivation levels based on the distance-agnostic accessibility metric and predominant road surface type. The maps reveal consistent spatial patterns across the three cities. Low deprivation areas are predominant, and generally concentrated in the urban core, as well as on the outskirts of the cities. The latter are partly due to areas without any built-up area, which is also considered low-deprivation according to the conceptual model (Section 2). Pockets of high deprivation are dispersed throughout the study areas. Medium deprivation zones, typically associated with unpaved but direct access to motorable roads, are found in peripheral neighborhoods outside the urban core.

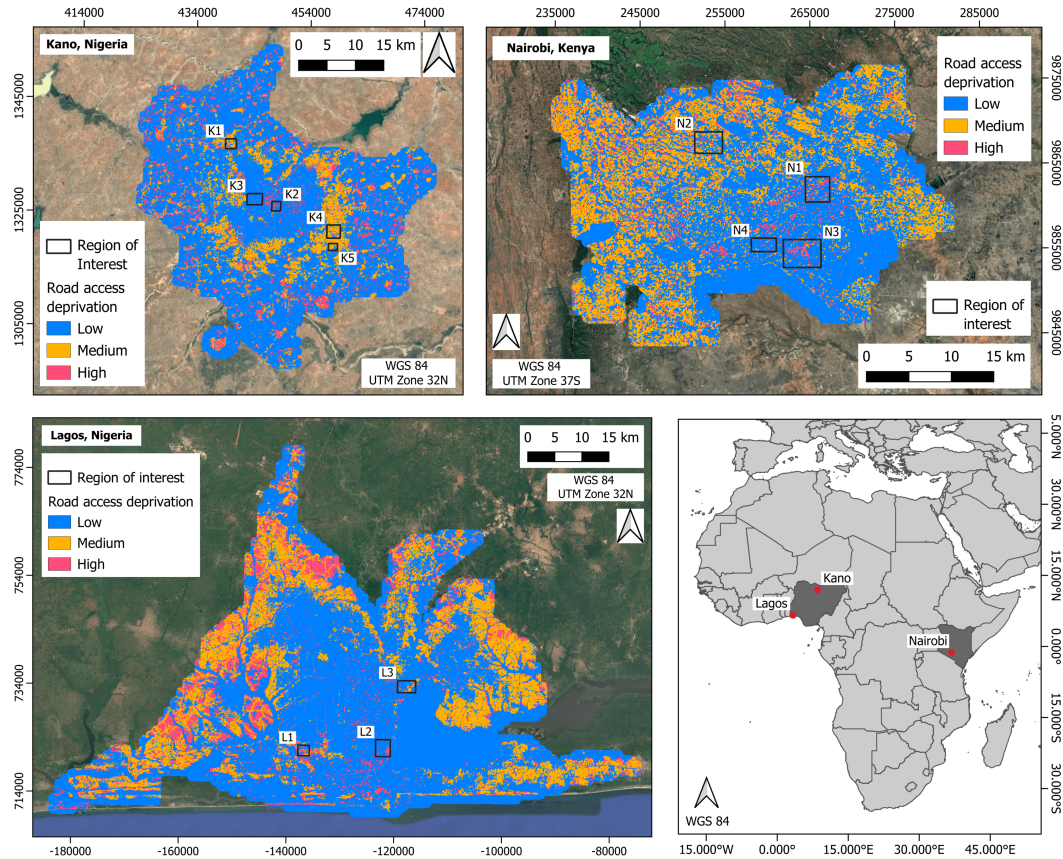


Figure 6: Road access deprivation modeling results for Nairobi, Kenya, Lagos, Nigeria, and Kano, Nigeria.

In addition, Figure 7 shows the road access deprivation level distributions for the three cities. It should be noted that cells with no buildings were excluded to isolate road access statistics for built-up areas. In Nairobi, Lagos, and Kano, the numbers of no-building cells excluded from this analysis were 33,491 (30.3 %), 118,471 (36.4 %), and 99,132 (59.3 %), respectively. In other analyses, these are included in the low-deprivation category following the conceptual model (Section 3.1). Across all three cities, low and medium deprivation areas are the most prevalent, with similar proportions of grid cells falling within either category. High-deprivation areas, on the other hand, are less common. In particular in Nairobi, relatively few grid cells fall within the high category (11.8 %). In contrast,

the percentage of highly deprived grid cells is high in Kano (27.7 %), and only slightly lower than those for low and medium deprivation.

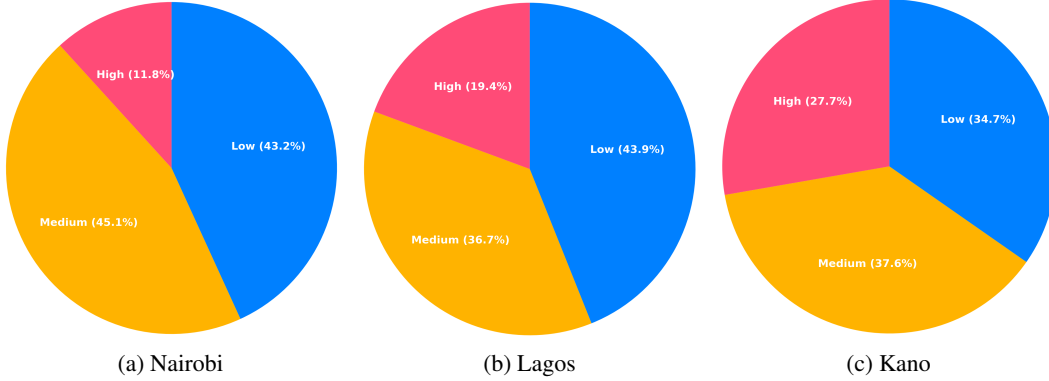


Figure 7: Pie charts showing distribution of modeled road access deprivation levels, excluding cells with no buildings in the low-deprivation category.

4.2 Quantitative Assessment

Table 2 presents the quantitative evaluation of the road access deprivation model against community-sourced validation data. Overall model accuracy is highest in Nairobi (79.2 %) and Lagos (77.9 %), followed by Kano (63.2 %). The deprivation level-specific F1 scores indicate that the models perform best in identifying low deprivation, especially in Nairobi (F1 = 0.873) and Lagos (F1 = 0.871). Performance is poorer in identifying medium deprivation, particularly in Lagos (F1 = 0.378). For high deprivation, we observe the highest performance variance, ranging from poor results in Kano (F1 = 0.258) to good results in Nairobi (F1 = 0.691).

Table 2: Quantitative accuracy assessment of the road access deprivation model based on community-sourced validation data.

	Acc (%)	F1 score		
		Low	Medium	High
Nairobi	79.2	0.873	0.528	0.691
Lagos	77.9	0.871	0.378	0.452
Kano	63.2	0.747	0.618	0.258

Figure 8 uses alluvial plots to visualize the relationship between model outputs and community-sourced validation data. In Nairobi and Lagos, the agreement between modeled and validated road access deprivation levels is strong for low deprivation, confirming the model’s ability to identify areas with good road accessibility. Most confusion in Nairobi exists between the low and medium levels. Similarly, many of the grid cells validated as medium-deprived in Lagos were incorrectly modeled as low-deprived. We observe in Lagos that a large fraction of grid cells modeled as highly deprived were validated as low. Considerable disagreement with the modeled highly deprived grid cells is also apparent in Kano, with similar fractions of them being validated as low, medium, and high. On the other hand, the modeled representation of medium road access deprivation is most accurate in Kano.

The confusion matrices in Figure 9 provide a detailed view of classification performance across the three deprivation levels, complementing the accuracy metrics in Table 2 and the patterns seen in the alluvial plots (Figure 8). In Nairobi, the model achieves balanced performance across low and high deprivation, reflected in high F1 scores for both classes (0.873 and 0.691, respectively). The confusion matrix shows strong agreement for low deprivation, but medium deprivation is often misclassified as low, mirroring the substantial low–medium mixing seen in the alluvial plot. Lagos displays a similar pattern, with good identification of low deprivation but weaker results for medium and high deprivation. Many cells validated as medium are predicted as low, and a notable share of predicted high deprivation is validated as low, indicating overestimation of severe deprivation in

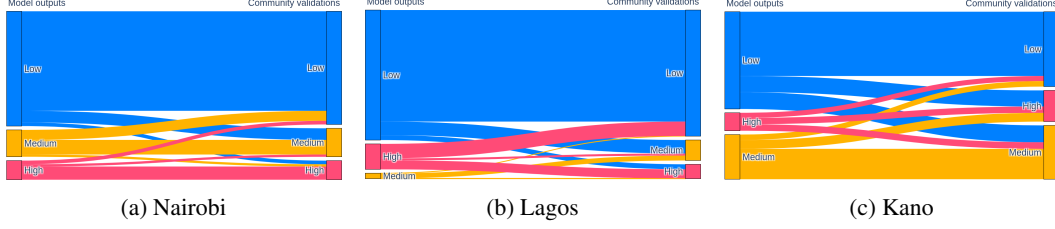


Figure 8: Alluvial plots illustrating the relationship between modeled and community-validated road access deprivation levels. The left side represents modeled levels, and the right side shows the corresponding community-sourced validations. The width of each flow indicates the number of validations with a given pair of levels, highlighting areas of agreement and disagreement across deprivation levels.

some well-connected areas. Kano shows the lowest overall accuracy (63.2 %), with high deprivation performing poorly ($F1 = 0.258$). The confusion matrix confirms widespread disagreement: high deprivation is almost equally redistributed across all three validated classes, while medium deprivation is captured more accurately ($F1 = 0.618$) than in the other cities. These patterns indicate that while the model reliably detects low deprivation in all contexts, the medium class remains the most challenging to model consistently, and high deprivation performance is highly context-dependent.

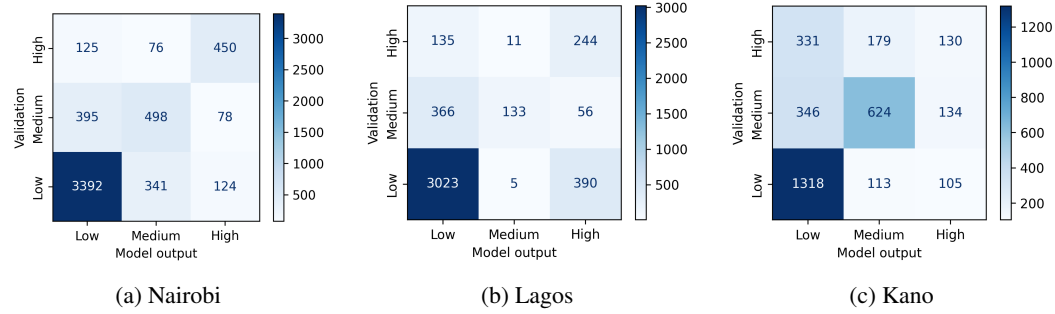


Figure 9: Confusion matrix showing the road access deprivation classification performance across modeled and validated deprivation levels. Rows represent the community-sourced deprivation levels, and columns represent modeled deprivation levels. Diagonal cells indicate correct classifications, while off-diagonal cells represent misclassifications.

4.3 Qualitative Assessment

Figures 10 to 14 show visual comparisons between model results and community-sourced validation data for regions of interests in Nairobi (N1: Figure 10, N2: Figure 11), Lagos (L1: Figure 12, L2: Figure 13), and Kano (K1: Figure 14a, K2: Figure 14b). Grid cells with disagreement between validations and model outputs are hashed in the validation visualization. Additional qualitative comparisons are available in the supplementary material (see Figures 16 to 21 in Section A). Locations of all regions of interest are shown in Figure 6.

N1 and N2, the regions of interest in Nairobi, represent a dense urban area close to the city's core and a more sparse urban area on the periphery of Nairobi, respectively. The model aligns well with the community validation data in identifying low-deprivation areas in both regions, consistent with the high F1 score for low deprivation (0.873). However, road access deprivation is underestimated in dense urban areas (N1), which is also apparent in the additional regions of interest (N3 and N4) listed in the supplementary material. Specifically, large areas validated as medium-deprived were incorrectly modeled as low, indicating that while the connectivity was correctly modeled as good, i.e., buildings are directly connected to motorable roads, the surface type of roads is predominantly unpaved instead of paved. Additionally, a few areas modeled as low were validated as highly deprived. In the rural region of interest (N2), we also observe grid cells validated as medium that were incorrectly modeled

as low. In contrast to the dense urban areas, however, the model is overestimating high road access deprivation in this rural area.

L1 and L2 are both located in the periphery of Lagos. A strong agreement between modeling outputs and validation data is shown for L1, with the main confusion coming from low-deprivation model outputs that were validated as medium-deprived by community members. For L2, we also observe strong agreement for low and high road access deprivation. Disagreements in L2 primarily arise from low-deprivation areas validated as medium and from a small number of high-deprivation grid cells validated as low, the latter likely influenced by their location within clusters of low-deprivation areas.

For Kano, we show a community located in the periphery of the city (K1) and one in the urban core (K2). For K1, there is a moderate agreement between model outputs and validation data. Proximity to the primary road connecting the town to the city is validated as low road access deprivation. Medium and high road access deprivation levels are found with increasing distance to the primary road in the validation visualization. These observations are also consistent with the additional regions of interest showing peripheral neighborhoods in Kano from the supplementary material (K4 and K5). For K2, located in the urban core of Kano, road access deprivation is modeled primarily as low, which is a severe underestimation according to the community-sourced validation data. Specifically, areas with medium and high deprivation were incorrectly modeled as low, indicating modeling errors for road surface type and road connectivity, respectively. To a much lesser degree, these errors are also present in K3 from the supplementary material, which is also located in the urban core of Kano.

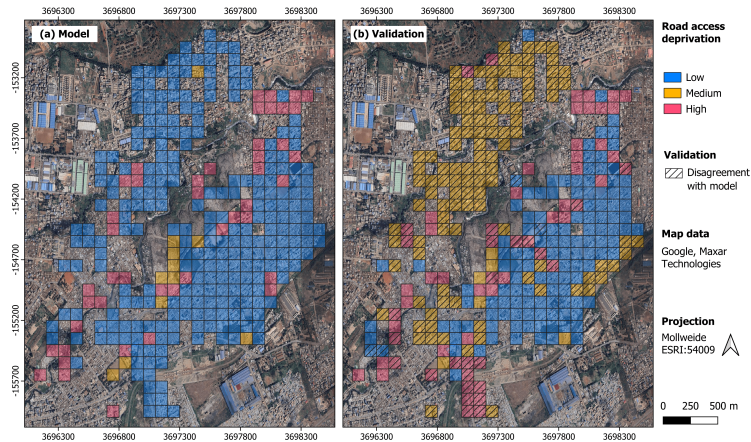


Figure 10: Region of interest N1: Visual comparison between (a) model results and (b) the data from validation with local communities in Nairobi, Kenya.

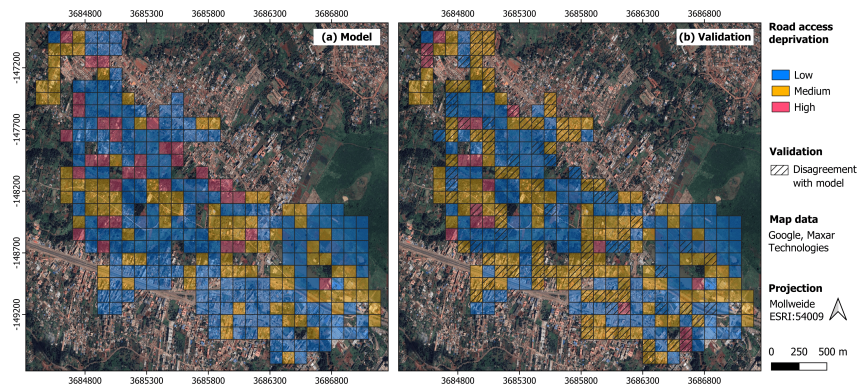


Figure 11: Region of interest N2: Visual comparison between (a) model results and (b) the data from validation with local communities in Nairobi, Kenya.

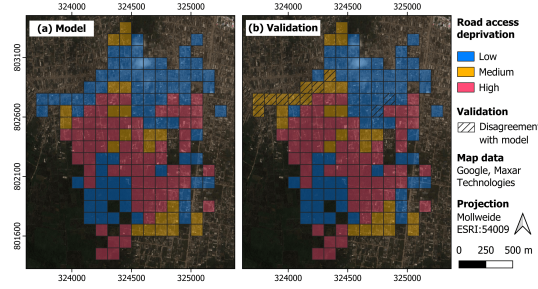


Figure 12: Region of interest L1: Visual comparison between (a) model results and (b) the data from validation with local communities in Lagos, Nigeria.

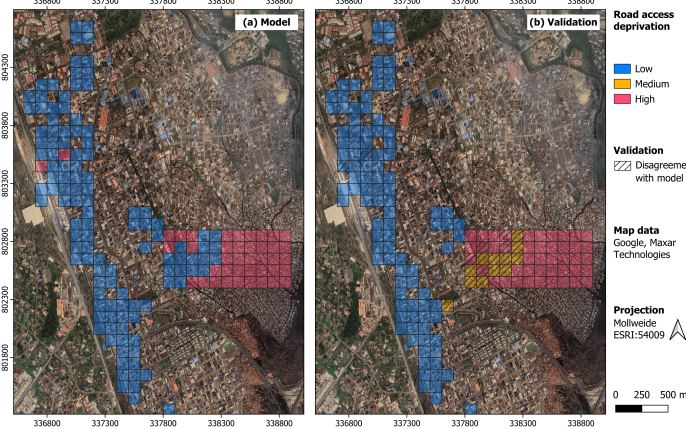


Figure 13: Region of interest L2: Visual comparison between (a) model results and (b) the data from validation with local communities in Lagos, Nigeria.

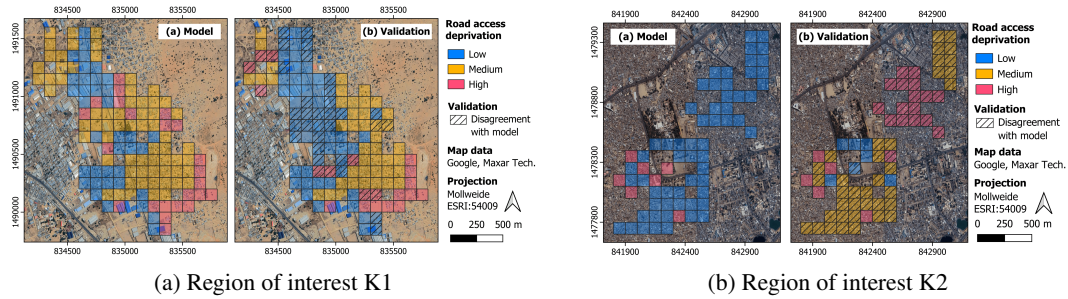


Figure 14: Visual comparison between (a) model results and (b) the data from validation with local communities in Kano, Nigeria.

4.4 Analysis of Grid Cells with Multiple Validations

Some grid cells received more than one validation by different community members. As described in Section 3.3, we applied majority voting to determine the final deprivation level of these multi-validation grid cells, assigning the most frequently chosen level. Grid cells with no clear majority (i.e., a tie between at least two deprivation levels) were disregarded for model evaluation.

To better understand the extent of agreement or disagreement among participants, we analyzed the variability in assigned deprivation levels for all multi-validated grid cells (including cells with no consensus). Figure 15 shows the variability using ternary plots. Each grid cell is positioned according to the proportions of its assigned validation levels. For example, a cell validated exclusively as low (low: 1, medium: 0, and high: 0) appears in the bottom-left corner of the triangle, while a cell with equal proportions of low and medium validations (low: 0.5, medium: 0.5, and high: 0) is placed

midway along the bottom axis. The color scale indicates the relative frequency of grid cells within different regions of the ternary plot.

For Nairobi (Figure 15a), most multi-validated grid cells were consistently validated as low deprivation, indicating strong agreement among the community members about low road access deprivation. We also observe a few grid cells with agreement for medium and high. On the other hand, disagreement mainly exists between low and medium, as indicated by the light blue area along the axis between low and medium. To a lesser extent, disagreement is observed between medium and high, and low and high. We observe similar patterns to Nairobi in Lagos (Figure 15b), with a strong agreement among community members for low deprivation and some disagreement between low and medium. For Kano, the ternary plot visualization is not applicable due to the small number of multi-validation grid cells ($n = 24$). For one-third of these cells, we observe agreement on low deprivation. For the other two-thirds, we observe disagreement between low and medium (3 cells) and medium and high (13 cells).

In summary, analyzing grid cells that were validated by more than one community member reveals that most often, there is agreement between validations. When there is disagreement, it typically exists between adjacent deprivation levels, meaning between low and medium or medium and high.

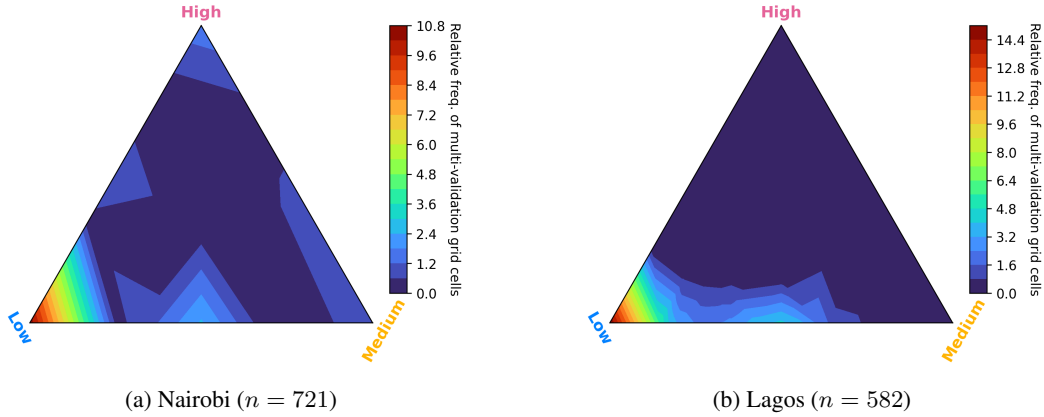


Figure 15: Ternary plots showing the variability in validations for grid cells that were validated more than once.

5 Discussion and Perspectives

5.1 Discussion

The results indicate that the proposed distance-agnostic road accessibility metric is effective at detecting extreme cases of deprivation, i.e., well-connected and severely disconnected areas. This is particularly apparent in Nairobi and Lagos, where overall accuracies and F1 scores for low deprivation exceed 77.0 % and 0.870, respectively. However, the consistent misclassification of medium road access deprivation suggests that modeling areas that fall between low and highly deprived areas remains challenging. This is most evident in Lagos, where nearly all medium-validated cells are predicted as low, and in Nairobi, where low-medium confusion is the dominant source of error despite comparatively balanced performance across classes. Two potential sources may explain these shortcomings for the medium class: (1) the conceptual model for medium deprivation does not fully align with community perceptions, and (2) the operationalization of this class, particularly its dependence on road surface type, introduces classification errors.

In the current conceptual model (Figure 2), medium road access deprivation mirrors the definition of low access deprivation in terms of road connectivity, i.e., most buildings are directly connected to motorable roads, but differs in that the quality of roads is lower. To capture this, we use the predominant surface type (paved or unpaved) within a grid cell as a proxy for road quality. The low F1 scores for medium deprivation, especially in Lagos (0.378), may indicate that surface type alone is insufficient to represent the quality dimension as perceived by local communities. Furthermore, the

qualitative results for towns located in the periphery of Kano indicate that with increasing distance to primary roads, road access deprivation levels are higher (Figures 14a, 20, and 21).

To discuss the shortcomings in the operationalization of the model, we assume that road surface type, indeed, differentiates low from medium road access deprivation. Then, the most likely source of error is the road surface type data. Specifically, we sourced automatically generated data from peer-reviewed studies (Zhou et al., 2024, 2025). However, mapping road surface types is challenging, leading to confusion between paved and unpaved road segments. In turn, these errors could have affected the distinction between low and medium road access deprivation in our model. The quantitative results largely support this hypothesis, showing that many grid cells modeled as low, were validated as medium by community members (see Figure 8). However, it should be noted that the validation data also indicates misclassifications beyond low-medium confusion.

Assuming that surface type is conceptually valid, a further explanation lies in the accuracy of the road surface type datasets used. While we sourced these from peer-reviewed studies (Zhou et al., 2024, 2025), mapping paved versus unpaved segments at scale remains difficult, and misclassifications may have directly affected the low-medium distinction. The confusion matrices (Figure 9) and alluvial plots (Figure 8) show this clearly: many cells modeled as low deprivation were validated as medium. In Kano, where medium deprivation achieves a comparatively high F1 score (0.618), this distinction appears more robust, though at the expense of very low performance for high deprivation ($F1 = 0.258$), which is frequently predicted as low or medium.

In general, our findings on the challenge of modeling deprivation beyond the dichotomy of slums and non-slums are also reflected in the existing literature on deprivation mapping. In particular, Dovey and Kamalipour (2017) emphasized the challenge of mapping areas that fall within strictly formal and informal categories. Furthermore, we encountered similar challenges when first developing a model to categorize areas into three levels of unplanned urbanization (Hafner et al., 2025). However, following the iterative model improvement framework proposed in IDEAMAPS Data Ecosystem Project Consortium (2025), this baseline road access deprivation model shows great potential to be improved based on the community validations.

Beyond the challenge of modeling medium deprivation, the current model does not consider natural barriers such as rivers, steep terrain, or constructed obstructions. In reality, such barriers can significantly affect accessibility even when road networks are technically present. Their exclusion may cause the model to underestimate access challenges in specific geographies.

Finally, the community validation data, though essential, is spatially limited and potentially influenced by subjective interpretations of deprivation. This is also reflected in the observed disagreements for some grid cells within the community-sourced validation data (see Section 3.3).

5.2 Perspectives

This study has several potential policy and community impacts, particularly for improving infrastructure planning and supporting locally grounded decision-making. The model provides actionable insights into where buildings lack adequate connection to motorable roads, allowing planners, civil society groups, and community advocates to prioritize upgrading efforts in the most deprived areas. By visualizing deprivation spatially, the approach also helps raise awareness of how limited road access affects daily life, including emergency response, mobility, and exposure to hazards, thereby strengthening efforts toward more inclusive and resilient urban development aligned with SDG 11. At the international level, the study demonstrates how open geospatial data, interpretable modeling, and community validation can be combined to produce context-sensitive assessments of infrastructure deprivation in data-scarce regions, offering a template for wider application while underscoring the importance of citizen-generated data and the need for continued capacity building and harmonization.

The findings of this study also highlight several promising directions for future research and application. First, there is considerable potential to model road access deprivation using Earth observation data, such as high-resolution satellite imagery. For example, Earth observation datasets, such as Sentinel-2 (Drusch et al., 2012) or PlanetScope (Roy et al., 2021), combined with transfer learning techniques, may enable continent-scale mapping of road quality with fewer data biases. Second, refinement of the conceptual road access deprivation model logic, particularly for the medium deprivation class, may better align model outputs with community perspectives and reduce the persistent low-medium confusion. This could include integrating additional proxies for road quality (e.g.,

drainage, width, or surface condition) or community-defined thresholds for what constitutes adequate access. Furthermore, the incorporation of road classes according to the formal road classification scheme, distinguishing between federal road (trunk A), state road (trunk B), and local government/community road (trunk C), could improve the model. Finally, the expansion of community validation mechanisms through participatory mapping campaigns remains essential. In this study, the extension of validation campaigns beyond the initial sessions substantially increased the amount of reference data available, improving both the spatial coverage and the robustness of the evaluation. Scaling such extended campaigns in future applications could not only enhance model accuracy but also ensure that local knowledge is more comprehensively integrated into deprivation assessments.

6 Conclusion

This study presents a road access deprivation model that combines a distance-agnostic accessibility metric, capturing how well buildings are connected to the road network, with road surface type data as a proxy for road quality. These two components together enable the classification of urban areas into low, medium, or high deprivation levels.

The model was applied to three cities, Nairobi, Lagos, and Kano, and evaluated using community-sourced validation data. In Nairobi, the model performs well for low and high deprivation, while medium remains challenging. In Lagos, performance is also high with strong results for low deprivation but weaker performance for medium and high levels. In Kano, accuracy is lower, with poor results for high deprivation.

These results suggest that, despite variations in performance across contexts, the approach provides a solid foundation for scalable road access deprivation mapping in data-scarce regions by leveraging openly available geospatial datasets. Furthermore, future work will focus on refining the model, particularly for medium deprivation and in accounting for informal paths or natural barriers. Consequently, this model lays the groundwork for future integration into broader informality and deprivation monitoring systems, supporting efforts to prioritize underserved communities and inform targeted interventions across rapidly urbanizing regions.

Acknowledgment

This report was generated as part of the IDEAMAPS Data Ecosystem project and the authors acknowledge the contributions from the stakeholders and partners communities in Nairobi (Kenya), Lagos (Nigeria), and Kano (Nigeria).

The work was supported, in whole or in part, by the Bill & Melinda Gates Foundation INV-045252. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

Computational resources for this work were provided by the Geospatial Computing Platform at the Center of Expertise in Big Geodata Science (CRIB), Faculty ITC, University of Twente. The authors thank the CRIB staff for maintaining the infrastructure and supporting scientific computing activities.

References

- Angela Abascal, Ignacio Rodríguez-Carreño, Sabine Vanhuyse, Stefanos Georganos, Richard Sliuzas, Eleonore Wolff, and Monika Kuffer. Identifying degrees of deprivation from space using deep learning and morphological spatial analysis of deprived urban areas. *Computers, environment and urban systems*, 95:101820, 2022.
- Shlomo Angel, Jason Parent, Daniel L Civco, Alejandro M Blei, et al. Making room for a planet of cities. 2011.
- Christopher Barrington-Leigh and Adam Millard-Ball. The world’s user-generated road map is more than 80% complete. *PloS one*, 12(8):e0180698, 2017.

- Luís MA Bettencourt and Nicholas Marchio. Infrastructure deficits and informal settlements in sub-saharan africa. *Nature*, pages 1–8, 2025.
- C Brelsford, T Martin, J Hand, and LMA Bettencourt. Toward cities without slums: Topology and the spatial evolution of neighborhoods. *sci. adv.* 4, eaar4644, 2018.
- Christa Brelsford, Taylor Martin, and Luís MA Bettencourt. Optimal reblocking as a practical tool for neighborhood development. *Environment and Planning B: Urban Analytics and City Science*, 46(2):303–321, 2019.
- M Burelle. Rural road connectivity and its effects on access to health care. *Int Affairs Rev*, 29(1): 1–22, 2021.
- Heather R Chamberlain, Edith Darin, Wole Ademola Adewole, Warren C Jochem, Attila N Lazar, and Andrew J Tatem. Building footprint data for countries in africa: to what extent are existing data products comparable? *Computers, Environment and Urban Systems*, 110:102104, 2024.
- Lewis Dijkstra, Hugo Poelman, and Paolo Veneri. The eu-oecd definition of a functional urban area. 2019.
- Kim Dovey and Hesam Kamalipour. Informal/formal morphologies. In *Mapping urbanities*, pages 223–248. Routledge, 2017.
- Matthias Drusch, Umberto Del Bello, Sébastien Carlier, Olivier Colin, Veronica Fernandez, Ferran Gascon, Bianca Hoersch, Claudia Isola, Paolo Laberinti, Philippe Martimort, et al. Sentinel-2: Esa’s optical high-resolution mission for gmes operational services. *Remote sensing of Environment*, 120:25–36, 2012.
- Sebastian Hafner, Qunshan Zhao, Angela Abascal, Manuella Comerio de Paulo, Grant Tregonning, Alexandra Middleton, Adenike Shonowo, Monika Kuffer, Ryan Engstrom, Dana R Thomson, et al. Towards an indicator-based morphological informality model for sub-saharan africa using open building footprint and road data (version 1). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 48:221–228, 2025.
- Mordechai Haklay and Patrick Weber. Openstreetmap: User-generated street maps. *IEEE Pervasive computing*, 7(4):12–18, 2008.
- IDEAMAPS Data Ecosystem Project Consortium. Ideamaps data ecosystem: incorporating justice in urban data science through participatory analytics. 2025.
- Kenya National Bureau of Statistics. 2019 kenya population and housing census volume 1: Population by county and sub-county. Technical report, Kenya National Bureau of Statistics, 2019.
- Divyani Kohli, Richard Sliuzas, Norman Kerle, and Alfred Stein. An ontology of slums for image-based classification. *Computers, environment and urban systems*, 36(2):154–163, 2012.
- Auwalu Faisal Koko, Muhammed Bello, and Muhammad Abubakar Sadiq. Understanding the challenges of 21st century urbanization in northern nigeria’s largest city, kano. In *Integrative Approaches in Urban Sustainability-Architectural Design, Technological Innovations and Social Dynamics in Global Contexts*. IntechOpen, 2023.
- Miljenko Lapaine. Mollweide map projection. *KoG*, 15(15.):7–16, 2011.
- Wanjing Li, Qi Zhou, Yuheng Zhang, and Yijun Chen. Visualising rural access index and not served rural population in africa. *Environment and Planning A: Economy and Space*, 54(2):215–218, 2022.
- Basirat Oyalowo. Implications of urban expansion: land, planning and housing in lagos. *Buildings & Cities*, 3(1), 2022.
- Martino Pesaresi, Marcello Schiavina, Panagiotis Politis, Sergio Freire, Katarzyna Krasnodebska, Johannes H Uhl, Alessandra Carioli, Christina Corbane, Lewis Dijkstra, Pietro Florio, et al. Advances on the global human settlement layer by joint assessment of earth observation and population survey data. *International Journal of Digital Earth*, 17(1):2390454, 2024.

- Sukanya Randhawa, Eren Aygün, Guntaj Randhawa, Benjamin Herfort, Sven Lautenbach, and Alexander Zipf. Paved or unpaved? a deep learning derived road surface global dataset from mapillary street-view imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 223: 362–374, 2025.
- David P Roy, Haiyan Huang, Rasmus Houborg, and Vitor S Martins. A global analysis of the temporal availability of planetscope high spatial resolution multi-spectral imagery. *Remote Sensing of Environment*, 264:112586, 2021.
- 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.
- Satej Soman, Anni Beukes, Cooper Nederhood, Nicholas Marchio, and Luís MA Bettencourt. Worldwide detection of informal settlements via topological analysis of crowdsourced digital maps. *ISPRS International Journal of Geo-Information*, 9(11):685, 2020.
- Dana R Thomson and Alejandro Blei. City Segments Layer v1, 2025. URL <https://doi.org/10.7910/DVN/XLRSF0>.
- UN-Habitat. Streets as tools for urban transformation in slums: A street-led approach to citywide slum upgrading, 2014. URL <https://unhabitat.org/streets-as-tools-for-urban-transformation-in-slums>.
- UN-Habitat. Lagos - city context report. Technical report, 2020. Part of the Global Future Cities Programme; Available online: <https://www.globalfuturecities.org/federal-republic-nigeria/cities/lagos>.
- Marylene Wamukoya, Damazo T Kadengye, Samuel Iddi, Claudious Chikozho, et al. The nairobi urban health and demographic surveillance of slum dwellers, 2002–2019: value, processes, and challenges. *Global Epidemiology*, 2:100024, 2020.
- Tayo Iyanda Yahaya and YM Suleiman. Urbanisation effect on the occurrence of urban heat island over kano metropolis, nigeria. 2017.
- Godwin Yeboah, João Porto de Albuquerque, Rafael Troilo, Grant Tregonning, Shanaka Perera, Syed AK Shifat Ahmed, Motunrayo Ajisola, Ornob Alam, Navneet Aujla, Syed Iqbal Azam, et al. Analysis of openstreetmap data quality at different stages of a participatory mapping process: Evidence from slums in africa and asia. *ISPRS International Journal of Geo-Information*, 10(4): 265, 2021.
- Qi Zhou, Zixian Liu, and Zesheng Huang. Mapping road surface type of kenya using openstreetmap and high-resolution google satellite imagery. *Scientific data*, 11(1):331, 2024.
- Qi Zhou, Yaoming Liu, and Zixian Liu. Mapping national-scale road surface types using multisource open data and deep learning model. *Transactions in GIS*, 29(1):e13305, 2025.

A Supplemental Material

A.1 OpenStreetMap Road Type Selection

In order to model road access deprivation, we consider roads that are accessible by a motorized vehicle in our model. Therefore, we subset the OpenStreetMap (OSM) road network to motorable roads. Specifically, we use the 'class' attribute to select the following road types: 'living_street', 'motorway', 'primary', 'residential', 'secondary', 'service', 'tertiary', 'trunk', 'unclassified'. Road types with class 'unknown' were also selected since the road type for many motorable roads is not specified in the data.

A.2 Additional Qualitative Results

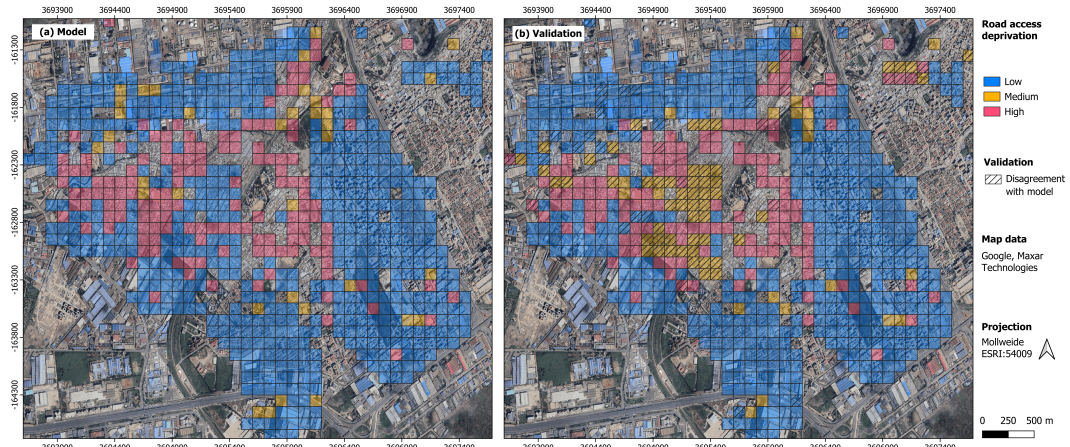


Figure 16: Region of interest N3: Visual comparison between (a) model results and (b) the data from validation with local communities in Nairobi, Kenya.

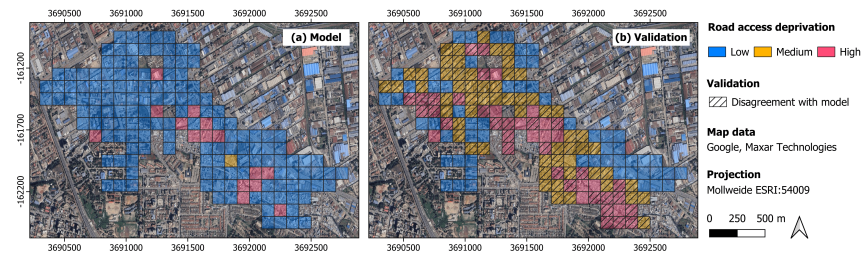


Figure 17: Region of interest N4: Visual comparison between (a) model results and (b) the data from validation with local communities in Nairobi, Kenya.

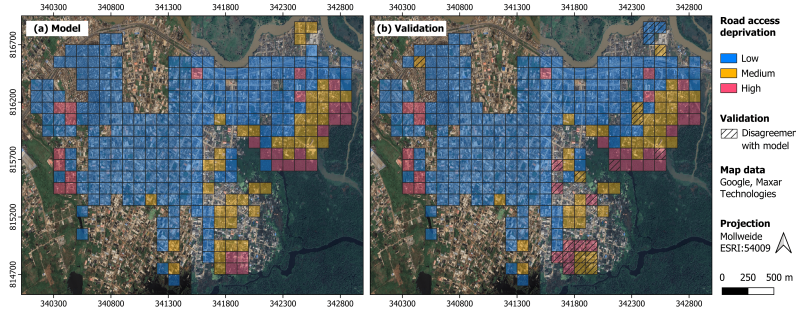


Figure 18: Region of interest L3: Visual comparison between (a) model results and (b) the data from validation with local communities in Lagos, Nigeria.

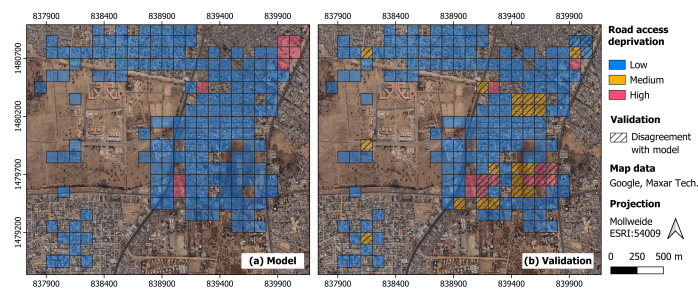


Figure 19: Region of interest K3: Visual comparison between (a) model results and (b) the data from validation with local communities in Kano, Nigeria.

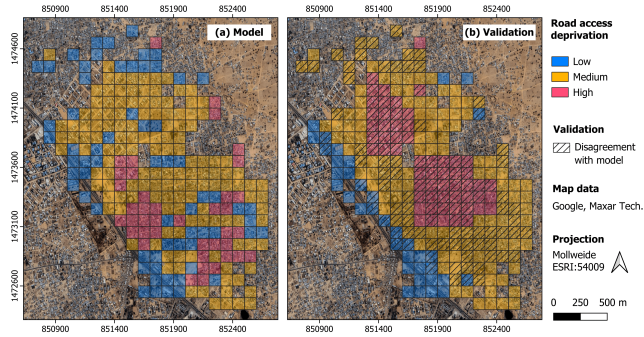


Figure 20: Region of interest K4: Visual comparison between (a) model results and (b) the data from validation with local communities in Kano, Nigeria.

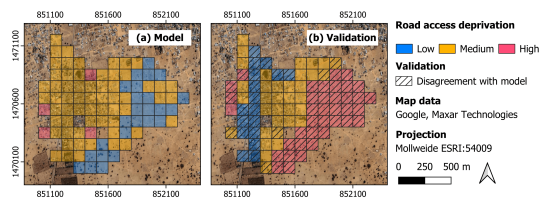


Figure 21: Region of interest K5: Visual comparison between (a) model results and (b) the data from validation with local communities in Kano, Nigeria.