Dataset Title

DeepWealth: A Generalizable Deep Learning Framework to use Satellite Images for Poverty Estimation

Abstract

The integration of Deep Learning (DL) and Earth Observation (EO) data is becoming increasingly prevalent in the estimation of socioeconomic conditions. This combination has provided promising estimations at regional and global levels. Socioeconomic conditions mapping is relevant for many reasons, from determining earth system and climate impacts to environmental and urban planning. It provides a good understanding of socioeconomic changes across time and space. The goal of this paper is to report on an end-to-end framework (DeepWealth) to estimate poverty using EO data and DL. DeepWealth aligns with the Sustainable Development Goal (SDG1) of ending poverty (Fig. 1). We use a multidisciplinary approach incorporating satellite imagery, socio-economic surveys, and DL models to estimate poverty. We demonstrate the effectiveness and generalisability of DeepWealth by training it in Africa and deploying it in three use cases across the world to test its flexibility. Our results show that DeepWealth provides accurate and stable poverty estimates. DeepWealth empowers computer-literate users, including those skilled in Python and R, to manage and visualize poverty-related information. This open-source framework follows FAIR principles, providing data, source code, metadata, and training checkpoints with its source code made available on Zenodo and GitHub. In this manner, we aim to provide a DL framework that is reproducible and replicable.

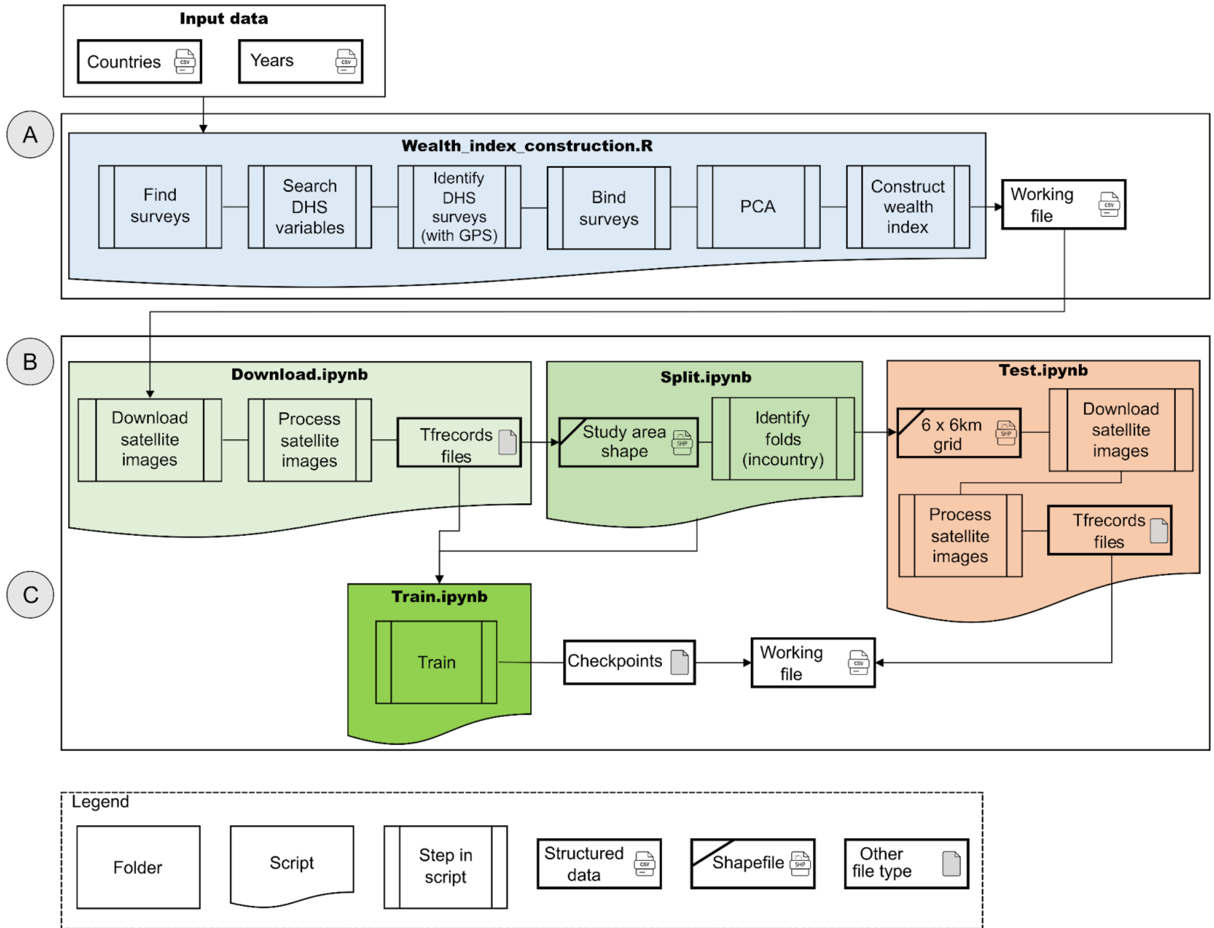

Figure 1: Overview of *DeepWealth* framework and code structure. The letters A, B and C denote folders for (A) 'R scripts', (B) 'Python scripts' and (C) 'Checkpoints'.

Creators

First Name	Middle Initial	Last Name	Organization	e-mail address	ORCID ID (optional)
Ali	-	Ben Abbes	FRB-Cesab	Ali.benabbes@fondationbiodiversite.fr Ali.benabbes@yahoo.fr	0000-0001-5714-7562
Jeaneth	-	Machicao	Escola Politécnica da Universidade de São Paulo (EPUSP), São Paulo, Brazil,	machicao@usp.br	0000-0002-1202-0194

Other personnel names and roles

First Name	Middle Initial	Last Name	Organization	e-mail address	ORCID ID (optional)	Role in project
Pedro	Pizzigatti	Corrêa	Escola Politécnica da Universidade de São Paulo (EPUSP), São Paulo, Brazil,	pedro.correa@usp.br	0000-0002-8743-4244	PI
Alison	-	Specht	Terrestrial Ecosystem Research Network, University of Queensland, Australia	a.specht@uq.edu.au	0000-0002-2623-0854	PI
Rodolphe	-	Devillers	Espace-Dev (IRD, Univ Réunion), La Réunion, France	rodolphe.devillers@ird.fr	0000-0003-0784-847X	Researcher
Jean	Pierre	Ometto	Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil	jean.ometto@inpe.br	0000-0002-4221-1039	Researcher
Yasuhisa	-	Kondo	Research Institute for Humanity and Nature (RIHN), Kyoto, Japan The Graduate University for Advanced Studies, SOKENDAI, Japan	kondo@chikyu.ac.jp	0000-0001-7670-4475	Researcher
David	-	Mouillot	MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Montpellier, France	david.mouillot@umontpellier.fr	0000-0003-0402-2605	PI

License

Keywords

Deep Learning, Socioeconomic Conditions Mapping, Satellite Images, Reproducibility

Funding of this work:

PI First Name	PI Middle Initial	PI Last Name	PI ORCID ID (optional)	Title of Grant	Funding Agency	Funding Identification Number
David		Mouillot	0000-0003-0402-2605	This project was conducted as part of the Belmont Forum PARSEC project, funded under the Collaborative Research Action (CRA) on Science-Driven		

				e-Infrastructures Innovation (SEI), with support from the synthesis center CESAB of the French Foundation for Research on Biodiversity In Brazil the PARSEC project is supported by the grant 2018/24017–3, of the São Paulo Research Foundation (FAPESP)		

Timeframe

- Begin date: 1996
- End date: 2019
- Data collection ongoing/completed: completed

Geographic location

24 countries in Africa (Angola, Benin, Burkina Faso, Cameroon, Cote d'Ivoire, Democratic Republic of Congo, Ethiopia, Ghana, Guinea, Kenya, Lesotho, Malawi, Mali, Mozambique, Nigeria, Madagascar, Rwanda, Senegal, Sierra Leone, Tanzania, Togo, Uganda, Zambia, and Zimbabwe).

Methods

Two separate CNNs based on the ResNet-18 architecture were trained on wealth index constructed from the Demographic and Health Surveys (DHS), Landsat and Night light imagery,

Data Provenance

Dataset title	Dataset DOI or URL	Creator (name & email)	Contact (name & email)
Wealth index constructed from the DHS	https://dhsprogram.com/data/available-datasets.cfm		
Trained models (output checkpoints)	https://doi.org/10.5281/zenodo.10575637	Ali Ben Abbes et al.	Ali.benabbes@yahoo.fr
LANDSAT/LC08/C01/T1_SR LANDSAT/LE07/C01/T1_SR LANDSAT/LT05/C01/T1_SR	https://landsat.gsfc.nasa.gov/	USGS/NASA Landsat Program	
NOAA/DMSP-OLS/CALIBRATED_LIGHTS_V4	https://doi.org/10.3390/rs70201855	Defense Meteorological Satellite Program (DMSP)	
NOAA/VIIRS/DNB/MONTHLY_V1/VCMSLCFG	https://doi.org/10.1080/01431161.2017.1342050	Defense Meteorological Satellite Program (DMSP)	
Imagery API	https://earthengine.google.com/	Google Earth Engine API	
ResNet-18 architecture (v2, with preactivation on Imagenet)	https://doi.org/10.48550/arXiv.1603.05027	Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun	

Data Table

Table 1 The remotely sensed sources used in the project.

Satellite	Sensor	Product
Landsat	Landsat 5 (1984-2013)	LANDSAT/LT05/C01/T1_SR
	Landsat 7 (1999-2022)	LANDSAT/LE07/C01/T1_SR
	Landsat 8 (2013-)	LANDSAT/LC08/C01/T1_SR
Nightlight	DMSP (Year <2011)	DMSP-OLS/CALIBRATED_LIGHTS_V4
	VIIRS (Year>2011)	VIIRS/DNB/MONTHLY_V1/VCMSLCFG

Scripts/code (software)

File name	Description	Scripting language
0_Download.ipynb	Download the satellite images based on the wealth index csv file	Python
1_process_tfrecords.ipynb	Split the downloaded TFrecords file by country_year_villages using the csv file	Python
2_create_incountry_folds.ipynb	Split the data into five folds using incountry configuration (based on the distance between villages in order to avoid the overlapping of the satellite image).	Python
3_dhs_baslines.ipynb	Generate a .npz file that resumes the Night Light images features (center, mean, etc.) Run the machine learning baselines	Python
4_dhs.ipynb	Generate a DataFrame that merges the .npz file generated in step 3 with the original csv file	Python
5_train.ipynb	Contains all training scripts for all configurations	Python
6_dhs_resnet_ridge.ipynb	Apply a ridge regression to concatenate the Resnet-MS and Resnet-NL	Python
6_dhs_incountry.ipynb	Calculate the performance metrics for all configuration	Python
8_test.ipynb	Test the training models for new villages	Python
Wealth_index_construction	Construct the wealth index from the DHS surveys	R
Grid_script	Calculate the (lat, lon) for a given country with a pixel ~6 km^2	R

Notes and Comments