Satellite images and deep learning for the prediction of socioeconomic indicators in the Vale do Ribeira

Jeaneth Machicao[1*], Iago Fava da Costa[1], Enrico Triñanes[1], Pedro Corrêa[1]

[1] University of São Paulo, School of Engineering, Computer Engineering Department, PB: 61548 São Paulo-SP, Brazil - Zip: 05508-900

machicao@usp.br

Abstract. *Due to low awareness and low investment in data collection and processing in developing countries, industry has struggled to plan and collaborate with public policy on data-driven socioeconomic decisions. This paper presents a deep learning approach to estimate socioeconomic indicators using satellite imagery. For this purpose, a region in southeastern Brazil called Vale do Ribeira was selected as a study case due to its data availability and environmental relevance. We used publicly available data to analyze three socioeconomic indicators: Longevity, Income and Literacy, the main representatives of the Human Development Index. Our preliminary results show a relevant correlation between satellite imagery and the income indicator, although they are not yet conclusive for longevity and literacy.*

1. Introduction

Socioeconomic indicators are a useful tool for data-driven decision making in government policy. However, such indicators can be expensive, difficult to obtain, and time consuming. The difficulty of obtaining reliable data is even more pronounced in developing countries, where the development aid community faces barriers to accessing trustworthy and up-to-date socioeconomic information [Blumenstock, 2016]. Tipically, census surveys are collected every ten years. This periodicity creates a data gap that affects the interpretability of the survey results, as they may not reflect the socioeconomic reality in real time. In this context, alternative data collections such as cell phone usage data, satellite imagery, or commercial translations [Blumenstock, 2016] can be widely, low-cost, and easily accessible.

The scientific community is well aware of this challenge, as evidenced by the numerous efforts to predict poverty in Africa using satellite imagery [Jean *et al*., 2016; Yeh *et al*., 2020; Ayush *et al*., 2020; Burke *et al*., 2021]. Other approaches use not only satellite imagery but also street view imagery to measure social, environmental, and health inequalities [Suel *et al*., 2019; Machicao *et al*., 2022]. In this context, the use of DL methods can help provide valuable information for environmental and socioeconomic monitoring, facilitating initiatives aimed at sustainable development and biodiversity conservation in the national territory.

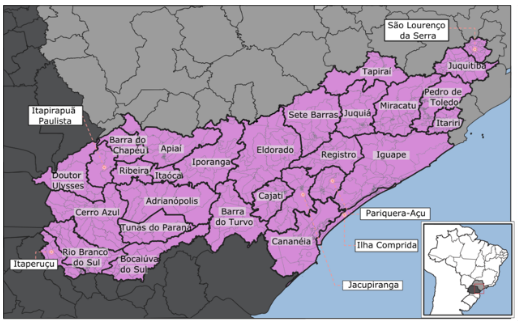
In this work, we adapted the DL methodology conducted by Jean *et al*. (2016), which aims to predict socioeconomic criteria in African countries based on satellite images of their territories. Our objective is to apply this methodology to the Brazilian context, more specifically to the Vale do Ribeira region as a case study for this approach due to the availability of data and its environmental and economic importance. Vale do Ribeira is the largest contiguous area of protected Atlantic Forest in Brazil, declared a

Natural Heritage of Humanity by UNESCO (United Nations Educational, Scientific and Cultural Organization) in 1999. We used a census survey conducted by the Brazilian Institute of Geography and Statistics (IBGE). We aim to obtain similar results to the experiments in Africa. This ongoing project is part of the PARSEC research group (https://parsecproject.org/), which aims to explore a method for estimating socioeconomic indicators in a cost-effective and easily accessible way, and to further develop satellite imagery analysis for use in developing countries.

2. Material and Methods

2.1. Data Gathering

Vale do Ribeira (VR) is located in the southeastern region of Brazil with an area of 28,306 km² within the states of São Paulo and Paraná. It includes 30 municipalities divided into 954 tracts, which are the subject of this experiment. 62% of this region is protected areas and its importance is determined by the well-preserved Brazilian Atlantic Forest, which represents 21% of the remaining forest in the country. Although it has a high economic development, this area is considered economically poor and has the lowest Human Development Index (HDI) in the state of São Paulo [Bueno *et al.*, 2020].



State	Municipality	Num. census sectors	State	Municipality	Num. census sectors
PR	Adrianópolis	21	SP	Itariri	61
SP	Apiaí	55	SP	Itaóca	9
SP	Barra Do Chapéu	11	SP	Jacupiranga	26
SP	Barra Do Turvo	14	SP	Juquitiba	54
PR	Bocaiúva Do Sul	23	SP	Juquiá	36
SP	Cajati	38	SP	Miracatu	48
SP	Cananéia	27	SP	Pariquera-açu	27
PR	Cerro Azul	42	SP	Pedro De Toledo	17
PR	Doutor Ulysses	13	SP	Registro	69
SP	Eldorado	30	SP	Ribeira	8
SP	Iguape	60	PR	Rio Branco Do Sul	58
SP	Ilha Comprida	28	SP	Sete Barras	26
SP	Iporanga	19	SP	São Lourenço Da Serra	27
PR	Itaperuçu	34	SP	Tapiraí	15
SP	Itapirapuã Paulista	9	PR	Tunas Do Paraná	12

Figure 1. The 30 municipalities of the Vale do Ribeira in southeastern Brazil (left). VR contains 954 census tracts, the boundaries of which are shown within each municipality. Statistics of census tracts, municipalities and states of VR (right).

2.1.1. Socioeconomic indicators

One of the most well-known and recognized indicators for interpreting socioeconomic development is the HDI adopted by the United Nations, which takes into account three dimensions: Health, Education, and Economic Development. Abreu *et al.* (2011) analyzed the documentation of social development at municipality level in Brazil using data from the IBGE to determine these three HDI dimensions. In this paper, we focused on the census tract level, so we use the method of Abreu *et al.* (2011) and adapt it for this purpose. We used 2010 census data from https://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_do_Universo/Agregados_por_Setores_Censitarios/ and created three proxy indicators: Longevity, Literacy, and Income, as follows:

Longevity: As a health dimension of the HDI, we calculated the longevity of the census tract using data on life expectancy at birth given by Eq. 1.

$$HDI - Longevity = \frac{(LF - 25)}{(85 - 25)} \tag{1}$$

Where LF represents life expectancy in years, 85 for maximum age, and 25 for minimum age, according to Atlas Brazil [PNUD 2012]. Abreu *et al*. (2011) calculated HDI-Longevity by census tract, assuming that the municipal HDI represents the average of each census tract. Therefore, for each census tract, we sum the population by age and make the distribution based on municipal life expectancy to determine the age of life expectancy for the tract.

Literacy: The education dimension of the HDI is composed of the period in which a population was formed. Abreu *et al*. (2011) suggested using a population's literacy as a proxy for this dimension because the IBGE census publications lack the specific question about the period of education. Therefore, we adopted the same concept by dividing the rate of alphabetized population into two different categories ($\boldsymbol{p}$ and $\boldsymbol{q}$), where $\boldsymbol{p}$ is the rate of alphabetized until 14 years old (end of elementary school), and $\boldsymbol{q}$ is the rate of alphabetization after 14 years old. These two values are calculated using Eq. 2.

$$HDI - Literacy = \frac{2p\ +\ q}{3} \tag{2}$$

Income: The income indicator represents the economic development of a region and is directly linked to gross domestic product (GDP). The IBGE publishes GDP results annually at the national and municipal levels. Abreu *et al*. (2011) proposed the use of household income as a good proxy for the distribution of municipal income (Eq. 3).

$$HDI - Income = \frac{ln(PC) - ln(min)}{ln(max) - ln(min)} \tag{3}$$

Where PC is the per capita income of a census tract, min = R$8.00 and max = R$4033.00 [PNUD 2012] are the reference values for minimum income and maximum income, respectively. PC is calculated as the division of the total nominal monthly income of responsible householders by the total resident population in the census tract.

2.2. Satellite imagery

We used two satellite data sources: Planet [Planet Team 2017] and Google Earth Engine (GEE) [Gorelick *et al*., 2017]. The Planet API was used to acquire daytime satellite imagery, while GEE API was used to obtain nightlight imagery. It should be noticed that the area of census tracts in VR is much smaller than the villages are of countries in Africa (original study by Jean *et al*., 2016). To aggregate the census data collected, it was necessary to create a cluster of images based on a 10 km x 10 km square centered on the smallest possible granularity: the census tracts. Each of these squares contained at least 10 randomly selected coordinate points. For each point, an image (centered on the selected point) was downloaded from the Planet API. Before downloading the data, a nightlight image of the VR region was acquired using GEE API. This nightlight image was used as a filter to select only cluster segments that had at least some nightlight level (greater than zero). The purpose was to obtain only segments with human activity to reduce noise when training the model. The dataset used was from the United States Air Force Defense Meteorological Satellite Program (DMSP) in collaboration with the National Oceanic and Atmospheric Administration's National Geophysical Data Center [NOAA-NGCD, 2010]. Note that NOAA-NGCD provided a separate nightlight after 2014 that is high resolution, but we did not use it because it does not match our survey years from IBGE 2010.

Despite the large variety of Planet images, no image was returned for some of the points requested via the API. For example, only 4296 of the expected 8800 images were downloaded. In addition, 510 were deemed unsuitable for the study and discarded because they were edges of photos taken from satellites, i.e., they did not contain any part of the desired image. Thus, a total of only 3786 images were used for the experiment in Vale do Ribera, representing 43% of the points originally selected. It should be noted that satellite imagery or census data are not available for all 954 census tracts, as they may be rural and semi-rural and IBGE hides the data for privacy reasons. For this reason, we selected only 629 census tracts for our experiments (Fig. 2).

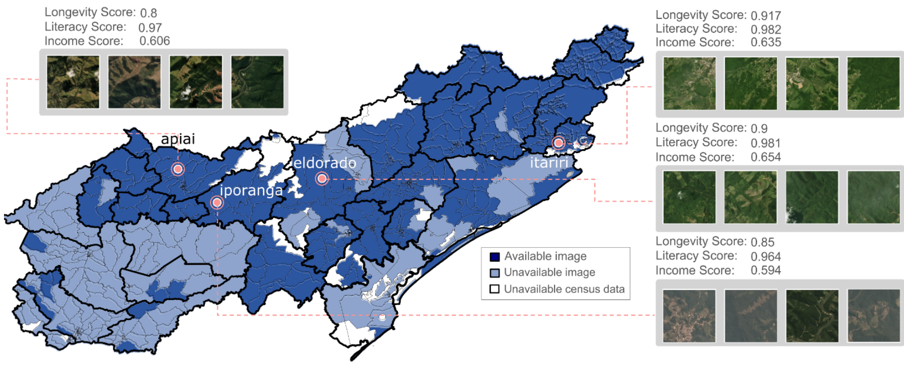


Figure 2. Map of Vale do Ribeira (VR). The 629 census tracts where satellite images were available are colored in blue, meanwhile, the light blue ones represent unavailable images, and the white correspond to census with no census data (for the three indicators). Random images samples were taken for some census tracts and respective socio-economic scores are shown.

3. Proposed Deep Learning replication

We proposed to replicate the experiments described in Jean *et al*. (2016) and to use as case study a Brazilian area, following the proposed workflow (Fig. 3). The methodology is as follows: (i) creation of a pre-trained model using a pre-trained Convolutional Neural Network (CNN) to learn the relationship between night light intensities; (ii) extracting a feature vector from the CNN output used for transfer learning so that each input diurnal satellite image corresponds to a feature vector and can be annotated with a socioeconomic indicator (HDIM dataset); and (iii) using a simpler regression model to predict socioeconomic indicators from the corresponding CNN feature vector output. Then, the prediction performance is quantified using the coefficient of determination (r^2) as known from Machine Learning (ML).

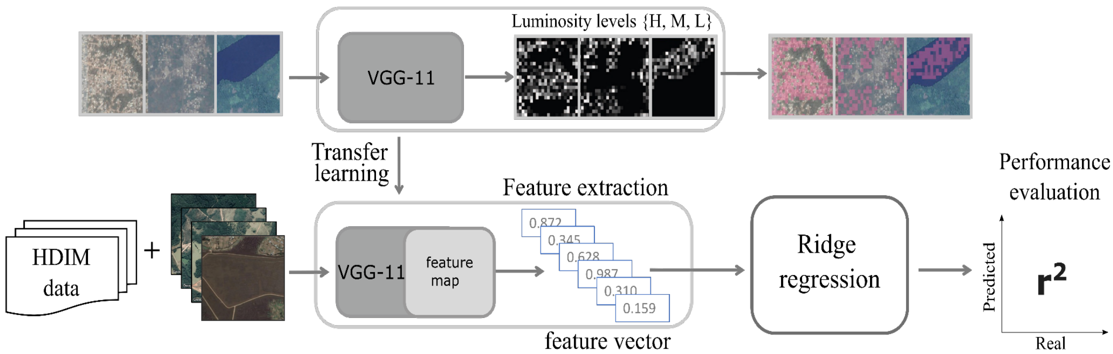


Figure 3. Proposed workflow for the replication of the Deep Learning method of [Jean *et al*., 2016] for the Vale do Ribeira case. First (from top to bottom), a pre-trained CNN is used

as a proxy for the poverty level by using nighttime satellite imagery for a supervised task of classification of luminosity levels. Second, the trained transfer learning network is used for daytime satellite imagery as feature extraction, such that a feature vector is obtained for each image sample. Finally, a ridge regression model is used to predict the socioeconomic values and evaluated by the coefficient of determination r^2.

3.1. Feature Extraction

We used the VGG-11 architecture [Simonyan and Zisserman, 2014] pre-trained with the ImageNet benchmark [Deng *et al*., 2009]. We applied the transfer learning technique by fine-tuning the CNN model using satellite images with nightlights as proxies for the socioeconomic indicators. First, we used Gaussian Mixture Models as proposed by Jean *et al*. (2016) to clusterize the nightlight images into three luminosity levels (low, medium, and high) of 0-5, 5-14, and 14-65, respectively. Then, the pre-trained neural network is applied to a supervised learning task with three classes. Note that the VGG-11 was adapted by replacing the last dense layer of 1000 (by default) with a dense layer of size 3 corresponding to the three luminosity levels.

After training the CNN model to predict nighttime light intensity, we used this learned model as a feature extractor from the daytime satellite images. In doing so, we discarded the last layer of the model, the nighttime light intensity classification layer, and kept the feature maps from the last convolutional layer to compose the feature vector. In this way, the dataset of daytime satellite images is fed into the pre-trained network and it returns a feature vector with 4096 dimensions as output for each image sample.

An interesting aspect of the method is that each census tract can have many images, where the position does not necessarily correspond to the actual location where the census was taken (for privacy reasons). Since there is one feature vector per image within a 10x10 km region, we averaged all these feature vectors as described by Jean *et al*. (2016). The final step is to actually train these feature vectors using Ridge regression to predict the socioeconomic indicators.

3.2 Training and Performance Evaluation

The dataset was divided into 80% for training and 20% for testing, according to the 5-fold cross-validation for all experiments. This validation is a statistical method used to evaluate the predictive ability of a model in general. To measure the performance of this method, we calculated the coefficient of determination in the cross-validation according to Eq. 4, where N is the number of census tracts with satellite images. y is the real HDIM value, $\hat{y}$ is the value predicted by the regression model, $\bar{y}$ is the average of $\hat{y}$ and $r^2 \in [-\infty, 1]$. In ML it is well established that closest r^2 values as close to 1 as possible indicate positive correlation, which quantifies the predictive quality of the model. Thus, the higher the value, the better the performance of the predictive model.

$$r^2(y, \hat{y}) = 1 - \frac{\sum_i^N (y-\hat{y})^2}{\sum_i^N (y-\bar{y})^2} \tag{4}$$

3. Results

We used the three socioeconomic indicators: HDI-Longevity, HDI-Literacy and HDI-Income, and we trained the DL model according to the proposed workflow. Figure. 4 shows a plot comparing real and predicted value, indicating the coefficient of determination (r^2) for each of the three indicators.

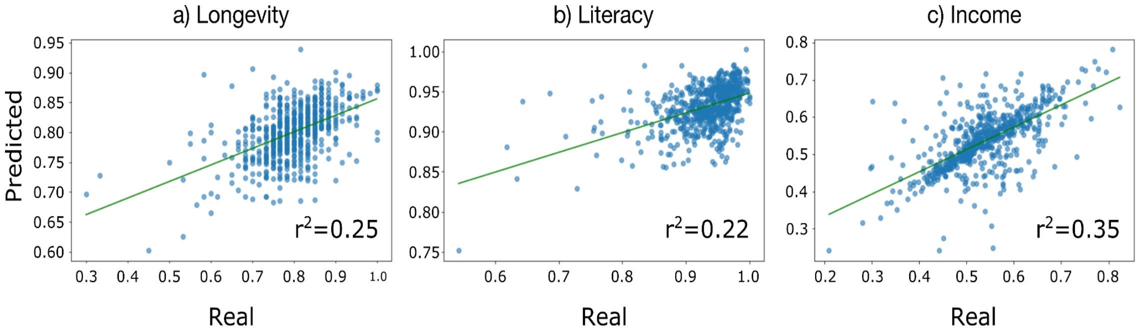


Figure 4. Results obtained for Vale do Ribeira. Each plot corresponds to the graph of predicted versus real for each of the three indicators. The coefficient of determination r^2 shown was calculated by cross-validation.

3.1. Longevity

Three different features can be observed for the longevity index (Fig. 4a). First, one can see that the data points appear in columns. This is due to the structure of the numbers in the real longevity index, as they are discrete and rounded values. At the same time, the predicted values are continuous representations. Second, the main cluster of points is right-sided. This is because VR has a high longevity rate, based on HDI-Longevity (Eq. 1), at least for the census tract where we collected images and census data. Third, we obtained a coefficient of determination ($r^2 = 0.25$), which is not a large value. In this particular case, the missing data could have a negative effect on the index in addition to the different data structures already described.

3.2. Literacy

For the literacy index, we obtained ($r^2 = 0.22$) (Fig. 4b), which is consistent with the results for longevity in two aspects. First, it is possible to explain the literacy index by the high alphabetization rate in the studied region, but we only considered cluster sections for which downloaded images and census data were available. Second, the literacy rate is also not ideal, with a coefficient value smaller than that of the first index. This can be explained by the same reasons mentioned earlier, namely the lack of downloaded images. Nevertheless, it would be interesting to repeat this index with more recent and reliable data sources to obtain more solid conclusions about the correlation.

3.3. Income

The income index (Fig. 4c) performed best in the experiments. For this index, we can note that the coefficient of determination ($r^2 = 0.35$) is the highest among the three studied. Compare this with the work of Jean *et al*. (2016), where the authors obtained the best case ($r^2 = 0.55$). The correlation coefficient is not ideal, but from the plots produced we can see that the real and predicted values are proportional. A clear crescent line can be seen, that is, when the real values are low or high, the predicted values are also proportionally low or high. These results show that our work can approximately predict the census index. It can be concluded that the algorithm is very promising in predicting socioeconomic indicators from satellite images.

4. Discussions and Conclusions

In this paper, we have presented an approach that is interesting not only for its computational potential, but also for its low resource consumption and the wide availability of satellite imagery, which can be collected more frequently compared to the census surveys taken every ten years. This periodicity limitation, is the main motivation to search for computational approaches based on machine learning.

Our experiments showed that the income estimation model produced the most promising result and was consistent with intuition: the wealth of a region is reflected on the visual aspects of the area captured by satellite image. Longevity could also be reflected in the region's photos, but the algorithm did not perform as well as the experiment on income. Literacy, however, remains inconclusive as it has no obvious relationship with the satellite imagery, but still corresponds to some degree with the income indicator as wealthier families tend to be literate.

We should mention our main limitations that could significantly affect the efficiency of our results and suggest possible ways to improve performance. Our dataset is sparse compared to the experiments of Jean *et al*. (2016) in Africa. This is mainly because the algorithm to download images from the region using the Planet API did not fully cover the region due to the gaps, and possibly the input to the training algorithm was biased by the availability of the images, which negatively affected the model accuracy. Perhaps the quality of Planet's output image was acceptable.

Further studies could focus on optimizing the coefficient of determination and covering a larger area to increase the accuracy of the model. It is hoped that when the 2021 IBGE census is released, the data collection will be adjusted to better utilize the newly obtained census survey and investigate time stamp estimation. Besides, in our experiment we used VGG architecture, mainly to compare with the results from the original work. On future work, it could be interesting to use other neural networks architecture, as well as the use of more diverse datasets to be extended to regions other than the Vale do Ribeira, possibly using different granularities (from census tracts up to the municipality level). To better understand the nuances of the learning process, it would be interesting to explore convolutional neural network visualization techniques to better understand the reliability of the algorithm and explore methods of explainable artificial intelligence.

5. Acknowledgments

PARSEC project is funded by the Belmont Forum, Collaborative Research Action on Science-Driven e-Infrastructures Innovation. J. M. acknowledges a scholarship from the São Paulo Research Foundation (FAPESP #2020/03514-9). P. C. thanks support from FAPESP (Grant #2018/24017-3).

6. References

Abreu, M., Oliveira, J., Andrade, V., & Meira, A. (2011). Methodological proposal for spatial calculation and analysis of the intra-urban HDI of Viçosa, Brazil. Revista Brasileira de Estudos de População. 28. 169–186.

Ayush, K., Uzkent, B., Tanmay, K., Burke, M., Lobell, D., & Ermon, S. (2021). Efficient Poverty Mapping from High Resolution Remote Sensing Images. Proceedings of the AAAI Conference on Artificial Intelligence, 35(1), 12–20.

Blumenstock, J. E. (2016). Fighting poverty with data. Science, 353(6301), 753–754.

Bueno, G.W., Leonardo, A.F.G., Machado, L.P., Brande, M.R., Godoy, E.M., & David, F.S. (2020). Indicadores de sustentabilidade socioambiental de pisciculturas familiares em área de Mata Atlântica, no Vale do Ribeira - SP. Arquivo Brasileiro de Medicina Veterinária e Zootecnia, 72(3), 901–910.

Burke, M., Driscoll, A., Lobell, D. B., & Ermon, S. (2021). Using satellite imagery to understand and promote sustainable development. Science, 371(6535).

Deng, J., Dong, W., Socher, R., Li, L-J., Kai, L., & Li F-F. (2009). ImageNet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18–27.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016) Combining satellite imagery and machine learning to predict poverty. Science, 353(6301), 790–794.

Machicao, J., Specht, A., Vellenich, D., Meneguzzi, L., David, R., Stall, S., Ferraz, K., Mabile, L., O'Brien, M. & Corrêa, P. (2022). A Deep-Learning Method for the Prediction of Socio-Economic Indicators from Street-View Imagery Using a Case Study from Brazil. Data Science Journal, 21(1), p.6.

National Geophysical Data Center. (2010). Version 4 DMSP-OLS Nighttime Lights Time Series.

Planet Team (2017). Planet Application Program Interface: In Space for Life on Earth. San Francisco, CA. https://api.planet.com

PNUD. Índice de Desenvolvimento Humano Municipal - IDHM: Metodologia. (2012). https://atlasbrasil.org.br/acervo/atlas

Simonyan, K.; Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. https://arxiv.org/abs/1409.1556.

Suel, E., Polak, J.W., Bennett, J.E., & Ezzati, M. (2019). Measuring social, environmental and health inequalities using deep learning and street imagery. Scientific Reports, 9(1), 6229.

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., & Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. Nature Communications, 11.