

<https://doi.org/10.1038/s42949-024-00156-x>

# AI perceives like a local: predicting citizen deprivation perception using satellite imagery

Check for updates

Angela Abascal<sup>1,8</sup>✉, Sabine Vanhuyse<sup>2,8</sup>, Taïs Grippa<sup>2,8</sup>, Ignacio Rodriguez-Carreño<sup>3,4</sup>, Stefanos Georganos<sup>5</sup>, Jiong Wang<sup>6</sup>, Monika Kuffer<sup>6</sup>, Pablo Martinez-Diez<sup>7</sup>, Mar Santamaria-Varas<sup>7</sup> & Eleonore Wolff<sup>2</sup>

Deprived urban areas, commonly referred to as ‘slums,’ are the consequence of unprecedented urbanisation. Previous studies have highlighted the potential of Artificial Intelligence (AI) and Earth Observation (EO) in capturing physical aspects of urban deprivation. However, little research has explored AI’s ability to predict how locals perceive deprivation. **This research aims to develop a method to predict citizens’ perception of deprivation using satellite imagery, citizen science, and AI.** A deprivation perception score was computed from slum-citizens’ votes. Then, AI was used to model this score, and results indicate that it can effectively predict perception, with deep learning outperforming conventional machine learning. By leveraging AI and EO, policymakers can comprehend the underlying patterns of urban deprivation, enabling targeted interventions based on citizens’ needs. As over a quarter of the global urban population resides in slums, this tool can help prioritise citizens’ requirements, providing evidence for implementing urban upgrading policies aligned with SDG-11.

Urban inequality stands as a challenging social problem. Deprived areas, commonly referred to as ‘slums,’ have emerged as a tangible consequence of unprecedented urbanisation in Low-and Medium-Income Countries (LMICs) cities and manifest high levels of physical deprivation. Moreover, lower-income residents, especially in institutional and economically weak contexts, face additional deprivations like energy poverty and environmental risks, such as heat waves<sup>1,2</sup>. This emphasizes how physical deprivation is interconnected with other domains of deprivation that affect sustainable urban life<sup>3</sup>. In recent years, the application of Earth Observation (EO) methods, leveraging labelled satellite imagery and Artificial Intelligence (AI), has made considerable progress in capturing elements of urban appearance, including mapping urban elements within slums. In response to the “Leave No One Behind” principle (the central promise of the 2030 Agenda for Sustainable Development), a reliable understanding of the

physical deprivation levels in slums is urgently needed but not available for some of the most vulnerable communities.

EO’s extensive spatial coverage, temporal frequency, and high resolution provide cost-effective means of obtaining a synoptic and gapless view of urban areas<sup>4</sup>. Recently there has been a notable acceleration in the development of processing methods with the adoption of AI by the EO community, particularly machine learning (ML) algorithms, including highly complex deep learning (DL) algorithms<sup>5,6</sup>. Furthermore, transfer learning techniques have been employed to enhance the efficiency of DL models, especially when training data is limited<sup>7</sup>. The combination of advancement in AI methods with increased availability of big EO data has led to the production of global open geospatial layers, including layers of built-up areas, and urban extents<sup>8</sup>, and building footprints (e.g., Microsoft and Google initiatives<sup>9,10</sup>). Nevertheless, due to (i) the variability of deprivation across cities<sup>11–14</sup>, (ii) the lack of reliable reference data<sup>15</sup>, and (iii) the

<sup>1</sup>Navarra Center for International Development, Institute for Culture and Society, University of Navarra, 31009 Pamplona, Spain. <sup>2</sup>Department of Geosciences, Environment and Society, Université libre de Bruxelles, 1050 Brussels, Belgium. <sup>3</sup>Business Department, School of Business Administration and Economics, University of Navarra, 31009 Pamplona, Spain. <sup>4</sup>Institute of Data Science and Artificial Intelligence, University of Navarra, 31009 Pamplona, Spain. <sup>5</sup>Geomatics Unit, Department of Environmental and Life Sciences, Karlstad University, 65188 Karlstad, Sweden. <sup>6</sup>Faculty of Geo-Information Science and Earth Observation, University of Twente, 7500 AE Enschede, The Netherlands. <sup>7</sup>300.000 Km/s, 08005 Barcelona, Spain. <sup>8</sup>These authors contributed equally: Angela Abascal, Sabine Vanhuyse, Taïs Grippa. ✉e-mail: [maabascal@unav.es](mailto:maabascal@unav.es)

limited representation of deprived areas in the training sets, the accuracy of global geospatial layers remains lower in areas of high deprivation<sup>16</sup>.

Slums, which account for more than half of the population in most African countries (Fig. 1), are often omitted in global EO-based datasets due to limited in situ data, causing challenges to the stratification of the urban sample (e.g., in health or demographic data collections). The availability of reliable field data for annotating satellite images is crucial for training and evaluating ML and DL-based methods designed for the automated interpretation of EO data<sup>16</sup>. Given the rapid pace of urban growth, conventional field data collection practices proved inadequate to meet demand<sup>17</sup>. This limitation not only hinders the development of accurate EO-based models but also exacerbates urban inequality, impeding progress towards the Sustainable Development Goals (SDGs), particularly the goals “No Poverty,” “Reducing Inequalities,” and “Eliminating Slums” (SDG 1, 10, and 11 respectively). To address this challenge, it is imperative to explore more efficient approaches that integrate EO data, citizen science, and AI in data-scarce environments, enabling comprehensive and equitable assessments of urban areas.

Citizen science processes have played a crucial role in evaluating various aspects of urban appearance, including safety, cleanliness, liveability, and wealth<sup>18–20</sup>. Such citizen science assessments of urban spaces have predominantly relied on street-level photographs rather than EO imagery, primarily focusing on high-income countries<sup>21,22</sup>. This disparity stems from the limited availability of street-view coverage in LMICs, particularly in deprived areas where narrow roads impede vehicle access (Fig.2). This geodata gap underlines the persistent global challenge of unequal access to data, in which EO can play an important role.

Notably, the integration of EO data into citizen science assessments remains unexplored, presenting an opportunity to bridge the gap between EO data (globally available) and urban perception analysis. Accordingly, AI methods have not yet been able to replicate the nuanced perceptions of citizens, particularly in relation to the varying levels of deprivation experienced in slum areas. Therefore, this research aims to explore the integration of EO data, citizen science, and AI to assess urban deprivation levels comprehensively and equitably. Our contribution seeks to answer the following research questions:

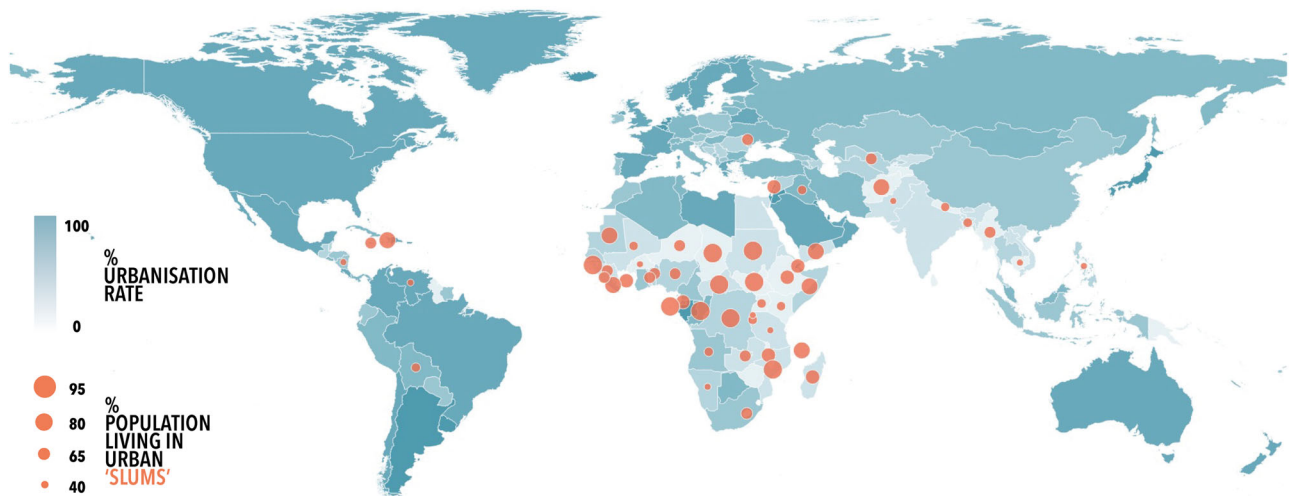
- (i) Can satellite imagery, rather than street-level imagery, serve as a reliable means of capturing perceived physical deprivation by citizens?
- (ii) Can AI, through satellite imagery, predict citizens’ deprivation perception?
- (iii) What are the features of the physical environment that most influence citizens’ perception of deprivation?

We have structured this manuscript for clarity and ease of navigation. Following this introduction, the results and discussion are presented synthesising our findings and interpretations. The subsequent sections delve into data and study area, and methods. Readers can refer to these sections for detailed insights.

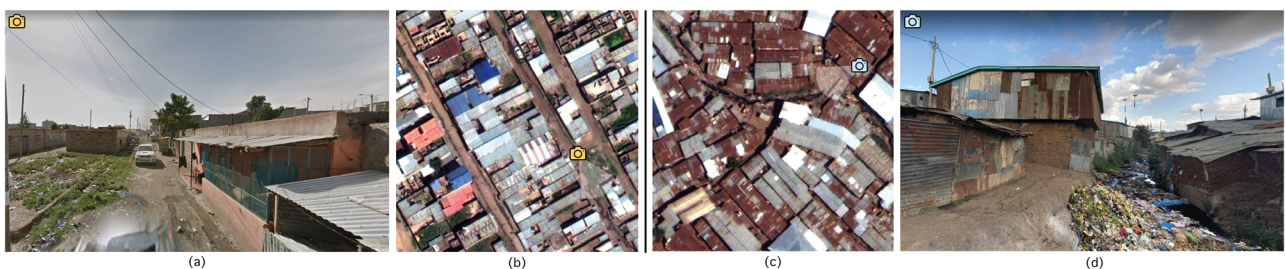
## Results

### Usage of EO imagery in citizen science assessments

The evaluation of urban spaces is generally based on the interpretation of street-level images<sup>21,22</sup>. However, due to limited global data coverage and greater scarcity of coverage in slum areas, we present satellite imagery as an



**Fig. 1 | Urbanisation rate and percentage of the urban population living in slums (<40% not represented).** The significant data gaps on slums in Africa pose a challenge to developing effective policies in the region where the issue is most pressing. Data source from United Nations Human Settlements Programme (UN-Habitat).

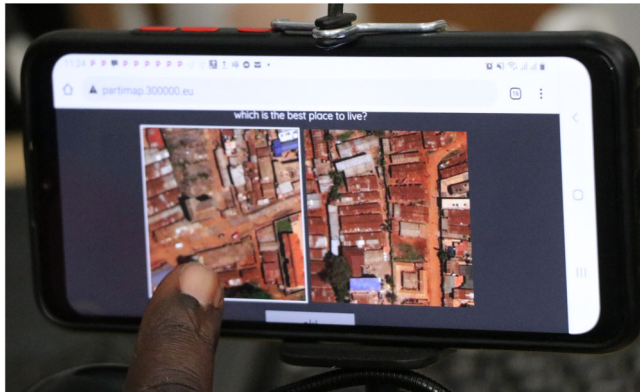


**Fig. 2 | Differences in the street network characteristics of Nairobi’s slums are visible in satellite images and street-level photos.** The street-level photo coverage varies depending on the accessibility by car, which is also an indicator of the level of deprivation. **a, b** A deprived area with a planned street network and hence good

street-level photo coverage. **(c, d)** A deprived area characterised by an extremely dense built-up fabric and narrow alleys, resulting in poor street-level photo coverage. Street-level photos: Google Street View © 2018 Google Maps. Satellite image subsets: WorldView-3 © 2019 Maxar Technologies.

effective alternative for assessing urban spaces (Fig. 2). In addition, image rating methods have been commonly used to identify differences in urban appearance<sup>23–28</sup>. However, these methods can introduce uncertainties when differences are subtle<sup>29</sup>. To overcome this issue, we employed a pairwise comparison method, which not only simplified the task for participants but also ensured a high level of consistency in their judgements<sup>30</sup>, as can be seen in Fig. 5.

We created a mobile website with a transactional database to enable several participants to vote simultaneously. A dataset comprising 1998 satellite image subsets (100 m × 100 m) was randomly paired, with each image compared to 15% of the total images in the set. As illustrated in Fig. 3, the spatial location and environment of the image displayed were not



**Fig. 3 | Pairwise comparison of satellite image subsets by citizen scientists.** Citizen scientists used a web application on their smartphones to visualize randomly paired satellite image subsets (100 m × 100 m) and vote, answering the question, ‘Which is the best place to live?’. Satellite image subsets: WorldView-3 © 2019 Maxar Technologies.

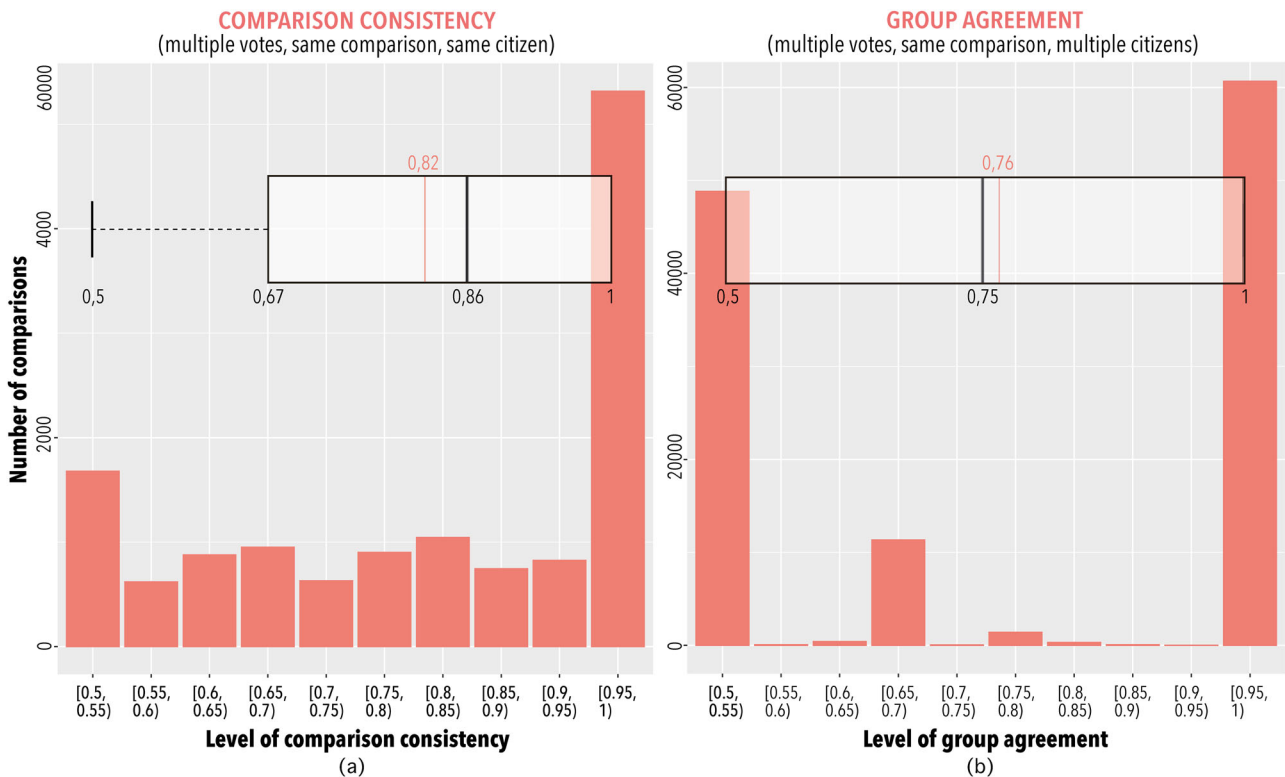
disclosed to the participants. In consultation with slum leaders, we excluded a tie option to prevent participants from intentionally casting ambiguous votes. Over one million votes were recorded, with each vote representing a comparison between a “winning” and a “losing” image.

Out of 1,089,302 pairwise comparisons, 629,027 were unique. Duplicates were used to assess *individual divergence* by comparing participants’ individual choices with the group’s opinion. Considerable divergence may indicate participants’ inattention during the workshop, as observed by the researchers. When there was inconsistency in individual choice regarding the same comparison (referred to as *comparison consistency*), a restrictive metric was applied. The group agreement was then checked and, if this also differed, *individual divergence* was not challenged. However, if the group agreed and individual choice varied, the individual was assumed to be voting inconsistently and the vote was penalised. In images with very subtle differences (usually due to spatial proximity), inconsistencies were seen in several of the individuals’ comparison consistency, as well as in the group agreement. Nevertheless, as can be seen in Fig. 4, there is a high individual consistency in each comparison, as well as a high consensus among citizens. Figure 4 shows the frequency distribution of the *comparison consistency*, and the *group agreement* metrics, with 0.5 being the least consistent to 1 being the most consistent. For a complete understanding of these metrics, see “Assessing citizen science” in the Methods section.

Figure 5 shows the frequency distribution of individual divergence, which follows an exponential pattern. Most participants showed low individual divergence and only five participants showed inconsistencies in their choices compared to the group’s opinion. Therefore, their votes were excluded as they could bias the final ranking of deprivation scores.

**From pairwise comparisons to a deprivation score**

Due to limitations in time and resources, conducting pairwise comparisons for all possible combinations of image subsets was not feasible, i.e., considering that for *t* objects and *n* judges, the required paired comparisons



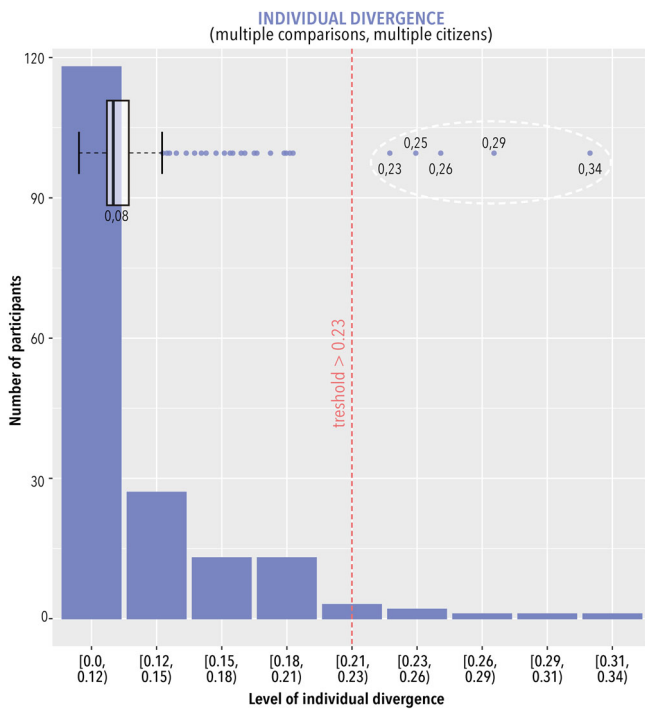
**Fig. 4 | Consistency of the citizens’ votes.** Consistency was assessed to determine if corrective measures were necessary. The results showed a high level of consistency at both the individual (a) and group levels (b). By merging bidirectional comparisons

(AB & BA) with unidirectional comparisons (AB), the x-axis represents the level of each metric, from 0.5 to 1.



would be  $n(t/2)$ . Instead, we adopted a robust Bayesian rating system called TrueSkill<sup>31</sup>. TrueSkill proved successful in previous studies for perception ranking and accommodates various match rules<sup>21,32</sup>. In our study, we implemented a free-for-all match rule, comparing all image subsets with each other without subgrouping. The pairwise comparison process ensures that each subset is shown an equal number of times to minimise bias and

maintain fairness in the assessment. Each subset was compared with about 315 others randomly selected. TrueSkill modelled the comparison as a  $N(\mu, 2)$  random variable, enabling both ranking and measuring the distance between subsets in the set. A normalisation process was applied to the ranking, transforming it into a standardised score ranging from 0 to 1. Each subset was subsequently assigned a deprivation score, with the highest score being identified as the “best place to live,” representing the least deprived area. Consequently, a detailed map depicting the comprehensive deprivation scores of slums was generated (Fig. 6).



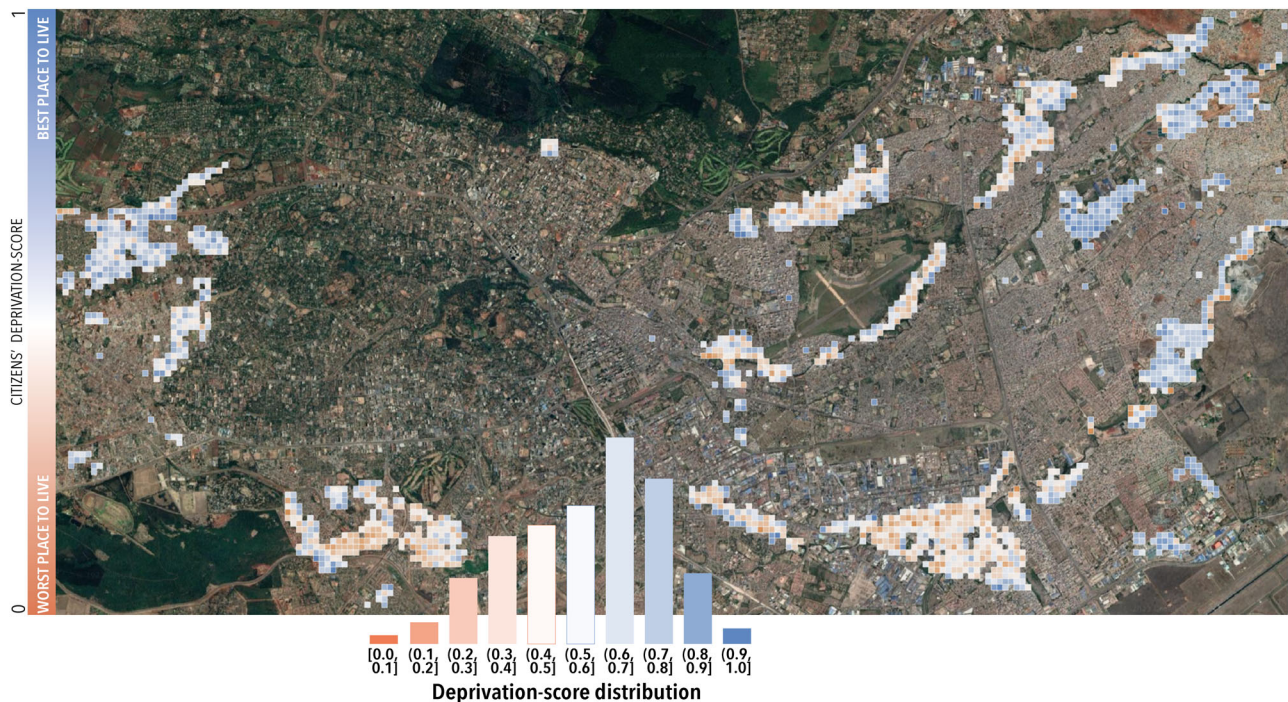
**Fig. 5 | Frequency distribution of individual divergence in the perception of urban deprivation.** i.e., disagreement between participants’ votes and the group’s opinion.

**AI to predict citizen deprivation perception**

We developed AI models employing satellite images to predict the scores derived from citizen votes. The results of the AI experiments confirm that “AI perceives as a local”, thus supporting the validity of the second research question. We compared the performance of DL and conventional ML models. The interpretation of the results was only feasible for conventional ML. Indeed, DL is generally recognised for its superior performance, but its main drawback for our application is its lack of interpretability.

While the accuracy of conventional ML models is comparable to that of DL models trained from scratch, employing a DL model pre-trained on ImageNet offered the advantage of automatic feature learning, which significantly enhanced the performance. Considering the limited number of observations, we took measures to prevent overfitting. We employed a 10-fold cross-validation to mitigate data scarcity, with a 90%-10% training-testing split. This ensured that all images were tested at least once while providing sufficient training data and improving performance evaluation while mitigating bias and overfitting. All experiments, both ML and DL, were repeated ten times for each fold. We report the highest score along with the mean, standard deviation, RMSE, and  $R^2$  for each fold, ensuring comprehensive testing of all images and minimising bias from a limited test set. Although DL methods are susceptible to overfitting with small data sets, this was not the case in our study. The loss plots of VGG trained from scratch and DenseNet121 fine-tuned with RGB channels show convergence of both curves and a flattened test trend indicating an absence of overfitting.

Table 1 presents the results of the DL models. The pre-trained DenseNet121 model has a clear advantage over the models trained from



**Fig. 6 | Deprivation score derived from citizens’ votes (spatial resolution: 100 m × 100 m).** The spatial variation reveals distinct geographic patterns of perceived deprivation. Satellite image: WorldView-3 © 2019 Maxar Technologies.



**Table 1 | Accuracy metrics relating to three DL models: VGG9 trained from scratch, DenseNet-121 trained from scratch, and pretrained DenseNet-121**

DL model scratch/pretrained (bands)	T max R <sup>2</sup>	T min RMSE	T mean R <sup>2</sup>	T mean RMSE	T sd R <sup>2</sup>	T sd RMSE	V max R <sup>2</sup>	V min RMSE	V mean R <sup>2</sup>	V mean RMSE	V sd R <sup>2</sup>	V sd RMSE
VGG9 scratch (RGB)	0.581	1.114	0.503	1.216	0.046	0.056	0.654	1.056	0.487	1.228	0.101	0.115
VGG9 scratch (RGNir)	0.615	1.067	0.532	1.179	0.066	0.083	0.716	0.958	0.520	1.185	0.101	0.110
Dense Net-121 scratch (RGB)	0.717	0.916	0.659	1.007	0.045	0.066	0.655	0.977	0.577	1.117	0.077	0.112
Dense Net-121 scratch (RGNir)	0.745	0.869	0.695	0.952	0.037	0.059	0.696	0.958	0.627	1.049	0.058	0.063
Dense Net-121 pretrained (RGB)	0.925	0.476	0.903	0.537	0.016	0.044	0.841	0.693	0.801	0.767	0.021	0.039
Dense Net-121 pretrained (RGNir)	0.950	0.383	0.886	0.576	0.038	0.101	0.822	0.734	0.789	0.790	0.021	0.027

For each model, two experiments were conducted with different band combinations (RGB and RGNir). The metrics include the maximum, mean and standard deviation of R<sup>2</sup> (coefficient of determination) and RMSE (root-mean square error), obtained on the training folds (represented with T) and on the validation fold (V).

**Table 2 | Accuracy metrics for the best-performing ML model in nine experiments (k-fold cross-validation)**

ML experiments – Best performing (features)	max R <sup>2</sup>	min RMSE	mean R <sup>2</sup>	mean RMSE	sd R <sup>2</sup>	sd RMSE
A – SVM rad (LOG)	0.577	1.158	0.468	1.262	0.063	0.067
B – SVM rad (LOG)	0.636	1.072	0.555	1.154	0.044	0.052
C – RF (STD)	0.636	1.089	0.559	1.150	0.043	0.056
D – SVM rad (LOG)	0.665	1.025	0.601	1.092	0.045	0.065
E – SVM rad (STD)	0.211	1.575	0.104	1.659	0.047	0.092
F – SVM rad (STD)	0.221	1.507	0.152	1.614	0.035	0.096
G – RF (LOG)	0.601	1.090	0.556	1.156	0.034	0.042
H – SVM rad (STD)	0.695	0.953	0.647	1.03	0.033	0.045
I – RF (LOG)	0.707	0.972	0.667	1.010	0.027	0.046

Each experiment (from A to I) involves a different set of input features, standardised (STD) or log-transformed (LOG), and the use of three types of ML algorithms (SVM, RF, XGBoost). In every experiment, the best-performing algorithm is either RF or SVM rad. The maximum (max), mean, and standard deviation (sd) of accuracy metrics (R<sup>2</sup> and RMSE) are calculated on the validation fold.

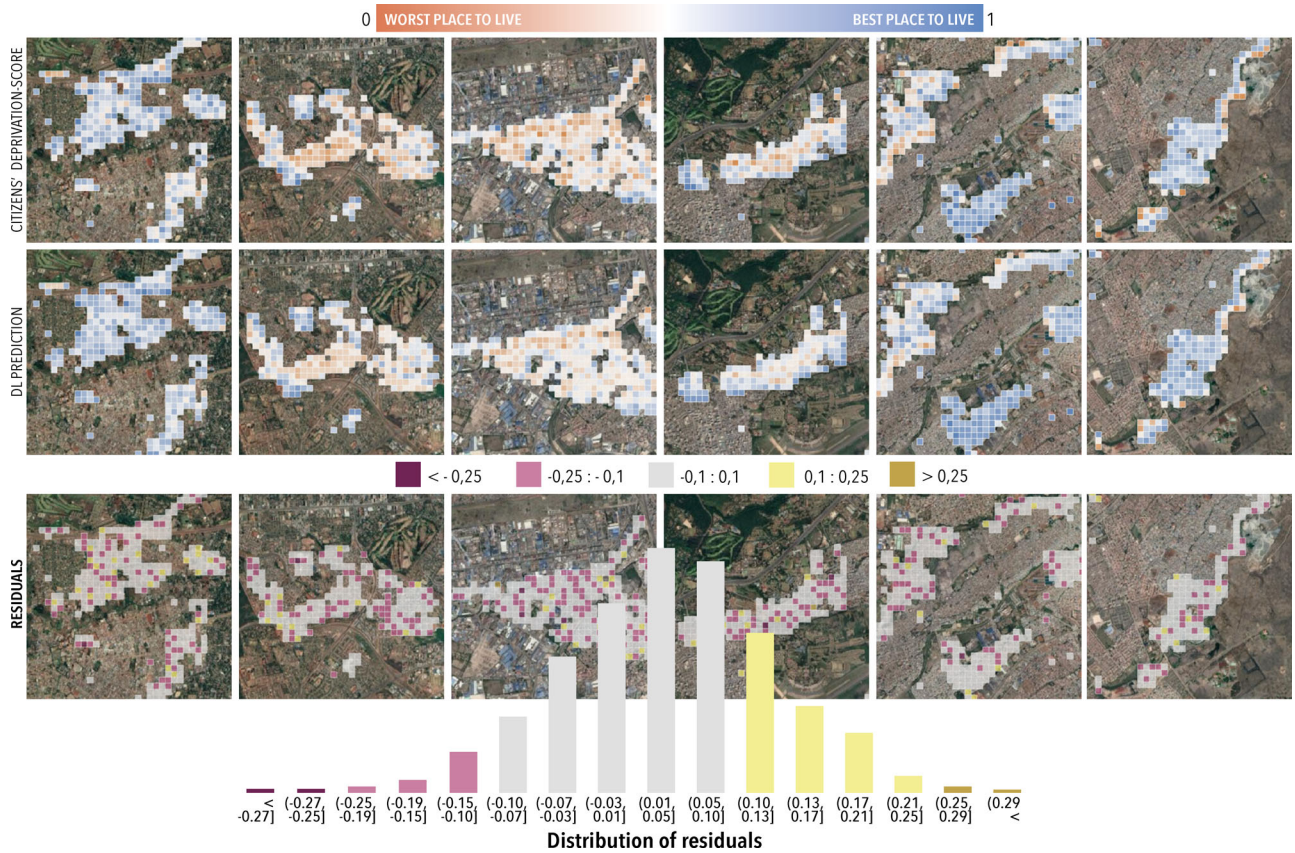
scratch, which notably show lower performance. The best result was obtained with the RGB band combination, despite the assumed importance of the near-infrared band. This can be attributed to either the difficulty of adjusting weights from the pre-training on an RGB dataset or citizens relying only on visible bands for assessing deprivation scores. It should be noted that the pretrained DenseNet-121 model exhibited significantly lower variance (R<sup>2</sup> standard deviation = 0.02) compared to the models trained from scratch. This suggests that the consistent initial weights from ImageNet and moderate modifications during training contributed to the pretrained model’s stability, while the random weight initialization of the VGG model made it more sensitive to partition-specific characteristics and prone to different local minima.

Compared to DL models, conventional ML models offer advantages such as lower computational requirements, the need for less training data, and higher interpretability. However, they require feature engineering, which is more time-consuming and may limit their transferability. Our study intentionally focused on a specific set of easily understandable features that represent physical variables of urban appearance. Regularisation and feature selection were conducted to reduce the number of features (168 in the larger set), aiming to prevent model overfitting and improve the interpretability of the results. Consistently, the results demonstrated that the best-performing models across all experiments were either Random Forest (RF) or Support Vector Machine with a radial kernel (SVM rad), as shown in Table 2. This finding aligns with previous research that achieved the best performance with the radial kernel<sup>33</sup>. Although the ML models did not reach the high R<sup>2</sup> values achieved by the pre-trained DL model, they exhibited significantly lower R<sup>2</sup> standard deviation, indicating reduced sensitivity to data partitioning and iterations and enhanced performance robustness.

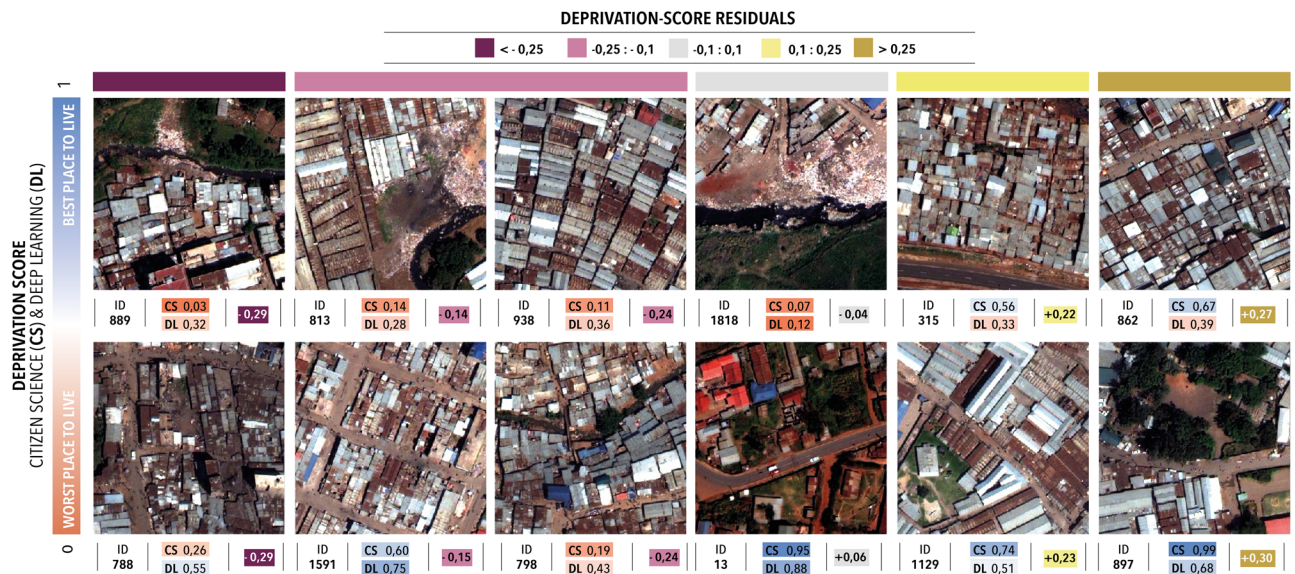
**AI geographical patterns of deprivation score**

In addition to the high accuracy in predicting the value of the deprivation scores (Table 1), the best DL model adequately extracts the overall geographic patterns of deprivation (middle line Fig. 7). For the DL prediction map, we used the validation set of each k-fold, i.e., 200 images not used in the k-fold training, to avoid representation bias with only the best model.

Finally, we analysed the performance of the DL model through the residuals. As shown in Fig. 7, the DL model predicts the citizen-based deprivation scores accurately. Generally, differences are smaller than 0.1, which is consistent across all deprived areas. We computed Morans’ I spatial correlation metrics and concluded that neither over- nor under-predictions follow a specific geographical pattern. However, some physical environment patterns are based on both under- and over-predictions. In the case of over-prediction, the residual (actual minus predicted) is negative. For instance,



**Fig. 7 | Geographic patterns of the deprivation score.** Top row: Score derived from citizen science. Middle row: Score predicted by the best DL model. Bottom row: Residuals (difference between citizen science-based and model-predicted scores). Satellite image subsets: WorldView-3 © 2019 Maxar Technologies.



**Fig. 8 | Satellite image subsets with their deprivation perception score derived from citizen science (CS) and deep learning (DL).** Negative residuals indicate that the deprivation score is higher according to citizens' votes than to model prediction,

and conversely. The highest divergences are illustrated, together with random samples from different slums. Satellite image subsets: WorldView-3 © 2019 Maxar Technologies.

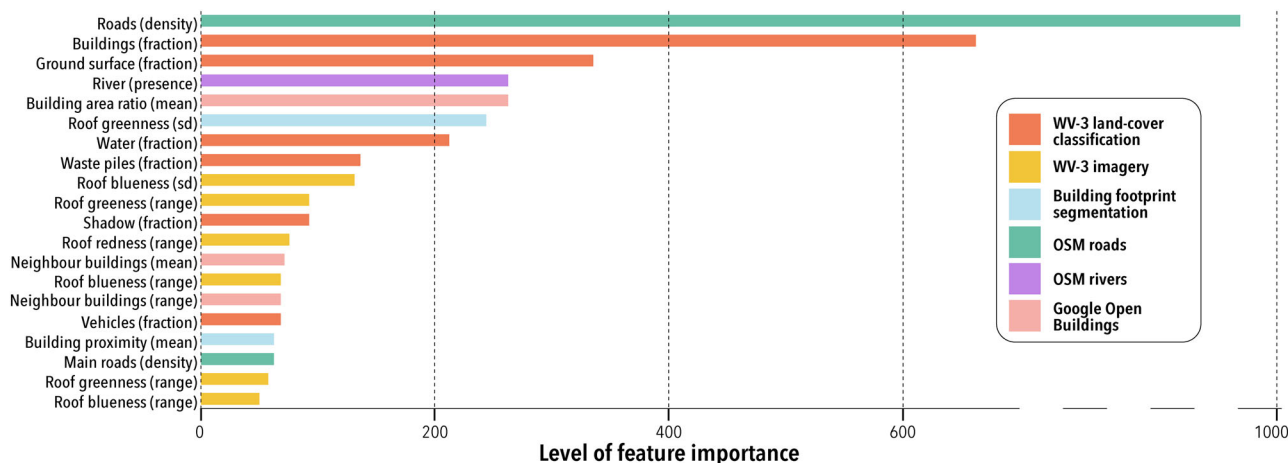
the over-predictions, represented at the left of Fig. 8, are mostly related to the presence of elements that citizens perceive as negative, e.g., very small buildings (ID 788 and ID 1591), the presence of a river (ID 889, ID 813 and ID 798 and) or the small waste sites (ID 889). As shown by the analysis of feature importance in the conventional ML experiments (Fig. 9), road density has a very strong influence on the deprivation score and could

weaken the influence of other relevant features captured by the DL model, such as building density, building size and roof colour (ID 862 and ID 315).

**Identifying deprivation features**

Feature importance was feasible only for conventional ML. It reveals the most influential factors driving predictions, which may help decision-





**Fig. 9 | Feature importance in the conventional ML experiment that provided the best accuracy.** This experiment involved a Random Forest (RF) model using the full set of log-transformed features. The colour indicates the dataset from which the features were derived. The features that have the strongest influence on the predicted

deprivation score were also identified by citizen scientists as having a strong influence on their perception of deprivation (e.g., street network quality, built-up fabric density, presence of open spaces, etc.).

makers and researchers identify areas in need of targeted interventions and allocate resources to address the underlying causes of deprivation and improve the well-being of the affected communities. In Experiment I, which used the full set of features and achieved the highest accuracy, we aimed to identify features having the strongest influence on the deprivation score, as illustrated in Fig. 9. Notably, relative road density emerges as the most influential, followed by building proportion and ground surface. These findings are in line with community opinion (collected via discussions) that well-connected streets, lower building density, and open spaces are indicators of lower deprivation. Additionally, the presence of rivers emerges as a significant predictor, given their association with flooding, waste accumulation, and potential insecurity. Remarkably, roof colour-related features are also good predictors, emphasising the significance of characterising building roofs in urban deprivation analysis. High visible greenness and blueness reflect the prevalence of corrugated iron sheets, common in deprived areas, while high visible redness indicates better-quality dwellings with tiled roofs. Moreover, the proportion of water bodies (linked to rivers) and waste piles stand out as strong predictors. These results validate the concerns raised by slum dwellers regarding waste piles, which are significant environmental and health threats to slum communities.

**Discussion**

Our research focuses on three key findings related to the connection between citizens’ perceptions of urban deprivation through EO data and the predictive capabilities of AI. First, we confirm that satellite imagery depicts important aspects of perceived urban deprivation in slums, overcoming the need for working with the incomplete coverage of street-level imagery. Second, automated DL and conventional ML tools accurately predict citizens’ deprivation scores. Finally, although the best-performing DL model outperforms the best-performing ML model, ML is essential for interpreting influential urban characteristics that define deprivation, such as road density, the presence of rivers and building density. Thus, the physical environment significantly influences citizens’ perceptions of deprivation, which can be accurately predicted using AI techniques. These insights indicate that the perception of deprivation in slums can be mapped and quantified on a global scale. Moreover, concerns also arise that are discussed in detail in the following subsections.

**EO and Citizen Science to address local needs**

EO, including satellite imagery and ready-to-use available global/continental geo-datasets, offers a consistent global coverage, opening opportunities to better understand and analyse urban deprivation in

LMICs. While implementing data-driven EO-based approaches, it is important to consider the digital divide in urban areas. Caragliu et al.<sup>34</sup> in their study ‘Smart Cities and the Urban Digital Divide,’ argue that, contrary to the common fear of smart cities increasing the urban digital divide, urban smartness is negatively associated with internal digital divides. This suggests that the use of advanced technologies in urban settings, when thoughtfully implemented, may not exacerbate but rather help bridge the digital divide, which is crucial for inclusive and effective citizen science approaches. EO combined with ML and DL techniques allows the characterisation of the physical environment, in particular deprivation levels, and, therefore, provides stakeholders and decision-makers with quantitative urban information. Nonetheless, implementing data-driven EO-based approaches without acknowledging the intrinsic needs of the citizens can hardly contribute to novel interdisciplinary knowledge<sup>16</sup>. This study provides insights into the importance of bridging the gap between EO state-of-the-art methods and citizen science by capturing perceived deprivation. This is highly relevant to EO research that involves socio-economic rather than biophysical measures. Testing the connection between citizens’ perceptions and the physical environment is a breakthrough for further developments, including the scalability and transferability of the proposed approach.

Citizen science guided us in understanding both the inputs and outputs of indicators, which can be challenging from only remote measurements. Participatory processes provide an essential local perspective, even more necessary in slums, whose residents are underrepresented in official data, documents, and measurements. Optimally, for participatory processes, a representative sample of the residents is needed. Our study worked with a non-random sample, which resulted in a potentially biased deprivation perception score. To mitigate this bias in future applications, it is recommended to include a random sample of participants.

Moreover, pairwise comparisons in a systematic experiment is a conceptually simple approach that allows for deriving a score with low levels of uncertainty. Overall, the derived deprivation score exhibits high individual consistency in comparisons (Fig. 4), although low levels of group agreement warrant attention (Fig. 5). Disagreements often arise due to the proximity of ranked images, posing challenges for participants to discern subtle differences equitably. Contextual factors, such as regional disparities and socio-economic variations, may also contribute to divergent perceptions of deprivation. Nevertheless, it is crucial to acknowledge that subjectivity and contextual influences do not compromise the objectivity and fairness of the deprivation assessment, as validated through objective AI predictions.



The user-friendliness of online platforms makes participatory processes very efficient and inclusive, with a broader geographic scope of participation and coverage than traditional methods (e.g., paper-based participatory mapping using satellite imagery). Indeed, there is high access to smartphones among the slum population, and consequently, the necessity to have access to a smartphone is currently not a barrier to participation. On the other hand, internet access can be an impediment as there is no free Wi-Fi access within the slums. This should be considered in online platform processes so that it does not prejudice certain participants.

### Decoding deprivation perception: from subjectivity to tangible metrics

The notion of urban deprivation has evolved over time, with different definitions influenced by historical periods and geographical contexts. Deprivation indices and frameworks first emerged in high-income countries, such as the UK and the US, before deprived areas proliferated in LMICs. Activists and researchers, such as Jane Jacobs or Alkire and Foster<sup>35</sup> have presented their visions of an ideal city, calling for equitable distribution and emphasising urban factors as characterising urban deprivation, such as narrow streets, lack of green space or lack of urban vitality. Objective metrics, such as green space per capita or population density, now inform urban planning decisions and strive to address urban deprivation. However, a global consensus on the validity of these metrics has yet to be reached, highlighting the highly subjective nature of deprivation shaped by cultural context and individual circumstances.

Intending to increase urban liveability, including access to essential services, safety, security, and environmental sustainability, it is essential to understand what factors define desirable urban places. The term urban deprivation means a state of lacking some of the urban factors, the opposite of liveable. To ensure clarity and avoid any potential negative connotations, the term “deprivation” was not explicitly used when working with the slum citizens. Instead, the question “Which is the best place to live?” was asked. However, through the research narrative, we are inclined towards the term deprivation, as it can support the development of citizen-centred policies and underlines our claim to fight urban inequalities.

Our work demonstrates that there is a consensus on the perceived ‘best places to live’ in slums among citizens. Only about 3% were “outliers”, i.e., participants whose opinion greatly diverged from the group’s opinion. These participants did not belong to a specific socio-economic group, which suggests that a lack of attention and/or commitment to the project was the cause of the divergences (which was observed by the research team during the workshops organised in Nairobi). Nevertheless, our findings of consensus contrast with Lenzi and Perucca’s<sup>36</sup> observation that city size influences individual discontent due to varying urban economies and diseconomies. This highlights the scaling of discontent across different urban scales and domains of urban life, adding complexity to our understanding of urban experiences.

ML indicates the relative importance of each feature in making a prediction of deprivation, relating to the quantification of citizens’ perceptions. Our findings open up new avenues of research on physical deprivation in a data-scarce context. DL lacks interpretability compared to ML techniques, nevertheless, we expect that this limitation will be transcended in the near future, considering the amount of attention it receives in current research<sup>37</sup>. Furthermore, the recent developments of explainable AI models that are meaningful and transparent for local users will be essential to optimize the usability of modelling results to guide urban policy development and monitoring of living qualities.

### AI predicted varying deprivation scores: limitations and strengths

Attempting to model urban deprivation, which is related to liveability, wealth, or even health outcomes, is vastly different from measuring concrete physical phenomena such as building footprints or land cover<sup>38</sup>. But, as citizens rely on tangible and measurable factors to vote for deprivation, it also implies a reliable and unbiased basis for AI modelling. Harnessing the

power of citizen science can be a rigorous approach to providing robust training data and overcoming AI data limitations. However, citizen science projects face the challenge of operating with limited resources, including low-cost and efficiency constraints.

Therefore, to overcome overfitting<sup>39</sup>, additional strategies, such as transfer learning, are necessary. We demonstrated that our pre-trained DL model largely overperformed the ML approaches with only 1998 labelled samples. Future research could explore the contribution of other pre-trained DL models and conduct specific benchmarking for this kind of task, which should account for complexity and parsimony along with predictive accuracy.

### Towards global scale mapping: replicability and transferability

Our research contributes to the study of slums in the city of Nairobi, and its coverage is limited to a single city. Future research will aim to apply our methods to other cities to assess their replicability and transferability and to produce additional datasets on urban deprivation, with the ultimate objective of achieving global coverage. While fieldwork should take place in a local context, EO models can be applied in areas with limited to no training data using domain adaptation and transfer learning methods. Nonetheless, attempting to transfer deprivation models from city to city is expected to be very challenging as urban deprivation can manifest very differently across spaces and cultures. However, many of the most important features are common factors that contribute to deprivation, e.g., lack of open space, high built-up density, or the presence of waste piles. To address the potential biases of AI systems for new cities, it is paramount to collect data that authentically reflects the diverse views of local residents. Therefore, when transferring this approach to other urban contexts, it will be important to ensure that participants represent the diversity of inhabitants, in terms of gender, age, socio-economic groups, and from diverse locations and urban contexts.

Moreover, the main bottleneck regarding EO data is the acquisition cost of VHR satellite imagery for covering large spatial areas with a relatively high temporal frequency (for frequent updates, considering the dynamic character of slums). Nonetheless, openly available datasets such as aerial maps<sup>40</sup>; and building footprints<sup>10,41</sup> can narrow the gap, as they can be combined with cost-free images (e.g., Sentinel 1 and 2), to map aspects of the urban morphology with remarkable accuracy. Moreover, although not publicly available, the potential of PlanetScope images for studying of urban deprivation is worth investigating, considering their spatial resolution and free access for research purposes. Satellite images can be processed with sophisticated models. Still, they can also be visually interpreted, even by non-experts with very little prior experience with satellite imagery, such as slum dwellers, as demonstrated in this study.

### Tool for policy and practical implications

Urban policies should embrace and encompass citizens’ knowledge to maximize their impact and benefits towards the most vulnerable communities. The developed AI tool, trained to predict citizens’ perceptions, has substantial relevance for policy and practice implications, as it offers the potential to highlight and understand urban inequalities. In this context, the insights from Caggiano et al.<sup>42</sup>, who emphasize the importance of measurable equity indicators in Climate Action Plans, resonate with our approach. Their findings align with our call for comprehensive tools that not only predict but also measure and address urban inequalities effectively. More concretely, the tool could facilitate a deeper understanding of residents’ needs, preferences, and concerns, thus encouraging greater citizen engagement and participation in urban development initiatives. This aspect reinforces the political discourse on citizen-centred approaches to urban planning and policymaking. Furthermore, the tool’s ability to analyse citizen perceptions enables urban planners to identify areas for improvement, allocate resources efficiently and design more liveable, accessible, and sustainable cities. Furthermore, the tool’s identification of citizens’ dissatisfaction or specific needs enables authorities to optimise the delivery of public services, allocate resources efficiently and improve the quality of

services. Moreover, by exploiting the predictions generated by the AI model, decision-makers gain valuable information on the potential impact of proposed urban development policies or projects. Finally, a possible continuous analysis of citizens' perceptions through the tool enables the monitoring and evaluation of urban policies and projects over time.

**Methods**

**Data and study area**

Since its early days, Nairobi has exhibited spatial patterns that reflect divisions based on social class, particularly as a result of urban colonial planning and "racial zoning"<sup>43</sup>. These divisions have led to strong and regulated segregation between the Central Business District (CBD) and residential areas for Europeans, Asians, and Africans. Presently, this urban segregation is evident in terms of spatial and economic deprivation<sup>44</sup>. Slums, where 60% of urban residents reside, are concentrated in a mere 4% of the total built-up area. The western part of Nairobi is characterised by greener areas and low building density and is home to the wealthier population<sup>45</sup>, as well as certain minority groups experiencing high levels of deprivation, such as the recognised slums of Waruku. In contrast, the central localities, where most of the slums are situated, are predominantly occupied by low- and middle-income groups.

Our study focuses on the complete coverage of slum areas within Nairobi, covering a total area of approximately 20 square kilometres. We adopted a grid-based strategy, dividing the slum areas into 1998 square grids of 1 ha each.

Five geospatial datasets were used: (1) A very-high resolution WorldView-3 (WV3) image (Table 3), (2) a land-cover classification

derived from this WV3 image<sup>38</sup>, (3) OpenStreetMap (OSM) roads and (4) OSM rivers<sup>46</sup>, and (5) Google Open Buildings<sup>10</sup>. Every dataset was used to derive features for the conventional ML models, while the DL models employed only WV3 image bands. A WV3 natural colour composite was clipped using the 100 m × 100 m grid to produce the image subsets shown in pairs to slum citizens.

**General workflow**

The research follows the workflow outlined in Fig. 10. Firstly, we demonstrate the feasibility of relying on citizen science and satellite data to assess physical deprivation levels and generate a deprivation score. Secondly, we assess the potential of AI to predict the deprivation score from EO features derived or directly from the satellite image. To do so, we build a set of DL and conventional ML models and compare their respective performances. Finally, we examine the relative importance of the features provided by the best ML model to identify the most relevant urban characteristics for defining deprivation levels.

**Engagement of slum dwellers in citizen science**

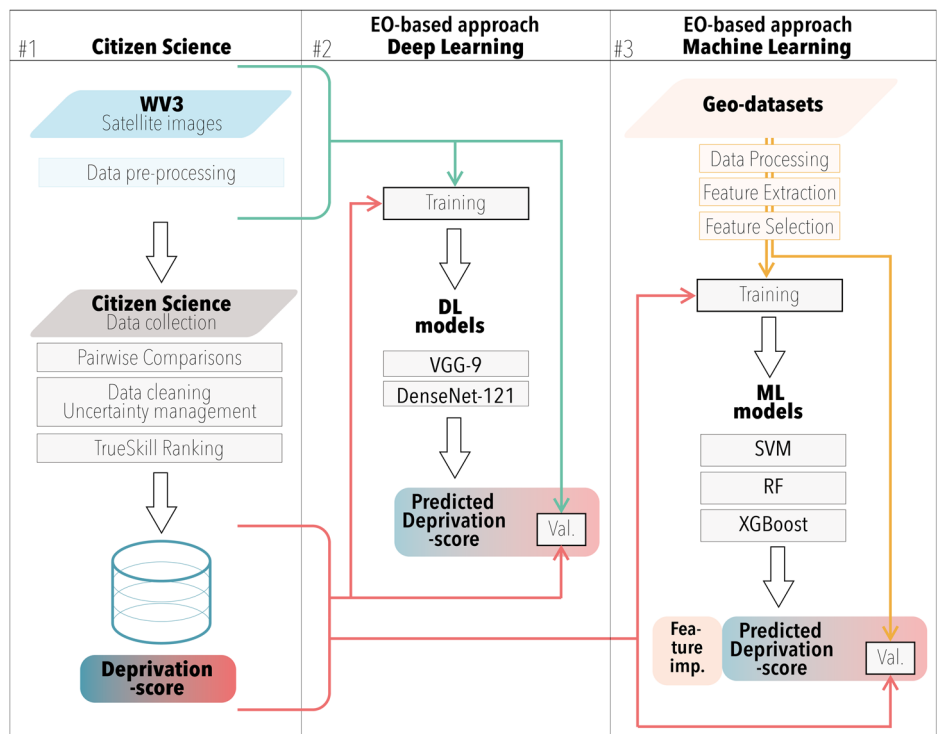
Due to the challenges associated with on-site data collection, slum citizens are often overlooked or under-sampled. This leads to incomplete census data in slums and hence a lack of reliability of global population datasets modelled from censuses<sup>47,48</sup>. The slum population is considered a hard-to-reach group for several reasons: (i) difficulty in reaching the participants, as working hours are often atypical and changing; (ii) complex urban morphology making data collection and general orientation in settlements

**Table 3 | World View-3 images specifications**

Bands	Spatial resolution	Geometric processing level	Radiometric processing level	Date
Panchromatic	0.30 m	Orthorectified	Calibrated	Mosaic of 2 images (13/01/2019 and 01/02/2019)
8 multispectral bands	1.20 m	Orthorectified	Calibrated	Mosaic of 2 images (13/01/2019 and 01/02/2019)

The table provides specifications for the WorldView-3 images, including details about the bands, spatial resolution, acquisition date, and levels of geometric and radiometric processing. The geometric processing level of the images is described as "Orthorectified," indicating that geometric distortions caused by sensor position and terrain variations have been corrected, to ensure accurate spatial representation. The radiometric processing level is mentioned as "Calibrated," indicating that the images have undergone atmospheric corrections to provide accurate and consistent pixel values.

**Fig. 10 | General workflow of the study.** In Phase 1, termed 'Citizen Science,' the deprivation-score database was generated through pairwise comparison of satellite images by slum residents. In Phase 2, 'EO-based Deep Learning Approach,' various deep learning models were developed, trained, and validated using the citizen science data alongside satellite imagery. Finally, in Phase 3, 'EO-based Machine Learning Approach,' different machine learning models were employed, trained, and validated with the citizen science data and additional geo-datasets.





**Fig. 11 | Participants’ socio-economic profiles for each of the seven studied slums.** The pie charts illustrate key socio-economic indicators that define the participants’ profiles, namely their gender, age, education level, employment, access to tap water, and access to toilets. Each slum is represented in a separate column.

challenging; (iii) difficulty to engage local communities as there is a lack of trust that the survey will produce any local benefit; (iv) potential security issues due to the presence of local street gangs<sup>49</sup>.

To enhance the reliability and validity of the citizen science protocol, proactive measures in the experimental design proved successful in addressing the challenges associated with reaching and engaging the slum population. These measures included early engagement with local community leaders, providing comprehensive project information, and overcoming educational and technological barriers. Over three months, we had a minimal drop-off in participation, with 186 participants actively engaged in the project. While our study included participants with diverse socio-economic profiles, considering gender diversity and ensuring representation from different areas within the slums, it is important to note that they were not randomly sampled. As a result, the findings do not provide a statistical representation of the entire slum population (Fig. 11).

**Citizen science assessment**

*Individual divergence* metric helps understand how much a participant’s opinion varies from the group’s opinion (*group agreement*) and whether the individual choices within the same comparison are consistent (*comparison consistency*). We calculate it by averaging the value of the absolute difference between group agreement and comparison consistency for the same comparison and individual:

$$\text{Individual divergence} = \frac{1}{n} \sum_{i=1}^n |(group\ agreement)_i - (comparison\ consistency)_i| \tag{1}$$

where *i* varies from 1 to *n*, and *n* stands for the comparisons done by an individual.

Comparison consistency examines whether each individual is consistent in their own choices. Participants are shown the same comparison

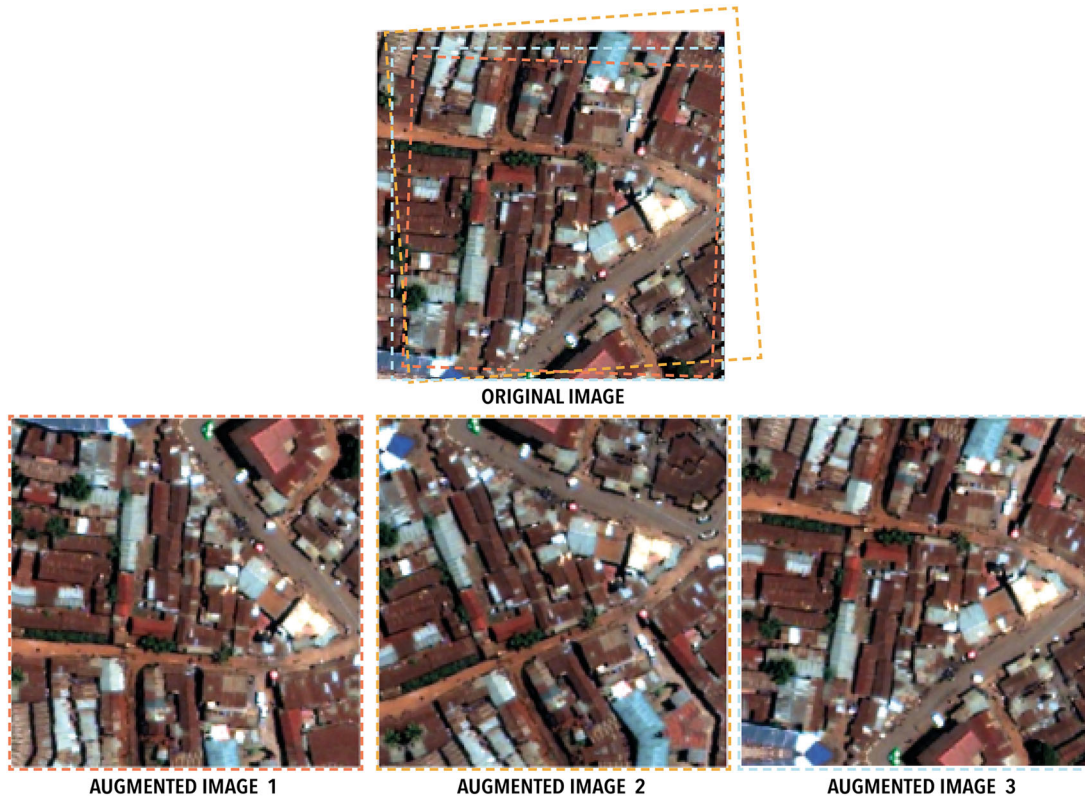
several times, and their consistency in upvoting the same image within a comparison is evaluated. This is defined as, comparison consistency =  $x / (x + y)$ , where *x* and *y* stand for the number of votes for each of the two images involved in a comparison, and *x* is the image with the larger share of votes in the comparison (avoiding double-counting, i.e.,  $x + y = 1$ , and making the range from 0.5 to 1). Only the comparisons repeated more than twice by an individual are used for the analysis.

Group agreement measures the agreement between participants in upvoting the same image in a comparison. The only difference between group agreement and comparison consistency is that the former is not calculated per individual. It is given by, group agreement =  $X / (X + Y)$ , where *X* and *Y* stand for the number of votes for each of the two images involved in a comparison, and *X* is the image with the larger share of votes in the comparison (which, similarly as before, allows for the measure to avoid double counting and makes it range from 0.5 to 1). Only comparisons voted for more than twice are considered for the analysis.

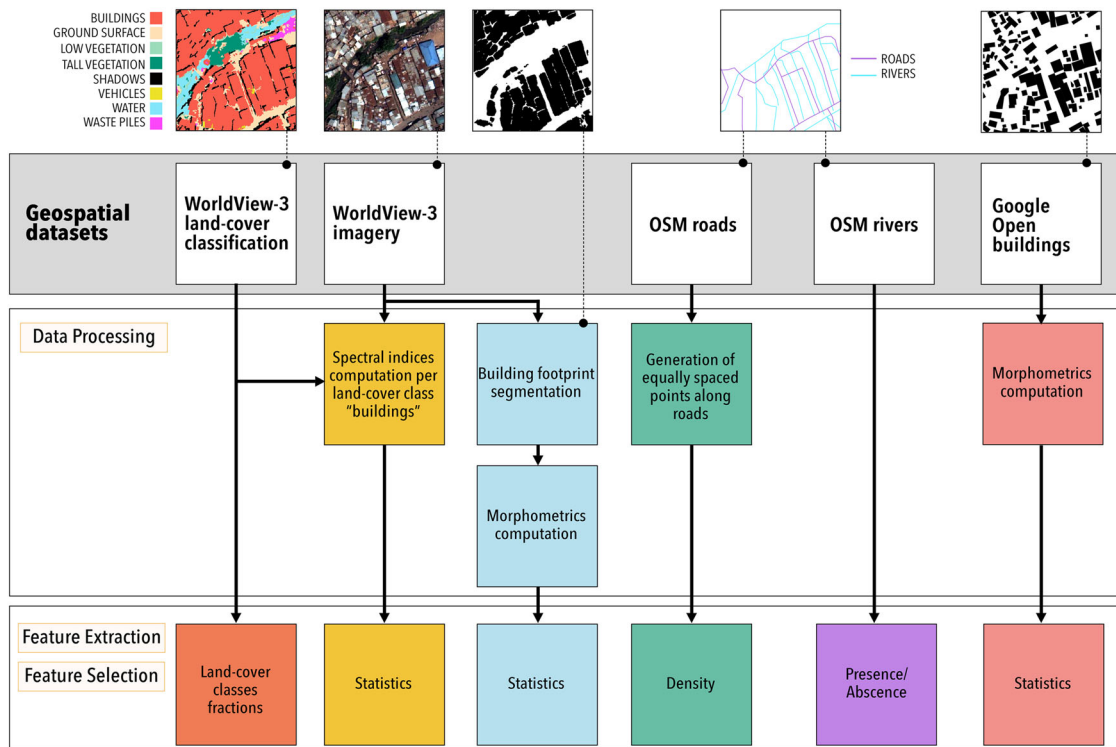
**Deep learning**

To model deprivation scores with DL models, we employed two different Convolutional Neural Networks (CNN) in our experiments: VGG and DenseNet-121. The first approach involved training a VGG-like model from scratch on the WV3 images. VGG is an architecture proposed by Simonyan et Zisserman<sup>50</sup>, based on a series of convolution blocks separated by max-pooling layers, and ending with flattened and dense layers. While it is considered a deep network, it has the advantage of being easily trainable without facing the vanishing gradient problem. For our experiments, we designed a small VGG-like model with a depth of nine layers: three convolution blocks (composed of two convolutions with 32 filters in the first block, two convolutions with 64 filters in the second, and three convolutions with 128 filters in the last block). This architecture was designed after several trial-and-error attempts. We used ReLU activation and batch normalisation after each convolution and a max-pooling layer at the end of each block. The





**Fig. 12 | Original image subset (100 m × 100 m) and its augmented versions.** Only moderate augmentations were used (vertical flip, 20% random rotation and 5% random translation, with missing values filled using the reflection mode) so that the augmented versions do not differ greatly from the original version. Satellite image subsets: WorldView-3 © 2019 Maxar Technologies.



**Fig. 13 | Workflow designed to produce the features used in the conventional ML experiments.** Feature selection is essential to prevent model overfitting and improve the interpretability of the results. Satellite image subset: WorldView-3 © 2019 Maxar Technologies.

convolution blocks are followed by two dense layers with 256 neurons each. For the regression task, the output layer consists of a dense layer with one neuron and a linear activation function. The model is trained for 400 epochs with a learning rate of 0.01 using the Adam optimizer. The mean squared error is used as the loss function, and the batch size is set to 128.

The other approach uses the DenseNet architecture, proposed by Huang et al. (2018), which is deeper and more complex than VGG. This architecture is composed of several dense blocks separated by transition layers. Each dense block is made of consecutive convolution layers whose output feature maps are concatenated to all the subsequent convolution layers in the block. The advantage of this approach is that it increases the depth (and thus, has the potential to improve the accuracy) of CNN models without facing the vanishing gradient problem. We used the implementation of DenseNet-121 available in the Keras Python library which provides a version of the model pre-trained on ImageNet. We added a dense layer of 1024 neurons to the main architecture, followed by a regression head with a single node and linear activation. The mean squared error was used as the loss function, and the batch size was set to 128. For our experiment, where the model was trained from scratch, the model was trained with a learning rate of 0.001 for 200 epochs, using the Adam optimizer. In the transfer learning approach, fine-tuning of our dataset was performed in two stages using the Adam optimizer: first, the pre-trained weights were frozen, and the regression head was trained for 15 epochs with a learning rate of 0.001. In the second stage, pre-trained weights were unfrozen and trained for 200 epochs with a smaller learning rate (0.0001) to avoid losing the information contained in the pre-trained weights.

Furthermore, on DenseNet121 architecture, we applied a transfer learning approach using a pre-trained model on ImageNet that was fine-tuned on our dataset. The initial input of 333×333 pixels (image patches of 1 ha, spatial resolution of 30 cm) was resampled to 224×224 pixels to match the patch size of the DenseNet121 model pre-trained on ImageNet. Input data were standardised, with zero mean and unit variance. Since our dataset is small and we must avoid overfitting, we used data augmentation, which replaces the original training set with a randomised version at each epoch (Fig. 12). We conducted experiments using two combinations of three image bands, namely RGB and RGNir. The RGB combination is spectrally similar to imagery that can be obtained from Google Earth, while the RGNir combination includes a near-infrared band known to be useful for discriminating active vegetation.

### Machine learning

ML experiments involve several stages: data processing, feature extraction, feature selection, ML regression modelling, and feature importance analysis (Fig. 10). Features were extracted from five geospatial datasets in 100 m × 100 m grid cells, as shown in Fig. 13. The datasets derived from very-high resolution satellite imagery and publicly available building footprints were created and outlined in earlier studies<sup>11,38,46,51</sup>.

Nine ML experiments were designed in which different combinations of features were used. In each experiment, feature selection was implemented as a two-step process to prevent overfitting and to improve the interpretability of the regression results. First, we regularised the models with the least absolute shrinkage and selection operator (Lasso)<sup>52</sup> using the glmnet R package for regularisation. Lasso is a type of linear regression that penalises the sum of coefficients' absolute values. A penalty hyperparameter, whose optimal value is found using cross-validation, determines the extent of coefficient shrinkage. Using Lasso, the coefficients of useless features were set to zero. In the second step, we performed feature selection on the features with a non-zero coefficient by computing Pearson's correlation matrix to identify feature pairs having a correlation coefficient greater than 0.8 in absolute value. In each pair, we discarded the feature having the lowest coefficient. As an additional selection criterion, we prioritised the mean over the median, as it better captures variability and is, therefore, more suitable for ML methods.

We created two sets of transformed features, the first through standardisation and the second through log-transformation, and we used both sets in the ML experiments. We compared three widely used conventional

ML regressors in each experiment: (i) Support Vector Machines (SVMs) with different kernels: Radial Basis Function (RBF), Linear Kernel and Polynomial Kernel<sup>53</sup>; (ii) Random Forest (RF)<sup>54</sup>; and (iii) eXtreme Gradient Boosting (XGBoost)<sup>55</sup>. We used the R caret package and optimised the hyperparameters of the models using the tuneLength parameter of the train function. Each ML regressor was run in each experiment, using, in turn, each transformed feature set.

We analysed the significance of urban characteristics defining urban deprivation through the ML experiment that provided the best accuracy. This involved employing the RF algorithm's capacity to assess the influence of each feature on the predictive outcomes.

### Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

### Data availability

The datasets generated during the study are available in the Zenodo repository (<https://doi.org/10.5281/zenodo.10881208>). This includes the deprivation perception scores derived from Citizen Science data and the best predictors (interpretable features) from the best conventional ML model.

### Code availability

The codes generated during the study are available in the Zenodo repository (<https://doi.org/10.5281/zenodo.10881208>). This includes the code of the Citizen Science web interface, the Citizen Science assessment, and the best DL and conventional ML models.

Received: 10 January 2023; Accepted: 14 March 2024;

Published online: 29 March 2024

### References

1. Furszyfer Del Rio, D. D., Sovacool, B. K., Griffiths, S., Foley, A. M. & Furszyfer Del Rio, J. A cross-country analysis of sustainability, transport and energy poverty. *npj Urban Sust.* **3**, 41 (2023).
2. Chen, M. et al. Rising vulnerability of compound risk inequality to ageing and extreme heatwave exposure in global cities. *npj Urban Sustainability* **3**, 38 (2023).
3. Abascal, A. et al. Domains of deprivation framework" for mapping slums, informal settlements, and other deprived areas in LMICs to improve urban planning and policy: A scoping review. *Comput. Environ. Urban Syst.* **93**, 101770 (2022).
4. Arribas-Bel, D., Rowe, F., Chen, M., & Comber, S. An image library: The potential of imagery in (quantitative) social sciences. Chapters, 528–543, (2022).
5. Camps-Valls, G., Tuia, D., Xiang Zhu, X., Reichstein, M. Deep Learning for Earth Sciences. Wiley: Hoboken, (2021).
6. Wu, G. et al. A boundary regulated network for accurate roof segmentation and outline extraction. *Remote Sens.* **10**, 1195 (2018).
7. Tan, C. et al. A Survey on Deep Transfer Learning. Lecture Notes in Computer Science, 270–279, (2018).
8. European Commission, GHSL- Global Human Settlement Layer, <https://ghsl.jrc.ec.europa.eu/dataToolsOverview.php> (2020).
9. Yang, S. How to extract building footprints from satellite images using deep learning, <https://azure.microsoft.com/en-us/blog/how-to-extract-building-footprints-from-satellite-images-using-deep-learning/> (2018).
10. Sirko, W. et al. Continental-Scale Building Detection from High Resolution Satellite Imagery, arXiv:2107.12283v2, (2021).
11. Abascal, A. et al. Identifying degrees of deprivation from space using deep learning and morphological spatial analysis of deprived urban areas. *Comput. Environ. Urban Syst.* **95**, 101820 (2022).
12. Duque, J. C., Patino, J. E., Ruiz, L. A. & Pardo-Pascual, J. E. Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. *Landscape Urban Plan.* **135**, 11–21 (2015).

13. Kuffer, M., Barros, J. & Sliuzas, R. V. The development of a morphological unplanned settlement index using very-high-resolution (VHR) imagery. *Comput. Environ. Urban Syst.* **48**, 138–152 (2014).
14. Sliuzas, R. & Kuffer, M. Analysing the spatial heterogeneity of poverty using remote sensing: typology of poverty areas using selected RS based indicators. *Remote Sensing—New Challenges of High Resolution*, Bochum, 5–7, (2008).
15. Kohli, D., Stein, A. & Sliuzas, R. Uncertainty analysis for image interpretations of urban slums. *Comput. Environ. Urban Syst.* **60**, 37–49 (2016).
16. Schmitt, M., Ahmadi, S. A. & Hänsch, R. There is no data like more data—current status of machine learning datasets in remote sensing. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 1206–1209, (2021).
17. Liu, Y., Liu, A., Wang, T., Liu, X. & Xiong, N. An intelligent incentive mechanism for coverage of data collection in cognitive Internet of Things. *Fut. Generat. Comput. Syst.* **100**, 701–714 (2019).
18. Jacobs, J. *The Death and Life of Great American Cities*. Vintage; Reissue edition, (1961).
19. Wilson, J. Q. & Kelling, G. L. Broken Windows: The Police and Neighborhood Safety. *The Atlantic* <https://www.theatlantic.com/magazine/archive/1982/03/broken-windows/304465/> (1982).
20. Keize, K., Lindenber, S. & Steg, L. The Spreading of Disorder. *Science* **322**, 1681–1685 (2008).
21. Salesses, P., Schechtner, K. & Hidalgo, C. A. The Collaborative Image of The City: Mapping the Inequality of Urban Perception. *PLoS ONE* **8**, e68400 (2013).
22. Martínez-Diez, P., & Santamaria-Varas, M. Arturo. How can citizens train an algorithm? [https://300000kms.net/case\\_study/arturo/](https://300000kms.net/case_study/arturo/) (2018).
23. Wilson, R. L. Livability of the city: attitudes and urban development. *Urban Growth Dynamics*, 359–399, (1962).
24. Herzog, T. R., Kaplan, S. & Kaplan, R. The prediction of preference for familiar urban places. *Environ. Behav.* **8**, 627–645 (1976).
25. Milgram, S. The image-freezing machine. *Society* **14**, 7–12 (1976).
26. Roth, M. Validating the use of Internet survey techniques in visual landscape assessment—An empirical study from Germany. *Landscape Urban Plan.* **78**, 179–192 (2006).
27. Wherrett, J. R. Creating landscape preference models using internet survey techniques. *Landscape Res.* **25**, 79–96 (2000).
28. Liu, L., Silva, E. A., Wu, C. & Wang, H. A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Comput. Environ. Urban Syst.* **65**, 113–125 (2017).
29. David, H. A. The method of paired comparisons. *Griffin's Stat. Monographs Courses* **12**, 120 (1988).
30. Clark, A. P., Howard, K. L., Woods, A. T., Penton-Voak, I. S. & Neumann, C. Why rate when you could compare? Using the “EloChoice” package to assess pairwise comparisons of perceived physical strength. *PLoS One* **13**, e0190393 (2018).
31. Herbrich, R., Minka, T., & Graepel, T. TrueSkill™: a Bayesian skill rating system, *Advances in neural information processing systems*, **19**, (2006).
32. Larkin, A., Gu, X., Chen, L. & Hystad, P. Predicting perceptions of the built environment using GIS, satellite and street view image approaches. *Landscape Urban Plan* **216**, 104257 (2021).
33. Savas, C. & Dervis, F. The Impact of Different Kernel Functions on the Performance of Scintillation Detection Based on Support Vector Machines. *Sensors* **19**, 5219 (2019).
34. Caragliu, A. & Del, Bo, C. F. Smart cities and the urban digital divide. *npj Urban Sustain.* **3**, 43 (2023).
35. Alkire, S., Kanagaratnam, U., & Suppa, N. The global Multidimensional Poverty Index (MPI) 2021. *Oxford Poverty and Human Development Initiative, OPHI MPI Methodological Note 51*, (2021).
36. Lenzi, C. & Perucca, G. Economic inequalities and discontent in European cities. *npj Urban Sustain.* **3**, 26 (2023).
37. Claudio, P. et al. Deep Learning and Earth Observation to Support the Sustainable Development Goals: Current approaches, open challenges, and future opportunities. *IEEE Geosci. Remote Sens. Magazine* **10**, (2022).
38. Georganos, S. et al. Is it all the same? Mapping and characterizing deprived urban areas using Worldview-3 superspectral imagery. A case study in Nairobi, Kenya. *Remote Sens.* **13**, 4986 (2021).
39. Alzubaidi, L. et al. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **8**, 1–74 (2021).
40. OpenAerialMap, The open collection of aerial imagery, <https://openaerialmap.org/> (2023).
41. Microsoft, microsoft/KenyaNigeriaBuildingFootprints, <https://github.com/microsoft/KenyaNigeriaBuildingFootprints> (2022).
42. Caggiano, H., Kocakuşak, D., Kumar, P. & Tier, M. O. U.S. cities' integration and evaluation of equity considerations into climate action plans. *npj Urban Sustain.* **3**, 50 (2023).
43. Morgan, W. T. W. *Nairobi: City and Region*. Oxford University Press, (1967).
44. K'akumu, O. A. & Olima, W. H. The dynamics and implications of residential segregation in Nairobi. *Habitat Int.* **31**, 87–99 (2007).
45. Kenya National Bureau of Statistics *2019 Kenya Population and Housing Census Reports* <https://www.knbs.or.ke/?s=census+2019> (2019).
46. Geofabrik, OpenStreetMap Contributors, <http://download.geofabrik.de> (2019).
47. Thomson, D. R. et al. Improving the accuracy of gridded population estimates in cities and slums to monitor SDG 11: Evidence from a simulation study in Namibia. *Land Use Policy* **123**, 106392 (2022).
48. Grippa, T. et al. Improving Urban Population Distribution Models with Very-High Resolution Satellite Information. *Data* **4**, 13 (2019).
49. Pizzol, B., Strambi, O., Giannotti, M., Arbex, R. O. & Alves, B. B. Activity behavior of residents of Paraisópolis slum: Analysis of multiday activity patterns using data collected with smartphones. *J. Choice Model.* **39**, 100287 (2021).
50. Simonyan, K., & Zisserman, A. Very deep convolutional networks for large-scale image recognition, (2014). Retrieved from <https://doi.org/10.48550/arXiv.1409.1556> (accessed 2 February 2021).
51. Wang, J. et al. EO+ Morphometrics: Understanding cities through urban morphology at large scale. *Landscape Urban Plan.* **233**, 104691 (2023).
52. Tibshirani, R. Regression shrinkage and selection via the lasso. *J. Royal Stat. Soc Series B: Stat. Methodol.* **58**, 267–288 (1996).
53. Boser, B. E., Guyon, I. M. & Vapnik, V. N. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, 144–152, (1992).
54. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
55. Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd international conference on knowledge discovery and data mining*, 785–794, (2016).

## Acknowledgements

Special thanks to all community participants from Kibera, Kariobangi, Korogocho, Mathare, Mukuru, Kahawa Soweto and Waruku. Notably to all the community leaders: Nicera Wanjiru from Kibera (<https://www.communitymappers.com>); Erick Chonelwa from Kariobangi; Anko Wanjiru from Korogocho; Elizabeth Kamene from Mathare; Daudi Anami from Mukuru; Susan Mwangi from Kahawa Soweto; Annastaciah Wanjiru from Waruku; for their commitment, trust, and generosity in sharing their indispensable ground knowledge. The research pertaining to these results received financial aid from the Belgian Federal Science Policy (BELSPO) according to the agreement of subsidy no. SR/11/217 (PARTIMAP). We would like to acknowledge Maxwell Owusu, from George Washington University, for supporting the participants' survey preparation, and Constant Cap and Dianne Wakonyo for organising the stakeholder workshops held in



Nairobi. In addition, we express our gratitude to the reviewers, whose expertise and thoughtful comments were invaluable in refining our arguments and strengthening the quality of our work.

### Author contributions

Conceptualisation, A.A., S.V., S.G., J.W., M.K. and E.W.; methodology, A.A., S.V., T.G., I.R. and P.M; writing—original draft preparation, A.A., S.V., and T.G.; writing—review and editing, A.A., S.V., T.G., I.R., S.G., J.W., M.K., P.M., M.S. and E.W; visualisation, A.A., S.V. and T.G.; project administration, A.A. and S.V; funding acquisition, A.A., S.V., S.G., J.W., M.K. and E.W. All authors have read and agreed to the published version of the manuscript. A.A., S.V. and T.G. contributed equally to the work and are considered “co-first authors”.

### Competing interest

The authors declare no competing interests.

### Ethical approval

The research prioritizes inclusion through the active involvement of local leaders in designing the citizen science protocol, ensuring local relevance and community engagement. Ethical considerations involve protecting participants' identities and obtaining informed consent for public sharing. Participant names remain confidential to prevent stigmatization or incrimination. Participants voluntarily agreed to be photographed and publicly share workshop photos, comparison votes data, and derived results.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42949-024-00156-x>.

**Correspondence** and requests for materials should be addressed to Angela Abascal.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024