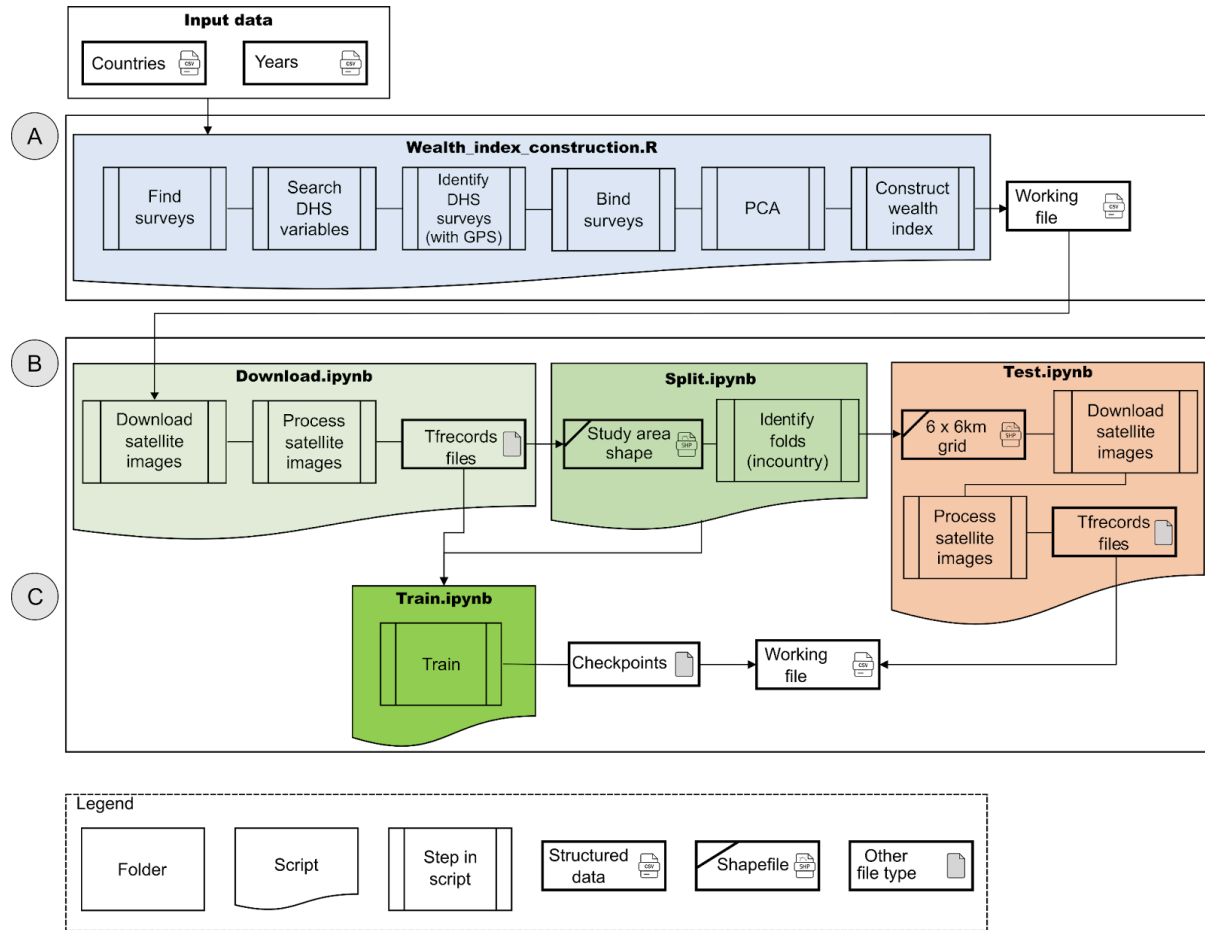


## Dataset Title

**DeepWealth:** DeepWealth: A Generalizable Open-Source Deep Learning Framework using Satellite Images for Well-Being Estimation

## Abstract

Measuring and representing socioeconomic variables at the scale of regions or countries are required in various contexts, in particular to inform public policies. If this is traditionally done using surveys, the integration of Deep Learning (DL) and Earth Observation (EO) data is becoming increasingly prevalent to estimate specific variables like overall wealth, offering better spatial and temporal coverages. This paper presents the end-to-end framework 'DeepWealth' that allows estimating a wealth index using open-source EO data and DL. DeepWealth uses a multidisciplinary approach incorporating satellite imagery, socio-economic surveys, and DL models to estimate a wealth index. We demonstrate the effectiveness and generalisability of DeepWealth by training it on 24 African countries and deploying it in Madagascar, Brazil and Japan to test its flexibility in different contexts. Our results show that DeepWealth provides accurate and stable wealth index estimates with an  $R^2$  of 0.69. DeepWealth empowers computer-literate users skilled in Python and R to estimate and visualize well-being-related data. 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	<a href="mailto:Ali.benabbes@fondationbiodiversite.fr">Ali.benabbes@fondationbiodiversite.fr</a> <a href="mailto:Ali.benabbes@yahoo.fr">Ali.benabbes@yahoo.fr</a>	0000-0001-5714-7562
Jeaneth	-	Machicao	Escola Politécnica da Universidade de São Paulo (EPUSP), São Paulo, Brazil,	<a href="mailto:machicao@usp.br">machicao@usp.br</a>	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,	<a href="mailto:pedro.correa@usp.br">pedro.correa@usp.br</a>	0000-0002-8743-4244	PI
Alison	-	Specht	Terrestrial Ecosystem Research Network, University of Queensland, Australia	<a href="mailto:a.specht@uq.edu.au">a.specht@uq.edu.au</a>	0000-0002-2623-0854	PI
Rodolphe	-	Devillers	Espace-Dev (IRD, Univ Réunion), La Réunion, France	<a href="mailto:rodolphe.devillers@ird.fr">rodolphe.devillers@ird.fr</a>	0000-0003-0784-847X	Researcher
Jean	Pierre	Ometto	Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brazil	<a href="mailto:jean.ometto@inp.br">jean.ometto@inp.br</a>	0000-0002-4221-1039	Researcher
Yasuhisa	-	Kondo	Research Institute for Humanity and Nature (RIHN), Kyoto, Japan  The Graduate University for Advanced Studies, SOKENDAI, Japan	<a href="mailto:kondo@chikyu.ac.jp">kondo@chikyu.ac.jp</a>	0000-0001-7670-4475	Researcher
David	-	Mouillot	MARBEC, Univ Montpellier, CNRS, Ifremer, IRD, Montpellier, France	<a href="mailto:david.mouillot@umontpellier.fr">david.mouillot@umontpellier.fr</a>	0000-0003-0402-2605	PI

## License

MIT

## 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	Belmont Forum	

				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: 2020
- 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	<a href="https://dhsprogram.com/data/available-datasets.cfm">https://dhsprogram.com/data/available-datasets.cfm</a>		
Trained models (output checkpoints)	<a href="https://doi.org/10.5281/zenodo.10575637">https://doi.org/10.5281/zenodo.10575637</a>	Ali Ben Abbes et al.	Ali.benabbes@yahoo.fr
LANDSAT/LC08/C01/T1_SR LANDSAT/LE07/C01/T1_SR LANDSAT/LT05/C01/T1_SR	<a href="https://landsat.gsfc.nasa.gov/">https://landsat.gsfc.nasa.gov/</a>	USGS/NASA Landsat Program	
NOAA/DMSP-OLS/CALIBRATED_LIGHTS_V4	<a href="https://doi.org/10.3390/rs70201855">https://doi.org/10.3390/rs70201855</a>	Defense Meteorological Satellite Program (DMSP)	
NOAA/VIIRS/DNB/MONTHLY_V1/VCMSLCFG	<a href="https://doi.org/10.1080/01431161.2017.1342050">https://doi.org/10.1080/01431161.2017.1342050</a>	Defense Meteorological Satellite Program (DMSP)	
Imagery API	<a href="https://earthengine.google.com/">https://earthengine.google.com/</a>	Google Earth Engine API	
ResNet-18 architecture (v2, with preactivation on Imagenet)	<a href="https://doi.org/10.48550/arXiv.1603.05027">https://doi.org/10.48550/arXiv.1603.05027</a>	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 >= 2012)	VIIRS/DNB/MONTHLY_V1/CMSSLCFG

## 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_baselines.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 <sup>2</sup>	R
visualizationBR.ipynb	Visualization maps of the estimated wealth index for Brazil	Python
visualizationJP.ipynb	Visualization maps of the estimated wealth index for Japan	Python

## Notes and Comments