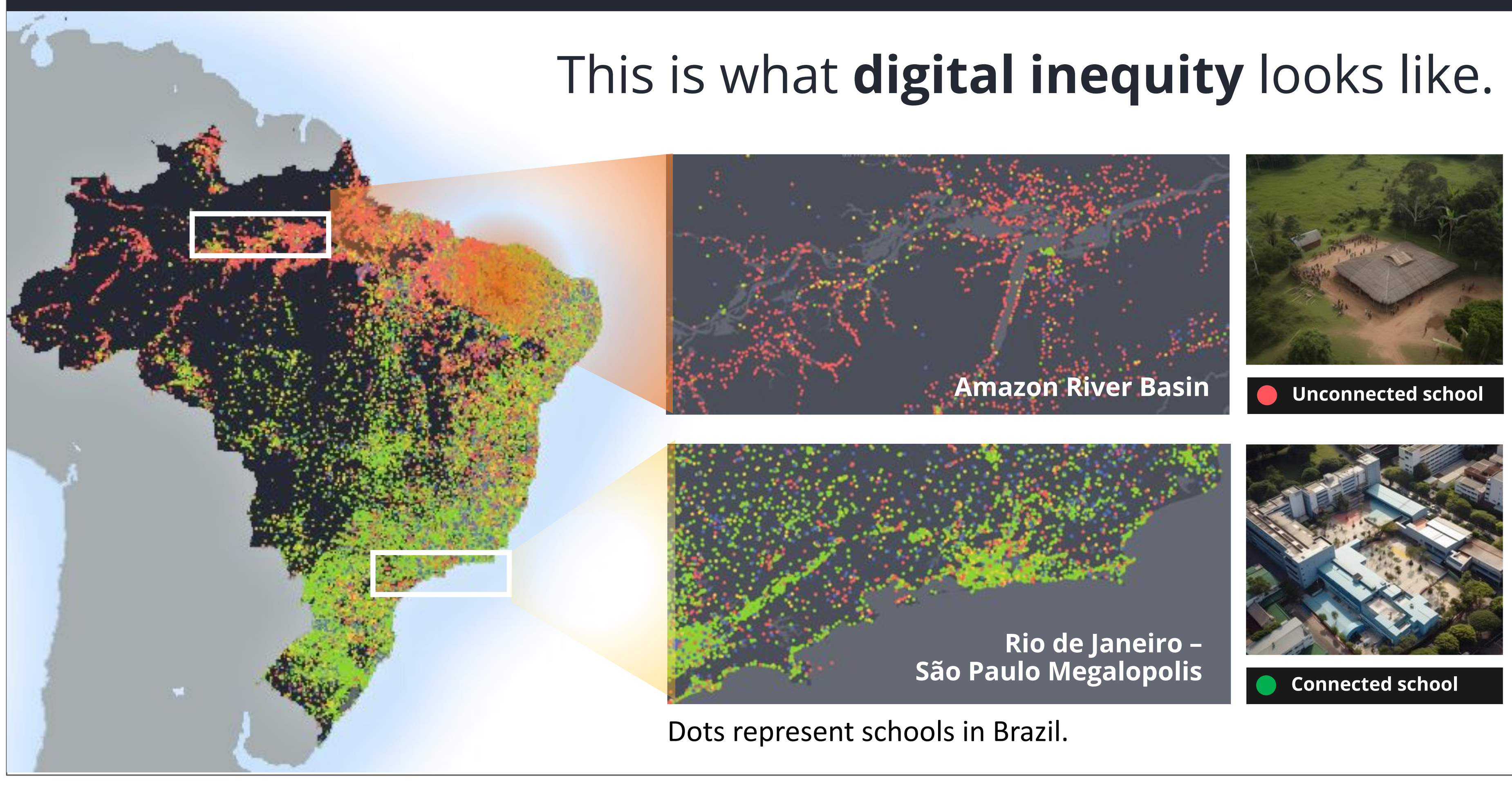
AI-powered School Mapping and Connectivity Status Prediction using Earth Observation

Kelsey Doerksen[1,2,3]*, Isabelle Tingzon[1]*, Casper Fibæk[3], Rochelle Schneider[3], Do-Hyung Kim[1]

[1]UNICEF, [2]University of Oxford Department of Computer Science, [3]European Space Agency Φ-lab

*equal contribution

Introduction



Dots represent schools in Brazil.

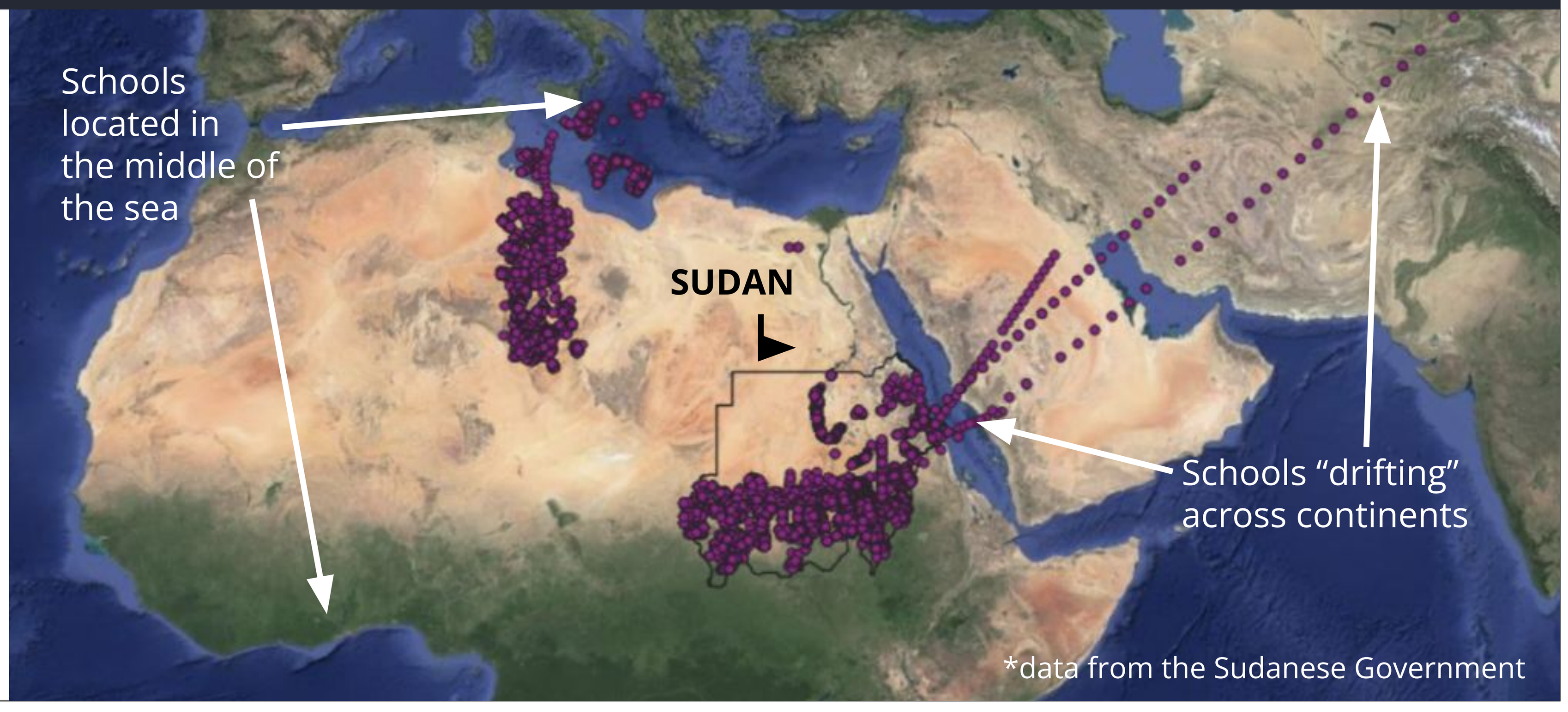
More than 500M students worldwide have no access to the internet. Not only does this limit learning opportunities but it also prevents students from developing the digital skills needed to compete in today's modern economy.

Giga's mission is to **connect all schools to the Internet by 2030** and every young person to **information, opportunity, and choice.**

Obtaining accurate information about school locations and their connectivity status is a critical first step to accelerating digital connectivity.

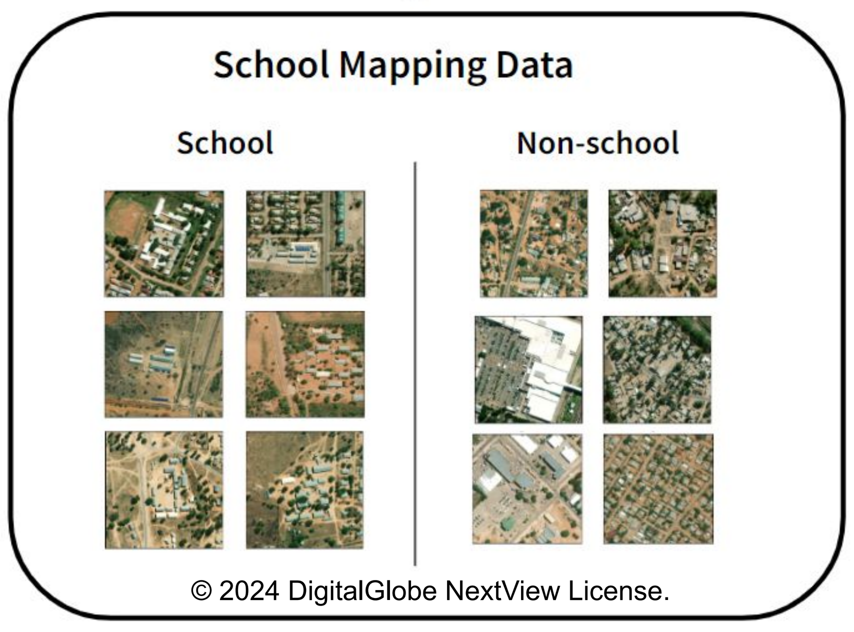
Challenges with School Location & Connectivity Data

- **Non-existent** "We know there are 20,000 schools in the country, but we don't know their exact coordinates."
- **Incomplete.** "We know the locations of 80% of the schools, and we need to find the missing 20%."
- **Inaccurate.** e.g. Recorded GPS coordinates are > 1 km away from the actual school building.
- **Invalid.** e.g. GPS coordinates point to non-school locations or uninhabited areas (e.g. forests, deserts).

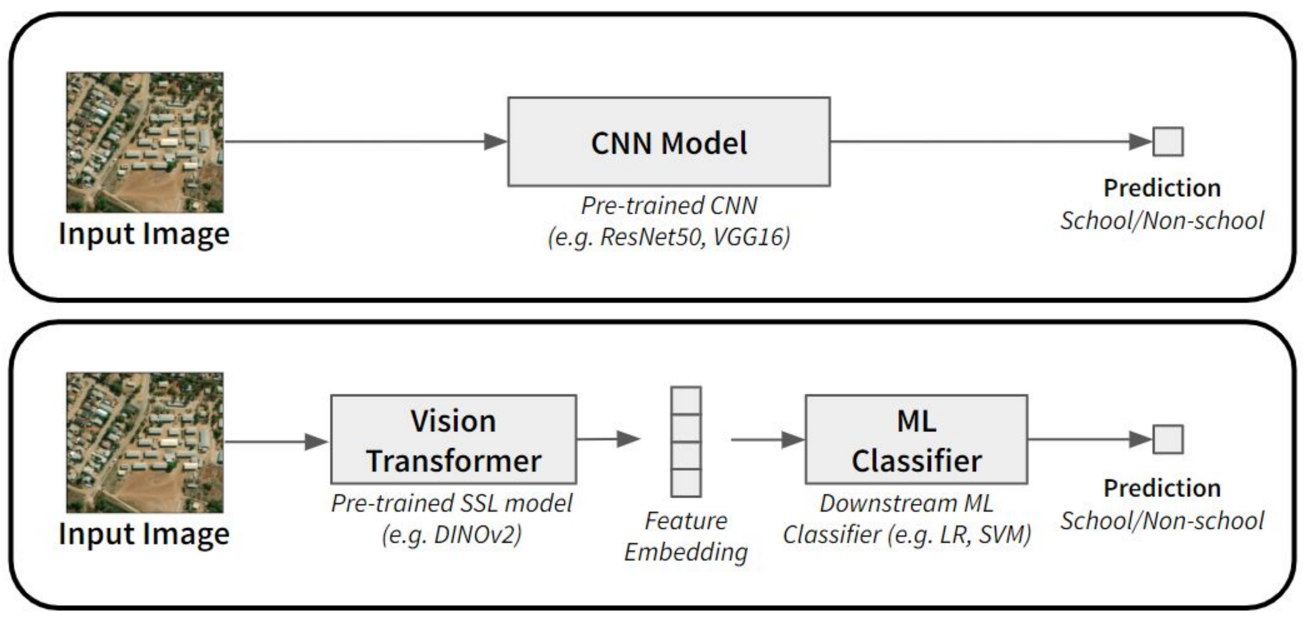


*data from the Sudanese Government

Methodology

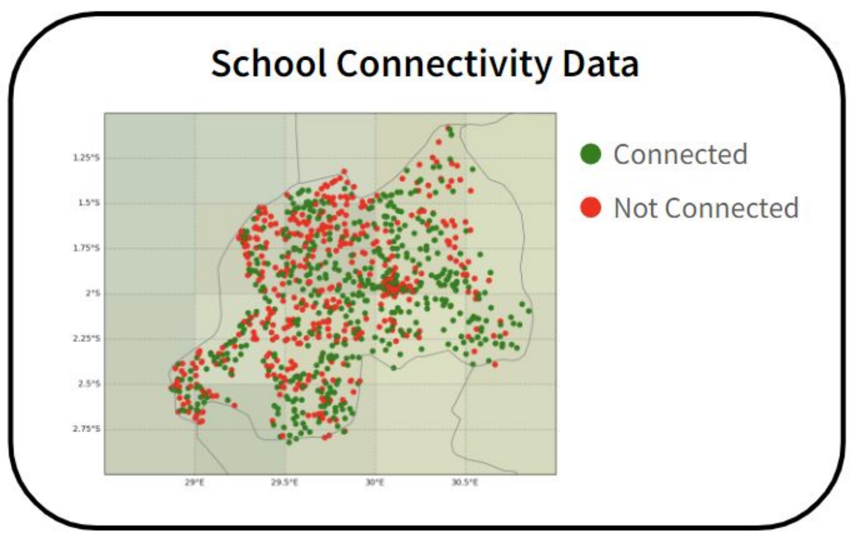


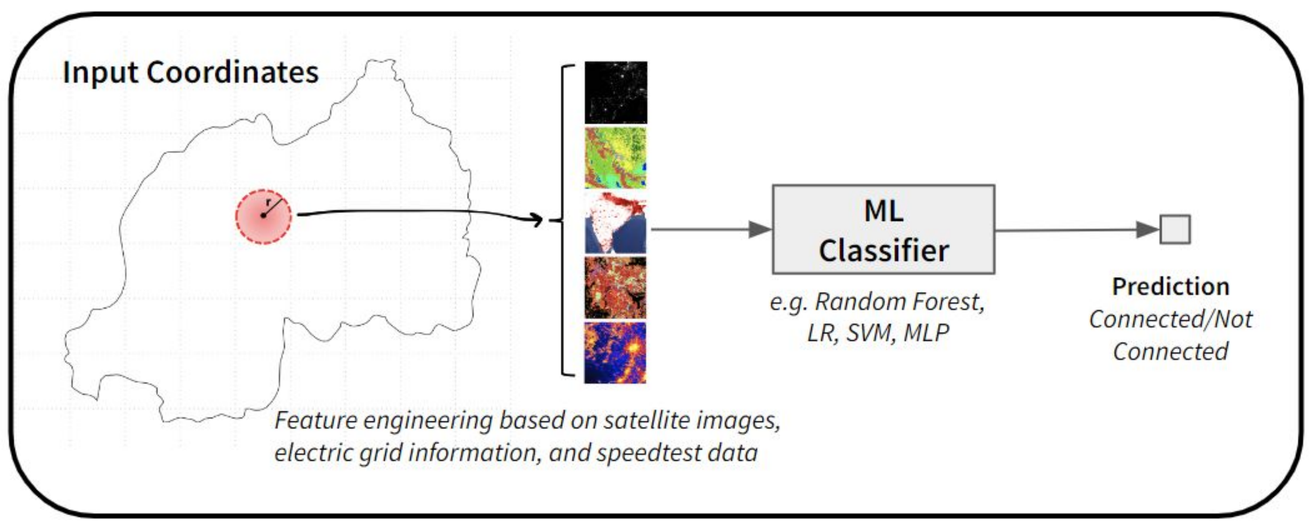
© 2024 DigitalGlobe NextView License.



School Mapping Model

We experiment with (a) fine-tuning pre-trained CNNS and (b) using pre-trained ViTs with downstream ML classifiers to classify 300 x 300 m Maxar satellite image tiles into either school or non-school.





Connectivity Prediction Model

We investigated the use of satellite imagery, electricity transmission network information, and internet speed test data to predict the internet connectivity status of schools with a suite of ML classifiers.

Results

School Mapping. We find that CNNs, specifically ConvNext models, generally outperform ViT-based models, with the best F1-scores ranging from 0.81 to 0.96 across the 5 pilot countries.

Connectivity Prediction. We show that we can predict the connectivity status of schools with the best F1 scores ranging from 0.72 to 0.90 across 5 countries.

(a) School Mapping F1 scores

		BIH	BLZ	BWA	GIN	RWA
CNN	ConvNext-S	0.80	0.80	0.95	**0.83**	**0.96**
	ConvNext-B	**0.81**	0.82	**0.96**	0.80	0.95
	ConvNext-L	0.79	**0.83**	0.95	0.82	0.96
	ResNet50	0.72	0.15	0.94	0.76	0.94
	VGG16	0.75	0.73	0.95	0.80	0.95
	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

(b) Connectivity Prediction F1 scores

	BIH	BLZ	BWA	GIN	RWA
RF	0.82	0.92	**0.73**	**0.74**	**0.72**
SVM	**0.83**	0.89	0.72	0.69	0.69
LR	**0.83**	0.88	0.71	0.66	0.70
GB	0.82	**0.90**	**0.73**	0.70	0.69
MLP	**0.83**	0.86	0.68	0.68	0.71

*Per-country macro-averaged **F1-scores** of ML classifiers for Bosnia and Herzegovina (**BIH**), Belize (**BLZ**), Botswana (**BWA**), Guinea (**GIN**), and Rwanda (**RWA**)