

AI-POWERED SCHOOL MAPPING AND CONNECTIVITY STATUS PREDICTION USING EARTH OBSERVATION

Kelsey Doerksen^{1,2,3*}, Isabelle Tingzon^{1*}, Casper Fibæk³, Rochelle Schneider³, & Do-Hyung Kim¹

¹United Nations Children’s Fund (UNICEF)

²University of Oxford Department of Computer Science

³ European Space Agency (ESA) Φ-lab

{kdoerksen, itingzon, dokim}@unicef.org

ABSTRACT

Obtaining complete and accurate information on schools and their internet connectivity status is a critical first step to accelerating digital connectivity and driving progress towards SDG4: Quality Education. However, records of school locations and digital connectivity are often inaccurate, incomplete, or even completely non-existent in many developing countries. To address this challenge, we combine machine learning, satellite imagery, and geospatial information for school location mapping and internet connectivity status prediction in support of Giga, a joint initiative by UNICEF and ITU to connect every school to the internet by 2030. First, we benchmark the performance of Convolutional Neural Networks against Vision Transformer-based models for automated school mapping. Next, we investigate the application of machine learning and remote sensing data for school connectivity prediction. We evaluate our approach in five pilot countries, namely Bosnia and Herzegovina, Belize, Botswana, Guinea, and Rwanda.

1 INTRODUCTION

According to a joint report by UNICEF and the International Telecommunication Union (ITU), approximately two-thirds of the world’s school-age children do not have access to the internet (Diallo, 2020). Not only does the lack of internet connectivity limit children’s opportunity to access online learning resources, but it also prevents them from developing the digital skills needed to thrive in today’s modern economy. Moreover, this digital divide only exacerbates existing inequalities, causing children from the most disadvantaged households to fall behind even further. In response to these challenges, UNICEF and ITU jointly launched Giga, a global initiative to connect every school to the internet by 2030. To achieve this ambitious goal, government agencies and connectivity providers require accurate and complete school location and internet connectivity data to better estimate the costs of digitally connecting schools and to plan the strategic allocation of their financial resources. However, while governments generally have comprehensive records of schools within their national register, the corresponding geographical coordinates and internet connectivity status information are often incomplete, inaccurate, invalid, or completely non-existent, especially in developing nations.

Recent advances in artificial intelligence (AI) and Earth Observation (EO) have led to promising new opportunities to fill data gaps in education infrastructure. In support of the Giga initiative, we leverage machine learning and remote sensing data to accelerate school mapping and enable internet connectivity prediction. First, we benchmark the performance of fine-tuned Convolutional Neural Networks (CNNs) against that of Machine Learning (ML) classifiers trained on features extracted from pre-trained Vision Transformers (ViTs) for school classification. Second, we investigate the use of open-source satellite imagery, electricity transmission network information, and internet speed test data to predict the internet connectivity status of schools. We evaluate our models in five pilot countries: Bosnia and Herzegovina, Belize, Botswana, Guinea, and Rwanda.

*These authors contributed equally to this work.

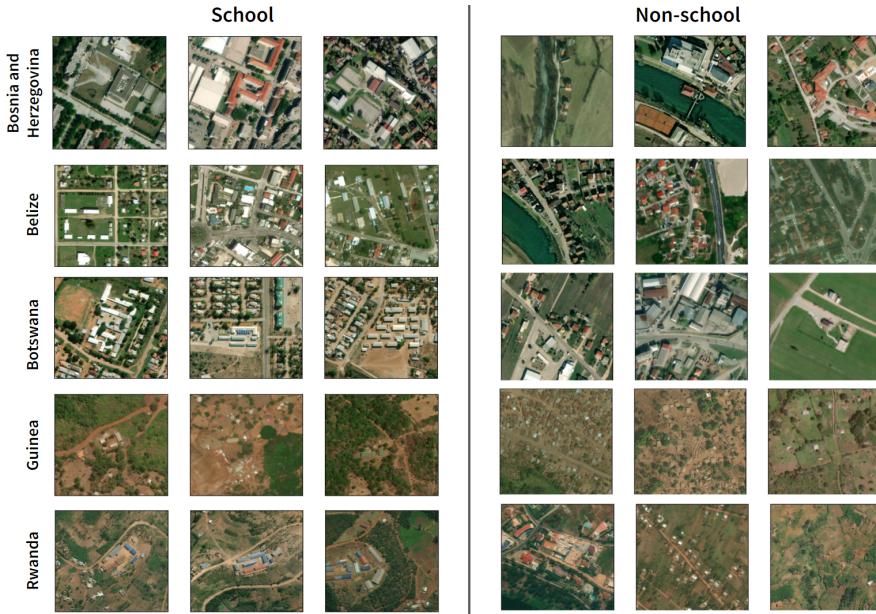


Figure 1: Examples of school and non-school satellite image tiles for each of the five countries of interest in this study. ©2024 DigitalGlobe NextView License.

2 DATA

2.1 SCHOOL MAPPING DATASET.

To create our school mapping dataset, we began with official school location data from government partners, made accessible through Project Connect¹, augmented with school point-of-interest (POI) locations from OpenStreetMap (OSM) and Overture Maps. For our negative examples, we queried the locations of non-school POIs such as hospitals, churches, warehouses, and offices from Overture Maps and OSM (see Appendix B for the full list of keywords). To ensure a high-quality dataset, we removed redundant entries, i.e. multiple points referring to the same building, and deduplicated names using the RapidFuzz library². We then leveraged Microsoft building footprints (Microsoft, 2023) and the Global Human Settlements Layer (Marcello et al., 2023) to filter out erroneous points located in uninhabited areas, e.g. forests, deserts, and bodies of water.

For each location in our dataset, we downloaded 300 x 300 m, 500 x 500 px RGB satellite images from Maxar with a spatial resolution of 0.6 m/px, centered on the corresponding GPS coordinate. To ensure the accuracy of the dataset, we manually reviewed the satellite images of known school locations and removed images where the school appeared to be either absent from the image or indistinguishable from surrounding buildings. We illustrate in Figure 1 examples of school and non-school image tiles for each country. For more details on the school mapping dataset creation and curation pipeline, see Appendix A.

2.2 CONNECTIVITY DATASET.

For the school connectivity dataset, we used a subset of the official school dataset from Project Connect containing school connectivity information. In line with the ethos of this work to provide open-source tools for good, we have leveraged openly available datasets to create our feature space, as summarized in Table 1. We derived a rich set of features from a combination of publicly available mid-resolution satellite data from Google Earth Engine (GEE), Electrical Power Grid information from World Bank Group (Arderne et al., 2020), and Speedtest data from Ookla (Ookla LLC., 2023).

¹<https://projectconnect.unicef.org/map/countries>

²<https://github.com/rapidfuzz/RapidFuzz>

Table 1: School connectivity feature space, including satellite-based feature (1-5), electricity grid and speed test features (6-7), and regional indicators (8).

Type	Source	Metrics
1. Land cover	MODIS MCD12Q1.006	% coverage per class, mode, variance
2. Population	Gridded Population of the World	mean, variance, max, min
3. Nightlight	VIIRS Nightlight	max, min, mean, variance
4. Human Settlement	GHS Built Up Characteristics	% coverage per class, mode, variance
5. Human Modification	Global Human Modification	mode, variance, mean, max, min
6. Electrical Transmission Grid	World Bank Group	distance from school to nearest transmission line
7. Fixed and Mobile Network Performance	Ookla Speedtest Data	mean download for mobile & fixed tests mean upload for mobile & fixed tests mean latency for mobile & fixed tests no. of mobile and fixed tests per tile distance of nearest tile to school
8. Region	geoBoundaries	no. of unique mobile & fixed devices one-hot encoded location of ADM2 boundary

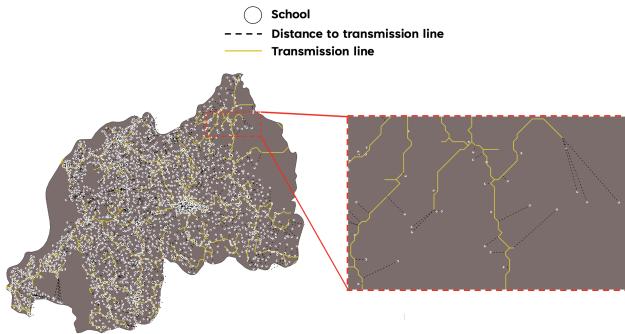


Figure 2: Examples of distance from school to nearest transmission line for Rwanda.

Satellite-based features. Open-source, satellite-based measurements are traditionally under-utilized in low-income contexts, where such data can provide great benefits (Yu, 2014; Haack & Ryerson, 2016). Taking each location with connectivity information in our dataset as the center point, we extract a 1,000 m radius extent of high-resolution satellite data from GEE including MODIS landcover (Sulla-Menashe & Friedl), VIIRS Nightlight (Elvidge et al.), Global Human Modification (Kennedy et al., 2019), Gridded Population of the World (CIESIN, 2018), and Global Human Settlement Layer (M. & Panagiotis, 2023) data products using the airPy³ data processing package.

Electrical transmission network & internet speed test features. Using the derived map of global electricity transmission and distribution lines dataset⁴, we calculate the distance from each school to the nearest transmission line, as shown in Figure 2. To incorporate speed test data, features are extracted for the nearest tile to each school for the latest data from October 2023.

Regional indicators. Regional indicators were incorporated by including the administrative level 2 location per school provided by geoBoundaries⁵ per country. For each school location, the one-hot encoded administrative boundaries were added as features with the region the school was within equal to one. For edge cases where the schools were slightly outside of a country’s boundary (occurring due to the label noise in the UNICEF dataset), the administrative boundary that was closest to the school was taken as the school’s region.

3 METHODS

3.1 SCHOOL MAPPING MODEL

We model school mapping as a binary image classification task wherein we classify each satellite image tile into school and non-school categories (Yi et al., 2019; Maduako et al., 2022). For all

³<https://github.com/kelsdoerksen/airPy>

⁴<https://energydata.info/dataset/derived-map-global-electricity-transmission-and-distribution-lines>

⁵<https://www.geoboundaries.org/countryDownloads.html>

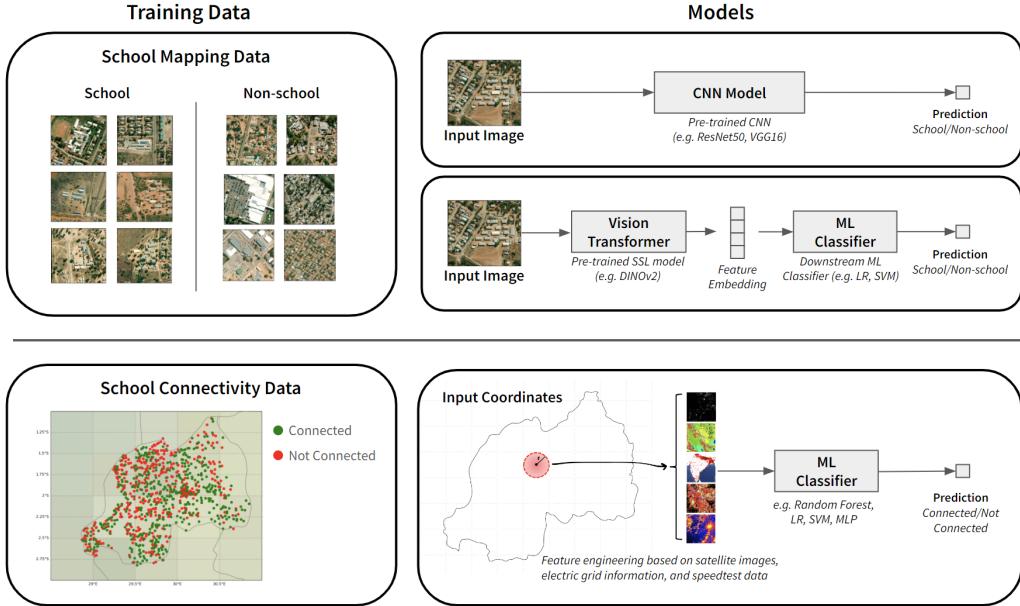


Figure 3: An overview of the training data and model types for AI-assisted school mapping (top) and connectivity prediction (bottom). ©2024 DigitalGlobe NextView License.

experiments, we used a consistent train (80%) and test (20%) split within each country-level dataset. Table 3 in the Appendix shows the class distribution for the train and test sets across the 5 countries.

Baseline CNNs. We selected three variations of ConvNext, i.e. small (S), base (B), and large (L) (Liu et al., 2022), VGG-16 (Simonyan & Zisserman, 2014), Xception (Chollet, 2017), and ResNet-50 (He et al., 2016) as our baseline CNN architectures based on their success in past works (Maduako et al., 2022; Yazdani et al., 2018; Yi et al., 2019). All models were pre-trained on the ImageNet dataset (Deng et al., 2009) and fine-tuned on each country-level dataset. Further details on CNN model configuration and data augmentation can be found in Appendix D.1.

DINOv2 ViT + ML classifiers. We used DINOv2 ViTs (Oquab et al., 2023) pre-trained on the LVD-142M dataset to generate dense feature embeddings, which were then used as input to shallow classifiers, including Logistic Regression (LR) and Support Vector Machines (SVM), for the downstream task of school classification. To determine the optimal model configuration, we implemented hyperparameter tuning on the training set using 5-fold cross-validation Grid Search, as described in further detail in Appendix D.1.

3.2 CONNECTIVITY PREDICTION

Similarly, we model school connectivity as a binary classification task wherein we classify each school, represented by an n-featured vector (where n ranges from 74-91 dependent on the number of one-hot encoded administrative boundaries in each country), based on its internet connectivity status. For all experiments, we used a consistent train (70%) and test (30%) split within each country-level dataset and 5-fold cross-validation. For model development, we feed the tabular features to ML classifiers including LR, SVM, Random Forest (RF), Gradient Boosting (GB), and Multi-Layer Perception (MLP), all trained and tested on a per-country level. Min-max scaling on the training and testing sets is done for the MLP, SVM and LR models to improve performance as implemented in *scikit-learn*. Features with a correlation score of greater than 0.9 are removed from the dataset. We use a 5-fold cross-validation Grid Search to determine the best model parameters within a given search space (see Appendix D.2).

Table 2: A comparison of model performance scores for (a) school mapping, given by the per-country binary F1 scores (%) of the CNN models and DINoV2 ViT-based models and (b) connectivity prediction, given by per-country F1 scores (%) of ML classifiers for Bosnia and Herzegovina (**BIH**), Belize (**BLZ**), Botswana (**BWA**), Guinea (**GIN**), and Rwanda (**RWA**).

(a) School Mapping F1 scores						(b) Connectivity Prediction F1 scores						
	BIH	BLZ	BWA	GIN	RWA		BIH	BLZ	BWA	GIN	RWA	
CNN	ConvNext-S	0.80	0.80	0.95	0.83	0.96	RF	0.82	0.92	0.73	0.74	0.72
	ConvNext-B	0.81	0.82	0.96	0.80	0.95	SVM	0.83	0.89	0.72	0.69	0.69
	ConvNext-L	0.79	0.83	0.95	0.82	0.96	LR	0.83	0.88	0.71	0.66	0.70
	ResNet50	0.72	0.15	0.94	0.76	0.94	GB	0.82	0.90	0.73	0.70	0.69
	VGG16	0.75	0.73	0.95	0.80	0.95	MLP	0.83	0.86	0.68	0.68	0.71
	Xception	0.63	0.07	0.95	0.73	0.94						
DINoV2	ViT-S/14-LR	0.61	0.56	0.91	0.63	0.94						
	ViT-S/14-SVM	0.65	0.65	0.91	0.67	0.94						
	ViT-B/14-LR	0.68	0.75	0.93	0.72	0.95						
	ViT-B/14-SVM	0.71	0.74	0.93	0.70	0.94						
	ViT-L/14-LR	0.64	0.69	0.92	0.69	0.93						
	ViT-L/14-SVM	0.69	0.71	0.94	0.70	0.94						

4 RESULTS & DISCUSSION

For school mapping, we find that the ConvNext model architecture consistently outperforms all other CNN and ViT-based models, with the highest F1 scores ranging from 0.81 to 0.96 across the five pilot countries, as detailed in Table 2. These results are generally consistent with past works benchmarking CNNs against ViT-based foundation models for domain-specific applications (Chen et al., 2023; Knott et al., 2023). Meanwhile, for school connectivity, we find that the RF classifier architecture yields the highest F1-score for Botswana, Guinea, and Rwanda, the GB classifier the highest F1 score for Belize, and the SVM, MLP, and LR classifiers all yield similar performances for Bosnia and Herzegovina as shown in Table 2. Consistently, all models struggle in Guinea, which is the only country within this study with a larger class representation of unconnected schools over connected schools (see Table 4 of the Appendix). Future work will include leveraging remote sensing-based encoders, e.g. SatCLIP, MOSAIKS (Klemmer et al., 2023; Rolf et al., 2021).

5 CONCLUSION

In alignment with the UN Sustainable Development Goals (SDGs), including SDG 4, which aims to “ensure inclusive and equitable quality education and promote lifelong learning opportunities for all”, we present in this study an end-to-end pipeline for accelerating school mapping and predicting school connectivity status using a combination of ML and Earth Observation data. Our results indicate that CNNs generally outperform ViT-based foundation models for satellite image-based school classification, achieving F1 scores between 0.80 and 0.96 across our pilot countries. We also demonstrate the viability of using tabular features engineered from satellite images, electric grid information, and speed test data to predict the connectivity status of schools, with the best F1 scores ranging from 0.64 to 0.89 across five countries. In line with Giga’s ethos to provide open-source school connectivity tools, we publicly release our code for AI-enabled school mapping⁶.

Future Work. One critical component in the adoption of AI-enabled school mapping and connectivity prediction solutions by governments and UNICEF country offices is the integration of human-in-the-loop validation protocols to ensure that strategic plans for delivering internet connectivity in schools are based on reliable ML predictions. To this end, we are currently developing interactive cloud-based platforms for post-classification validation via (1) remote validation by domain experts and (2) on-the-ground validation, as was done in Tingzon et al. (2020). In this work, we underscore the importance of field validation in the responsible applications of AI for humanitarian action.

⁶<https://github.com/unicef/giga-global-school-mapping>

6 ACKNOWLEDGEMENTS

Kelsey Doerksen's contributions to this paper were developed under a collaboration with the European Space Agency *Phi-lab* to support UNICEF and ITU's *Giga* initiative.

REFERENCES

- C. Arderne, C. Zorn, and C. et al. Nicolas. Predictive mapping of the global power system using open data. *Nature Scientific Data*, 7(19), 2020. doi: <https://doi.org/10.1038/s41597-019-0347-4>.
- Feng Chen, Mario Valerio Giuffrida, and Sotirios A Tsaftaris. Adapting vision foundation models for plant phenotyping. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 604–613, 2023.
- François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1251–1258, 2017.
- CIESIN. Gridded population of the world, version 4 (gpwv4): Population density. *New York: NASA Socioeconomic Data and Applications Center (SEDAC)*, 2018. doi: <https://doi.org/10.7927/H49C6VHW>.
- Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. IEEE, 2009.
- Georgina Diallo. Two thirds of the world's school-age children have no internet access at home, new UNICEF-ITU report says — unicef.org. <https://www.unicef.org/press-releases/two-thirds-worlds-school-age-children-have-no-internet-access-home-new-unicef-itu>, 2020. [Accessed 31-01-2024].
- Christopher D. Elvidge, Kimgerly Baugh, Mikhail Zhizhin, Feng Chi Hsu, and Tolottama Ghosh. Viirs night-time lights. *International Journal of Remote Sensing*, 38(21):5860–5879. doi: <https://doi.org/10.1080/01431161.2017.1342050>.
- Barry Haack and Robert Ryerson. Improving remote sensing research and education in developing countries: Approaches and recommendations. *International Journal of Applied Earth Observation and Geoinformation*, 45:77–83, 2016.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- C.M. Kennedy, J.R. Oakleaf, D.M. Theobald, S. Baurch-Murdo, and J. Kiesecker. Managing the middle: A shift in conservation priorities based on the global human modification gradient. *Global Change Biology*, 00:1:1–16, 2019. doi: [doi:10.1111/gcb.14549](https://doi.org/10.1111/gcb.14549).
- Konstantin Klemmer, Esther Rolf, Caleb Robinson, Lester Mackey, and Marc Rußwurm. Satclip: Global, general-purpose location embeddings with satellite imagery. *arXiv preprint arXiv:2311.17179*, 2023.
- Manuel Knott, Fernando Perez-Cruz, and Thijs Defraeye. Facilitated machine learning for image-based fruit quality assessment. *Journal of Food Engineering*, 345:111401, 2023.
- Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, and Saining Xie. A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11976–11986, 2022.
- Pesaresi M. and P. Panagiotis. GHS-BUILT-C R2023A - GHS Settlement Characteristics, derived from Sentinel2 composite (2018) and other GHS R2023A data. *European Commission, Joint Research Centre (JRC)*, PID: <http://data.europa.eu/89h/3c60ddf6-0586-4190-854b-f6aa0edc2a30>, 2023. doi: [10.2905/3c60ddf6-0586-4190-854b-f6aa0edc2a30](https://doi.org/10.2905/3c60ddf6-0586-4190-854b-f6aa0edc2a30).

- Iyke Maduako, Zhuangfang Yi, Naroa Zurutuza, Shilpa Arora, Christopher Fabian, and Do-Hyung Kim. Automated school location mapping at scale from satellite imagery based on deep learning. *Remote Sensing*, 14(4):897, 2022.
- Schiavina Marcello, Melchiorri Michele, Pesaresi Martino, Politis Panagiotis, Carneiro Freire Sergio Manuel, Maffenini Luca, Florio Pietro, Ehrlich Daniele, Goch Katarzyna, Carioli Alessandra, Uhl Johannes, Tommasi Pierpaolo, and Kemper Thomas. GHSL Data Package 2023. (KJ-03-23-103-EN-N (online)), 2023. doi: 10.2760/098587(online).
- Microsoft. Microsoft Global ML Building Footprints. <https://github.com/microsoft/GlobalMLBuildingFootprints>, 2023. [Accessed 31-01-2024].
- Ookla LLC. Global fixed broadband and mobile network maps. <https://www.ookla.com/ookla-for-good/open-data>, 2023. [Accessed 16-01-2024].
- Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.
- Esther Rolf, Jonathan Proctor, Tamara Carleton, Ian Bolliger, Vaishaal Shankar, Miyabi Ishihara, Benjamin Recht, and Solomon Hsiang. A generalizable and accessible approach to machine learning with global satellite imagery. *Nature communications*, 12(1):4392, 2021.
- Ribana Roscher, Marc Rußwurm, Caroline Gevaert, Michael Kampffmeyer, Jefersson A dos Santos, Maria Vakalopoulou, Ronny Hänsch, Stine Hansen, Keiller Nogueira, Jonathan Prexl, et al. Data-centric machine learning for geospatial remote sensing data. *arXiv preprint arXiv:2312.05327*, 2023.
- Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- Damien Sulla-Menashe and Mark A Friedl. Mcd12q1 modis/terra+aqua land cover type yearly l3 global 500m sin grid v006. *NASA EOSDIS Land Processes DAAC*. doi: <https://doi.org/10.5067/MODIS/MCD12Q1.006>.
- Isabelle Tingzon, Niccolo Dejito, Ren Avell Flores, Rodolfo De Guzman, Liliana Carvajal, Katherine Zapata Erazo, Ivan Enrique Contreras Cala, Jeffrey Villaveces, Daniela Rubio, and Rayid Ghani. Mapping new informal settlements using machine learning and time series satellite images: An application in the venezuelan migration crisis. In *2020 IEEE / ITU International Conference on Artificial Intelligence for Good (AI4G)*, pp. 198–203, 2020. doi: 10.1109/AI4G50087.2020.9311041.
- Mehrdad Yazdani, Mai H Nguyen, Jessica Block, Daniel Crawl, Naroa Zurutuza, Dohyung Kim, Gordon Hanson, and Ilkay Altintas. Scalable detection of rural schools in Africa using convolutional neural networks and satellite imagery. In *2018 IEEE/ACM International Conference on Utility and Cloud Computing Companion (UCC Companion)*, pp. 142–147. IEEE, 2018.
- Zhuangfang Yi, Naroa Zurutuza, Drew Bollinger, Manuel Garcia-Herranz, and Dohyung Kim. Towards equitable access to information and opportunity for all: mapping schools with high-resolution satellite imagery and machine learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 60–66, 2019.
- L. Yu. Meta-discoveries from a synthesis of satellite-based land-cover mapping research. *International Journal of Remote Sensing*, 35(4573), 2014.

APPENDIX

A SCHOOL MAPPING DATASET CREATION

Because data quality substantially impacts model performance, we adopt a data-centric approach that aims to systematically maximize the quality and richness of our school mapping dataset. We describe below the steps in our school mapping data creation and curation pipeline, with a particular focus on increasing the diversity and completeness, accuracy, consistency, unbiasedness, and relevance of the dataset (Roscher et al., 2023).

Data sources. To create a diverse and representative school mapping dataset, we collated school location information across multiple data sources. We began with existing school data from government partners across the five countries which contain the names and GPS coordinates of the schools. We then augmented this data with point-of-interest (POI) information of schools from Overture Maps⁷ and OpenStreetMap (OSM)⁸, retrieved using DuckDB and the Overpass API, respectively. In line with the interests of partner government stakeholders, we focus primarily on locating primary and secondary schools; we thus excluded schools containing keywords related to early childhood education (e.g. “preschool”, “nursery”, “kindergarten”), tertiary education (e.g. “university”, “college”), sports academies (e.g. “swimming”, “dance”, “taekwondo”), and other types of educational centers (e.g. “driving”, “aviation”, “cosmetics”, “business”,).

Deduplication. Combining school location data across multiple sources can lead to duplicate points, i.e. points that are within very close proximity of another school location and thus likely pertain to the same school building. We identified these redundant entries by creating 25 m buffers around each point and grouping together points that shared overlapping buffer zones; within each group of points, we retained one and discarded the rest. We also identified data points with similar names by applying fuzzy string matching for pairs of schools that are within a distance of 300 m of each other. Using the RapidFuzz library⁹, we deduplicated school names by identifying strings that match with a normalized Levenshtein similarity of at least 85% based on the optimal alignment of the two strings. For example, if two school locations with the names “Sashe” and “Sashe Primary School” match with a score of 86.2%, we would keep one of the points and discard the other.

Removal of uninhabited locations. To remove erroneous data points in uninhabited locations we leveraged open-source settlement data including the Microsoft building footprints dataset (Microsoft, 2023), retrieved per country and rasterized to a 10 m resolution GeoTIFF, and the Global Human Settlements Layer (GHSL) BUILT-C product (Marcello et al., 2023). We created 150 m buffers around each point and calculated the sum of the pixels within the buffer area using the Microsoft building footprints and GHSL rasters separately. We retained only the school locations where the pixel sum from the rasterized Microsoft building footprints and GHSL were both nonzero.

Negative sample retrieval. To generate our set of negative examples, we queried the locations of non-school POIs from Overture Maps and OSM (see Section B of the Appendix). For countries where the number of non-school locations n is less than the number of known school locations s , we augmented the non-school dataset by randomly sampling $(s - n)$ points from a set of points in inhabited locations, based on the Microsoft buildings dataset, spaced 300 m apart within the country boundary. We applied the same deduplication and filtering workflow for non-school locations as with school locations but with the additional step of removing non-school locations within 300 m of known school locations. This step is done to ensure that no school buildings appear in the periphery of satellite images labeled as non-school.

Manual review of satellite images. To improve the correctness and precision of the data, we manually reviewed the satellite images of known school locations and removed images where the school appeared to be either absent from the image or indistinguishable from surrounding buildings. Guided by auxiliary information from the Google Satellite Hybrid base map, we also manually repositioned the GPS coordinates of schools that were located more than 300 m away from the actual school building as a way to resolve location-related discrepancies.

⁷<https://github.com/OvertureMaps/data>

⁸<https://www.openstreetmap.org/>

⁹<https://github.com/rapidfuzz/RapidFuzz>

B NON-SCHOOL KEYWORDS

We used the following groups of keys and values to query OSM and Overture Maps for non-school POIs, based on OSM’s tagging system¹⁰: (a) **amenities**, including fast food, food court, cafe, restaurant, pub, bar, bank, clinic, doctors, hospital, pharmacy, dentist, veterinary, arts center, cinema, casino, community center, conference center, events venue, fuel, exhibition center, planetarium, theatre, nightclub, courthouse, fire station, police, post office, prison, townhall, crematorium, funeral hall, internet cafe, marketplace, place of mourning, and place of worship; (b) **buildings**, including commercial, industrial, office, retail, warehouse, church, cathedral, chapel, mosque, temple, synagogue, shrine, supermarket, fire station, police, prison, hospital, museum, and military; (c) **craft**, including agricultural engines, atelier, bakery, blacksmith, brewery, cabinet maker, carpenter, electronics repair, distillery, and oil mill; (d) **healthcare**, including audiologist, birthing center, chiropractor, dentist, midwife, occupational therapist, optometrist, physiotherapist, psychologist, speech therapist, blood bank, blood donation, and vaccination center; (e) **historic**, including church, cathedral, castle, mosque, and tower; (f) **land use**, including commercial, retail, industrial, warehouse, cemetery, and religious; (g) **shopping**, including bakery, beverages, brewing supplies, butcher, cheese, chocolate, coffee, confectionery, convenience, farm, food, general, department store, kiosk, mall, supermarket, wholesale, beauty, fabric, fashion, electronics, garden centre; and (h) **tourism**, including guest house, hostel, hotel, motel, museum, chalet, apartment, zoo.

C CLASS DISTRIBUTIONS

Tables 3 and 4 show the class distributions for school mapping and connectivity prediction, respectively. We note that the usage of supervised learning for the connectivity prediction task, namely, requiring connectivity labels for school samples, results in a lower number of school samples for the connectivity prediction task than the school mapping task.

Table 3: The class distribution across the training and test sets of Bosnia and Herzegovina (**BIH**), Belize (**BLZ**), Botswana (**BWA**), Guinea (**GIN**), and Rwanda (**RWA**).

	Training Set (80%)			Test Set (20%)			Total
	School	Non-school	Total	School	Non-school	Total	
BIH	576	3,283	3,859	142	821	963	4,822
BLZ	178	674	852	55	156	211	1,063
BWA	757	1,329	2,086	182	338	520	2,606
GIN	787	1,336	2,123	202	327	529	2,652
RWA	2,262	2,667	4,929	562	668	1,230	6,159

Table 4: The class distribution across the training and test sets of Bosnia and Herzegovina (**BIH**), Belize (**BLZ**), Botswana (**BWA**), Guinea (**GIN**), and Rwanda (**RWA**).

	Training Set (70%)			Test Set (30%)			Total
	Connected	Not Connected	Total	Connected	Not Connected	Total	
BIH	651	284	935	278	123	401	1336
BLZ	168	52	220	75	20	95	315
BWA	327	307	634	149	124	273	907
GIN	286	373	659	113	170	283	942
RWA	1337	1011	2348	551	456	1007	3355

D MODEL CONFIGURATIONS

D.1 SCHOOL MAPPING MODEL

Baseline CNNs. All CNN models used in this study were pre-trained on the ImageNet dataset (Deng et al., 2009) and fine-tuned on the designated training sets of each country using cross-entropy loss with label smoothing set to 0.1 for regularization. We resized all images to 224 x 224 px and

¹⁰https://wiki.openstreetmap.org/wiki/Map_features

implemented data augmentation in the form of vertical and horizontal flips for the training set. Across all models, we used an Adam optimizer, set the batch size to 32, and used an initial learning rate of $1e^{-5}$ that decays by a factor of 0.1 after every 7 epochs of no improvement. For the ResNet models, we also applied a dropout layer with a probability of 0.5 before the fully connected layer to prevent overfitting. We set the maximum number of epochs to 60, with early stopping if the learning rate dipped below $1e^{-10}$.

DINOv2 ViT + ML classifiers. We used the DINOv2 small (ViT-S), base (ViT-B), and large (ViT-L) architectures with a 14 x 14 patch size to extract dense feature embeddings of size 385, 768, and 1024, respectively from images resized to 560 x 560 px. These feature embeddings are then used as input to the downstream ML classifiers LR and SVM. We implemented hyperparameter tuning through a 5-fold grid search cross-validation scheme. For LR, our search space included the norm of the penalty (L1 and L2) and the regularization parameter C , (0.001, 0.01, 0.1, 1.0, and 10). For SVM, our search space included the kernel type (linear, polynomial, radial basis function, and sigmoid), the kernel coefficient gamma (1, 0.1, 0.01, 0.001, and 0.0001), and the regularization parameter C (0.001, 0.01, 0.1, 1.0, and 10). We also experimented with different scaling techniques including standard scaling, min-max scaling, and robust scaling as implemented in *scikit-learn*.

D.2 CONNECTIVITY MODEL

Each model leverages hyperparameter tuning through a 5-fold grid search cross-validation scheme, with search spaces defined in Table 5.

Table 5: School connectivity model parameter search spaces for Random Forest (**RF**) Support Vector Machines (**SVM**), Logistic Regression (**LR**), Gradient Boosting Classifier (**GB**), and Multilayer Perceptron (**MLP**).

Model	Parameters
RF	<code>max depth</code> : 80, 90, 100 <code>max features</code> : 2, 3, 4 <code>min samples leaf</code> : 3, 4, 5 <code>min samples split</code> : 4, 6, 8 <code>n estimators</code> : 100, 200, 300, 500
SVM	<code>C</code> : 0.001, 0.01, 0.1, 1.0, 10.0 <code>kernel</code> : linear, poly, rbf, sigmoid <code>degree</code> : 1, 2, 3, 4 <code>gamma</code> : scale, auto
LR	<code>penalty</code> : l2, None <code>C</code> : 0.01, 0.1, 1.0
GB	<code>loss</code> : log loss, exponential <code>learning rate</code> : 0.05, 0.1, 0.5, 1 <code>n estimators</code> : 100, 200, 300 <code>criterion</code> : freidman-mse, squared-error <code>min samples split</code> : 2, 4, 6 <code>min samples leaf</code> : 1, 3, 5 <code>max features</code> : sqrt, log2, None
MLP	<code>hidden layer sizes</code> : (100,), (150,), (200,) <code>activation</code> : logistic, tanh, relu <code>solver</code> : lbfgs, sgd, adam <code>alpha</code> : 0.0001, 0.005, 0.001 <code>learning rate</code> : constant, invscaling, adaptive