A Deep-Learning Method for the Prediction of Socio-Economic Indicators from Street-View Imagery Using a Case Study from Brazil

RESEARCH PAPER

JEANETH MACHICAO
ALISON SPECHT
DANTON VELLENICH
LEANDRO MENEGUZZI
ROMAIN DAVID
SHELLEY STALL
KATIA FERRAZ
LAURENCE MABILE
MARGARET O'BRIEN
PEDRO CORRÊA

*Author affiliations can be found in the back matter of this article

ubiquity press

ABSTRACT

Socioeconomic indicators are essential to help design and monitor the impact of public policies on society. Such indicators are usually obtained through census data collected at 10-year intervals, which are not only temporally coarse but expensive. Over recent years other ways of collecting data and producing these indicators have been explored, in particular using the new surveillance capabilities that remote observations can provide. The objective of this paper is to evaluate the assessment of socioeconomic indicators using street-view imagery, through a case study conducted in a region of Brazil, the Vale do Ribeira, one of the poorest semi-rural regions in Brazil. In this study we used socioeconomic indicators collected by the Brazilian Institute of Geography and Statistics (IBGE) and used Google Street View (GSV) images as our source of remote observations. A pre-trained convolutional neural network (CNN) was used to predict socio-economic indicators from GSV. To evaluate the performance of the classifier, we performed five-fold cross-validation between the predicted indicator and its true value. The best performance was obtained for the highest income class, with 80% of correct prediction. We conclude that the method has the potential to predict socioeconomic indicators across a large area with social challenges such as Vale do Ribeira, and that the network model is general enough to be used even when the imagery dataset is from semi-rural areas. This demonstrates the applicability of GSV datasets for similar settings and perhaps ensuring their replicability, which is a scientific requirement that requires further experimentation/evaluation.

CORRESPONDING AUTHOR:

Jeaneth Machicao

University of São Paulo, BR

machicao@usp.br

KEYWORDS:

socioeconomic indicators; deep-learning; data science; google street view

TO CITE THIS ARTICLE:

Machicao, J, Specht, A, Vellenich, D, Meneguzzi, L, David, R, Stall, S, Ferraz, K, Mabile, L, O'Brien, M and 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: 6, pp. 1–15. DOI: *https://doi.org/10.5334/dsj-2022-006*

1 INTRODUCTION

The primary goal of the United Nations 2030 Agenda for Sustainable Development (2015) is to 'eradicate extreme poverty for all people everywhere' (*https://sdgs.un.org/goals/goal1*). This goal proposes international and national targets and standards for various indicators of economic well-being. Establishing the thresholds for these indicators is a challenge but obtaining the basic information on these variables around the world is an equal, if not greater, challenge.

Most economic and population statistics used to define a socio-economic indicator are obtained through a national census. A census is taken at regular intervals, usually every 10 years. In order to make international comparisons, the censuses of different countries need considerable harmonisation, as there is no global standard for their content. There is also no synchronisation in the dates when the censuses are conducted in different parts of the world. This variability makes global predictions highly inaccurate. The Covid-19 pandemic has illustrated this inadequacy, predictions depending on data collected at widely varying spatio-temporal scales (Franch-Pardo et al., 2020). In recent years other means to collect socio-economic data have been explored, including using the new monitoring capacities that remote observations, such as satellite imagery, can provide.

Satellite imagery has been used with some success to collect information on socio-economic status, particularly using machine learning (e.g. Jean et al., 2016; Xie et al. 2016, Ayush et al., 2020 and Burke et al., 2021). Although satellite images have good resolution with broad global coverage and availability to the public at a reasonable cost, they only contain information on the vertical perspective. It is therefore worth exploring other aspects of remote detection from a horizontal perspective. Google Street View (GSV) is one possible source of such data. It offers geolocated images along roads allowing virtual exploration of an area.

There are a number of examples of the use of GSV to discover information about various attributes of human activity, such as detecting the spatial occurrence of ertain car models (Gebru et al., 2017), assessing the amount of green space for health outcomes (Larkin and Hystad, 2019), detecting urban conditions that predispose criminal activity (He et al., 2017), and in detecting the socio-economic status of urban neighbourhoods (Diou et al., 2018 in Greece and Suel et al., 2019 in London).

These images are usually trained with a deep learning network such as convolutional neural networks (CNN). These networks are known for their excellent performance in tackling various machine learning and artificial intelligence tasks. Using these relatively modern techniques, a large number of images can be scanned in an objective manner. The machine is trained using a series of images to recognise patterns that may be an abstraction to the human eye. Understanding what is in the 'black box' of the deep neural network is an area of current research (Li, et al., 2021; Abitbol, et al., 2020 and Dai, et al., 2021).

Monitoring socio-economic inequalities at the global scale requires the use of standardized global composite indicators. Most of them have been developed by the statistics departments of global organisations such as the United Nations, the Organisation for Economic Co-operation and Development (OECD), and the World Bank (e.g., Stiglitz et al., 2009; United Nations Department of Economic and Social Affairs, 2015; Statistical Office of the European Union, 2017). They are used in various international surveillance programmes (e.g. Household Surveys, Living Standards Measurement Study, Demographic & Health Surveys), or have been developed by national statistical offices. They rely on a wide range of data, some of which are provided at the granular, people-centred level and therefore not easily collected in a standardised way.

The United Nations Human Development Index (HDI), for example, is a summary measure of average performance around key dimensions of human development: a long and healthy life, level of education, and a decent standard of living. The HDI is the geometric mean of the normalized indices for each of these three dimensions. The health dimension is assessed by life expectancy at birth, and the education dimension is measured by the average number of years of schooling for adults aged 25 and over and the expected number of years of schooling for children at school entry age. The standard of living dimension is measured by gross national income (GNI) per capita. The HDI uses the logarithm of income to reflect the decreasing importance of income as GNI increases. The scores for the three HDI dimensions are then aggregated into a composite index using a geometric mean.

Well-documented data, methods and workflows help ensure that research can be well evaluated, reproduced and replicated (National Academies of Sciences, Engineering, and Medicine, 2019). We aim to ensure that our methods are fully reproducible as well as replicable. The data, software, and workflow used for the paper is preserved in a trusted repository, described in the availability statement and cited in the references. Our methods are detailed in a linked Jupyter notebook.

Most of the previous applications of GSV have been in urban environments, where the density of images and available indicators match. In this paper, (a) we will replicate the Suel et al. (2019) use of GSV to assess socio-economic indicators in London to (b) assess the efficacy of the use of GSV in detection of some selected socio-economic indicators in a semi-rural environment (Vale do Ribeira in Brazil), and (c) to document our research workflow in a replicable manner. To our knowledge no previous study has covered such a large area, and a formal assessment of replicability is rarely documented.

2 CASE STUDY: VALE DO RIBEIRA

Our case study is situated in the Vale do Ribeira (VR), one of the most biodiverse areas in the world, declared as a Natural Heritage of Humanity site by UNESCO (United Nations Educational, Scientific and Cultural Organization) in 1999. It covers 28,306 square kilometers and includes 30 municipalities distributed across two different Brazilian states, São Paulo and Paraná, in southeastern Brazil (***Figure 1***).

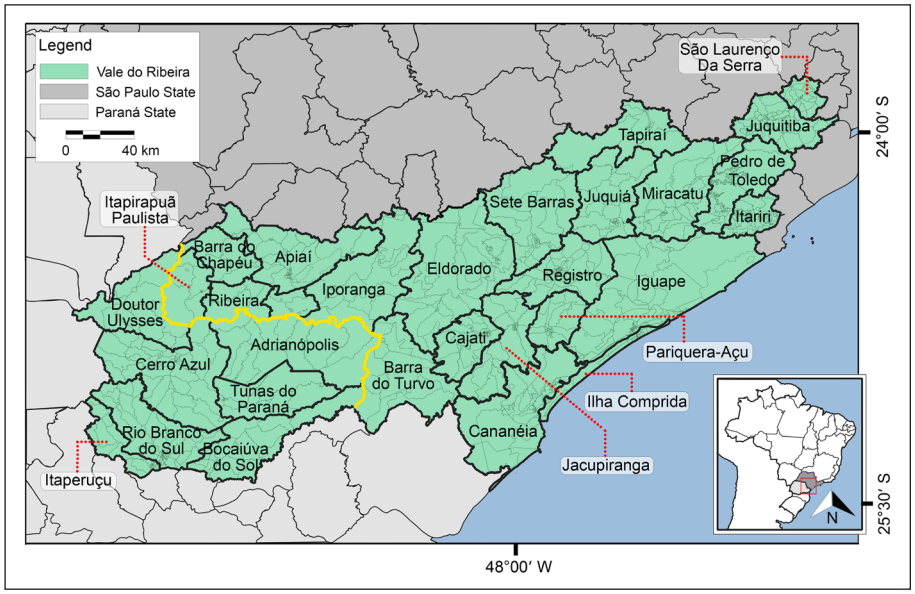

Figure 1 The 30 municipalities of the Vale do Ribeira in southeastern Brazil. The boundary between the States within which the Vale do Ribeira lies is shown by a yellow line. Smaller divisions within each municipality are 'census sectors', each containing about 300 households. These census sectors are the finest subdivision available in IBGE publications (*https://ibge.gov.br*).

This region is noted for the preservation of its forests and its great ecological diversity. The area contains a mosaic of protected areas covering a total of 4,700 square kilometers covering 62% of the area of the Vale do Ribeira. These protected areas–part of the World Heritage-listed Atlantic Forest South-East Reserves–comprise one of the largest and best-preserved areas of Brazilian Atlantic Forest, one of the most threatened biomes in the world (*https://whc.unesco.org/en/list/893*). The Vale do Ribeira holds 21% of the total area of this forest in Brazil, making it the largest contiguous area of any ecosystem in Brazil (Dias et al., 2015). Besides that, the area is host to important indigenous communities.

The Vale do Ribeira, in contrast to the otherwise wealthy São Paulo State, is economically poor having the lowest HDI of the state (Bueno et al., 2020). The region has low population density, with 0.66% of the population of São Paulo State. The sewage network is below the São Paulo State average and there is a high illiteracy rate. On the other hand, it has the best health conditions and health surveillance in the state of São Paulo (Mendes et al., 2015).

3 METHODS

3.1 OVERALL WORKFLOW

In this section, we present the workflow used to replicate the work of Suel et al. (2019). The workflow was divided into four components (***Figure 2***):

Step 1, GSV and IBGE data acquisition and aggregation. Socio-economic proxy indicators were calculated from the census data available through the IBGE. GSV images and socio-economic proxy indicators were then aggregated to obtain image-labeled datasets.

Step 2, feature extraction, in which a pre-trained VGG-16 network model (removing the last layers) was applied to extract one feature vector for each image.

Step 3, training, in which the four feature vectors (one per angle) with inputs averaged element-wise, were fed to the last fully connected layer of the VGG-16. The loss function used was the cross entropy. The training followed the k-fold method.

Step 4, validation, in which the predicted instances were compared with the actual, obtained in Step 1. Performance metrics were calculated to validate the training step, and graphs were obtained to better analyze the results, such as the differences between the real and predicted value and the confusion matrix.

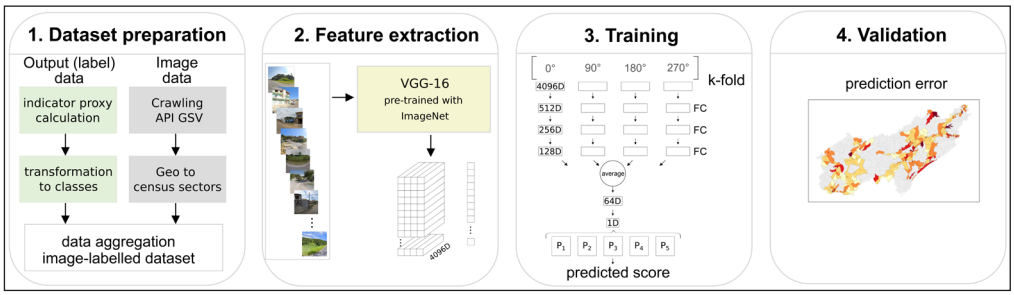

Figure 2 Workflow diagram showing the four main steps: 1) dataset preparation, 2) feature extraction, 3) training, and 4) validation. Street image samples were obtained from Google Street View.

3.2 DATA ACQUISITION AND AGGREGATION

3.2.1 Construction of socioeconomic indicators for Vale do Ribeira

Socio-economic data were obtained from the Brazilian Institute of Geography and Statistics (IBGE), which is responsible for official statistical and geoscientific information. The IBGE has been conducting surveys of the Brazilian population every ten years since 1872. In each of these surveys, all households in the 5,565 municipalities of Brazil are interviewed. The data are compiled, excluding data that could identify specific individuals, businesses, entities or products. We used data from the 2010 census, the latest available data at the time of our work. Details of the data sources we used are given in Section 6 of this paper.

There are there are 30 municipalities 954 census sectors in the Vale do Ribeira (***Figure 1***). However, it should be noted that IBGE does not publish the data if a defined census sector has a low population density or if there are too few answers to a question to ensure data protection. For the variables used in our analysis, 30 sectors fell into this category, resulting in 880 valid sectors for this work (***Table 1***).

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

Table 1 Statistics of census sectors, municipalities and states in the Vale do Ribeira.

The demographic census survey provides multiple variables for different domains. The HDI is an internationally recognised index to assess a country's development not only in economic terms, but also to include data on government policies and practices that affect well-being, health, and education. The HDI consists of three independently calculated dimensions: income, longevity and education.

The IBGE and Atlas Brazil (*https://atlasbrasil.org.br/acervo/atlas*) publish a municipal HDI every decade based on the census data and other surveys (Abreu et al., 2011). Gross Domestic Product per capita is used to measure HDI-Income. For municipal HDI-Income, IBGE uses per capita income, therefore we used the same strategy to construct an intra-municipal proxy. The formula used to calculate this index was as follows:

$$HDI - Income = \frac{ln(PC) - \ln(min)}{\ln(max) - \ln(min)}. \quad (1)$$

Where PC is the monthly per capita income of a census sector, min and max are the reference values for minimum and maximum income respectively. PC is calculated as dividing the total nominal monthly income of responsible household heads by the total resident population of the census sector. Source files can be found on the IBGE website and detailed in Section 6. The minimum and maximum income is set to min = R$ 8.00 and max = R$ 4033.00, respectively (Atlas Brazil, 2013).

To construct the outcome labels for the predictive model, we focused on the intra-municipal HDI-Income indicator, which is divided into five classes according to the HDI value. The first class contains the bottom fifth of the population on the income scale (i.e., the population with the lowest income), the second class contains the population with income levels between 20% and 40% of the maximum, and so on, with the top class containing the population with the highest income levels (80–100%) (***Table 2***).

INCOME SCORE	HDI–INCOME VALUE	ABSOLUTE INCOME REFERENCE	USD (2021 6 APRIL)
1	HDI [0.00–0.20]	R$ 8.00–R$ 813.00	$1.41–$143.52
2	HDI [0.20–0.40]	R$ 813.00–R$ 1618.00	$ 143.52–$285.63
3	HDI [0.40–0.60]	R$ 1618.00–R$ 2423.00	$ 285.63–$427.74
4	HDI [0.60–0.80]	R$ 2423.00–R$ 3228.00	$ 427.74–$569.85
5	HDI [0.80–1.00]	R$ 3228.00–R$ 4033.00	$569.85–$711.97

Table 2 The Income Score and income range for each HDI income class calculated on a monthly basis. The details of the source data are found in section 6.

3.2.2 Street View Images

The GSV images of Vale do Ribeira were acquired using a crawling algorithm that automated the acquisition of information using the Application Programming Interface (API) service of Google Street View (*https://developers.google.com/maps/documentation/streetview/overview*). We defined a bounding box for the area and established uniformly distributed (in two-minute intervals) geolocated points as shown in ***Figure 3*** for a sample census sector.

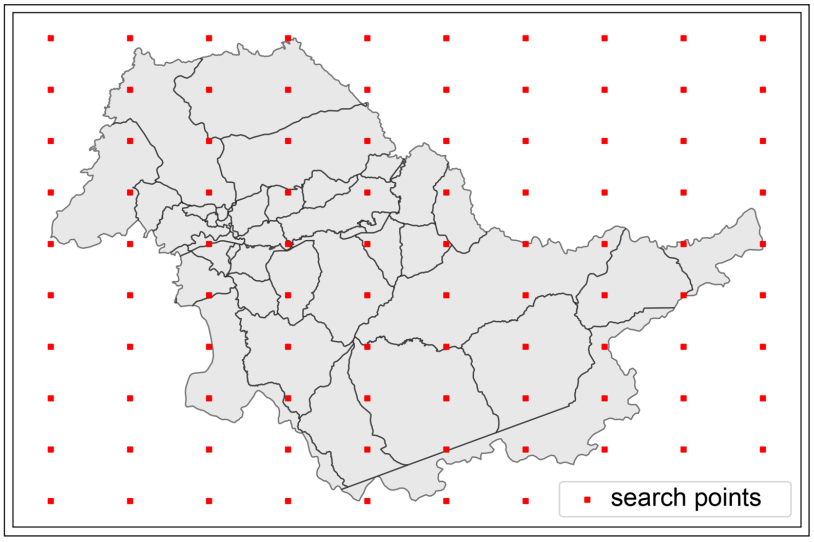

Figure 3 The sampling regime using an example of the Juquitiba municipality. The red dots represent the defined geolocated points searched by the crawler algorithm.

The first step in using the API is to determine if GSV images for each defined geolocation are available. If they are, the service returns the unique identifier for the nearest available panoramic image. It should be noted that these identifiers correspond to the most recently acquired images and their timestamp varied from 2011 to 2019 when accessed in March 2021.

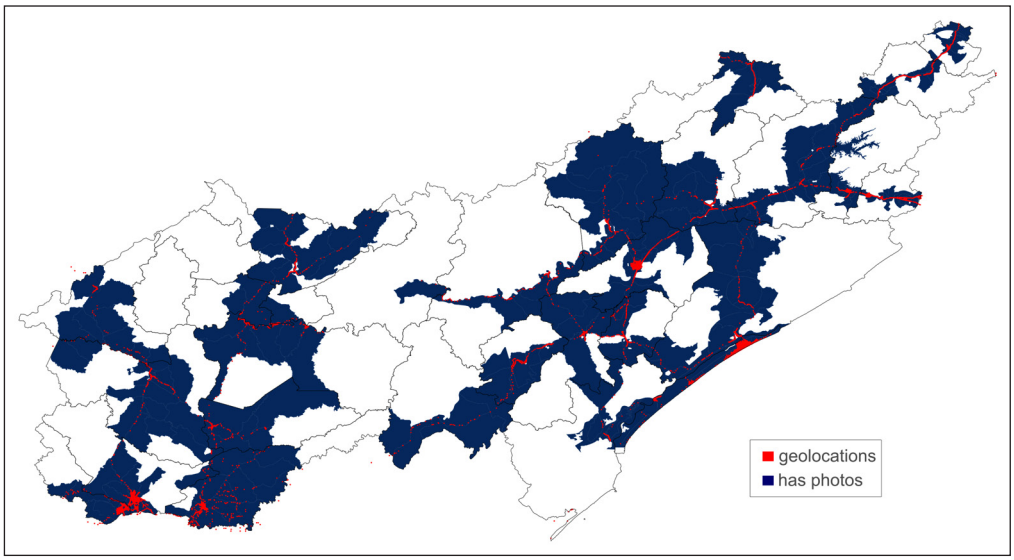

Figure 4 Geolocations at which GSV images were available, the census sectors shown in blue. In white the sectors in which there was no GSV available.

Since GSV images are panoramic ('panoid') images with cylindrical projection, we requested four picture orientations (0°, 90°, 180° and 270°) to fully cover the image view. In this manner we obtained 112,368 unique images from the Google Street View API for our study area, corresponding to 28,092 postcodes (also known as CEP in Brazil). It should be noted that only 500 census sectors had images. We were unable, for example, to acquire any images for the municipalities of Barra do Chapéu and Itapirapuã Paulista, both composed mainly of rural and semi-rural areas. We therefore used the ground truth (census) classification for these census sectors.

3.3 FEATURE EXTRACTION

Following the method proposed by Suel et al. (2019), we used a VGG-16 network model (Simonyan and Zisserman, 2015), where weights are initialized with the pre-trained model using an ImageNet benchmark (Russakovsky et al., 2015). This method focuses on large objects in millions of images and behaves as a feature extractor and afterward transfers the learning to a dataset of images. This is a common method for fine-tuning a pre-trained CNN to other novel tasks with a small dataset (Yosinski et al. 2014; Zhuang et al., 2021).

When a CNN is used as a feature extractor, usually the last three fully connected layers are removed and the remaining feature maps of the last convolutional layer are used to compose the feature vector (also called latent space). The street view dataset is fed into the pre-trained network, which then returns a feature vector with 4096 dimensions as output for each image. Since four angles are presented at each time, then the feature representation for each panoramic location consists of four 4096-D vectors, related to the four images at each geolocation. The input images were reduced in size from 512 × 512 to 224 × 224 pixels, the default input size for the VGG-16 model.

3.4 TRAINING

After applying the deep learning model, the three remaining fully connected layers of the VGG-16 network are used to predict each output label, performing an ordinal classification from four angle images for each CEP. All of our source code was implemented using TensorFlow 2.2 in Python 2.7. Besides that, we conducted all the experiments (Section 4) on a workstation with an 8-core CPU Intel(R) Core(TM) i7 and 1TB RAM and it took approximately 96hs to complete the training process. The network was trained for 1,000 iterations and a batch size of 400 images per step. Batch normalization was used in all layers of the network except the output layer. The Adam optimizer was used and set with a learning rate of 5×10^{-6}. The network that provided

the best validation error in the last iteration was kept as the final model. The categorical cross-entropy given by the following loss function was used.

$$\min_{w} \sum_{n} \sum_{m} y_n^m \ln p_n^m \tag{2}$$

Where w are the network weights, y_n is the label vector of the n^{th} sample (for one-hot encoding), and p_n^m is the probability of the m^{th} income score for the n^{th} sample. We paid special attention to maintain a 'natural' distribution in both the training and test datasets to prevent overfitting.

The network model used all four feature vectors corresponding to the image orientation as inputs for the last three fully connected layers (***Figure 2***). These four feature vectors are combined by calculating the average of all the elements of the vectors into a new vector. This neutralises the semantic information of the separate images, creating only one panoramic-view feature vector. This feature vector is then passed to the last dense layer which compute a single continuous value between 0 and 1 using the sigmoid function. This single continuous value was then used to compute probabilities (P_1, P_2, P_3, P_4, P_5), considering them as the probability of Bernoulli trials (coin toss) and transforming to a five-class problem, i.e. an indicator score between 1 and 5. Finally, the obtained scores are compared with the actual scores.

3.5 VALIDATION

To evaluate the performance of the classifier, we used five-fold cross-validation (Bishop, 2006), in which the predicted indicator for the census sector with the known but hidden score, was compared to its true value. This cross-validation method is a reliable strategy because it divides the data into two mutually exclusive sets: 80% for the training dataset (the set of instances used for training purposes) and the remaining 20% for the testing set (for testing purposes), thus ensuring the generalizability of the model.

We also evaluated performance in the same manner as Suel et al. (2019), using the Pearson correlation coefficient (r), which measures the correlation between true and predicted classes (the closer to 1, the higher the relationship), the Kendall-Tau's coefficient (τ), which is interpreted similarly to the Pearson coefficient, and the mean absolute error (MAE), which expresses the average prediction error of the model. For the final accuracy metrics, we measured the percentage of correctly predicted classes with an allowable error margin of ±1, ±2 per class, i.e., including the corresponding score ±1 or ±2.

A confusion matrix was used to evaluate the performance of the network model. In this matrix, the columns represent the true classes to which the elements belong and the rows correspond to the classes predicted by the model. For this we converted continuous values of the socio-economic indicator to an ordinal value, with an error margin of ±1, ±2. A perfect case would be a diagonal value of 100% showing a perfect match. Otherwise the prediction is expected to decrease as the score decreases.

4 RESULTS

4.1 INCOME PREDICTION FOR VALE DO RIBEIRA

The kinds of images we obtained from the geolocated positions are illustrated in ***Figure 5***. These samples were extracted from each geolocation according to their true income score derived from the census data. Each row shows the view from two different geolocations with four images corresponding to the set suite of angles (0°, 90°, 180° and 270°).

The distribution of images across income indicators was naturally unbalanced, from 4% of the total number of images for the lowest income value to 42% for the highest (***Figure 6***), which reflects the actual income situation of the residents of Vale do Ribeira. In turn we think this data imbalance will not affect the results, an assumption consistent with the work of Tetila et al. (2020).

The use of a VGG network pre-trained on the ImageNet benchmark gives us confidence in using an imbalanced dataset. As far as we know, there is yet no formal proof of the sensitivity of a VGG network to imbalanced data, but the work of Johnson and Khoshgoftaar (2019, page 30) has shown that using a VGG network as a baseline is more than adequate to handle imbalanced classes. The ability to deal with imbalances in this way shows the power of reusing strong feature extractors trained on large volumes of data such as ImageNet.

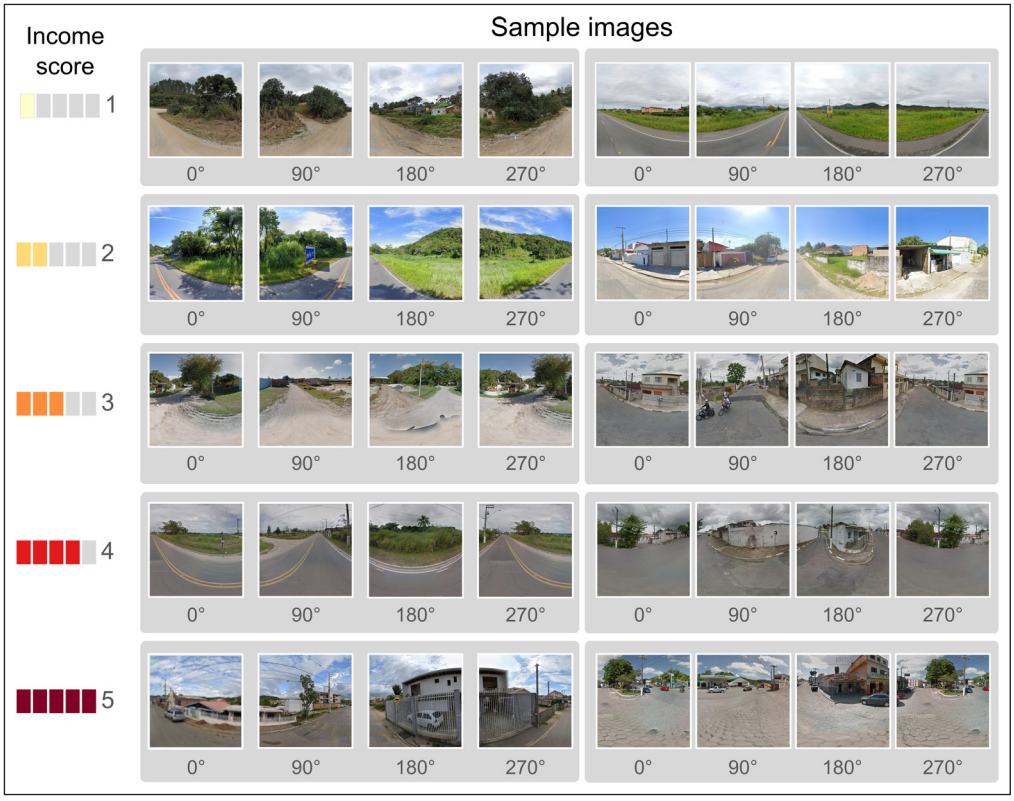

Figure 5 Examples of two panoramic images for each income score taken randomly from Vale do Ribeira. Each image sample was taken over four angles.

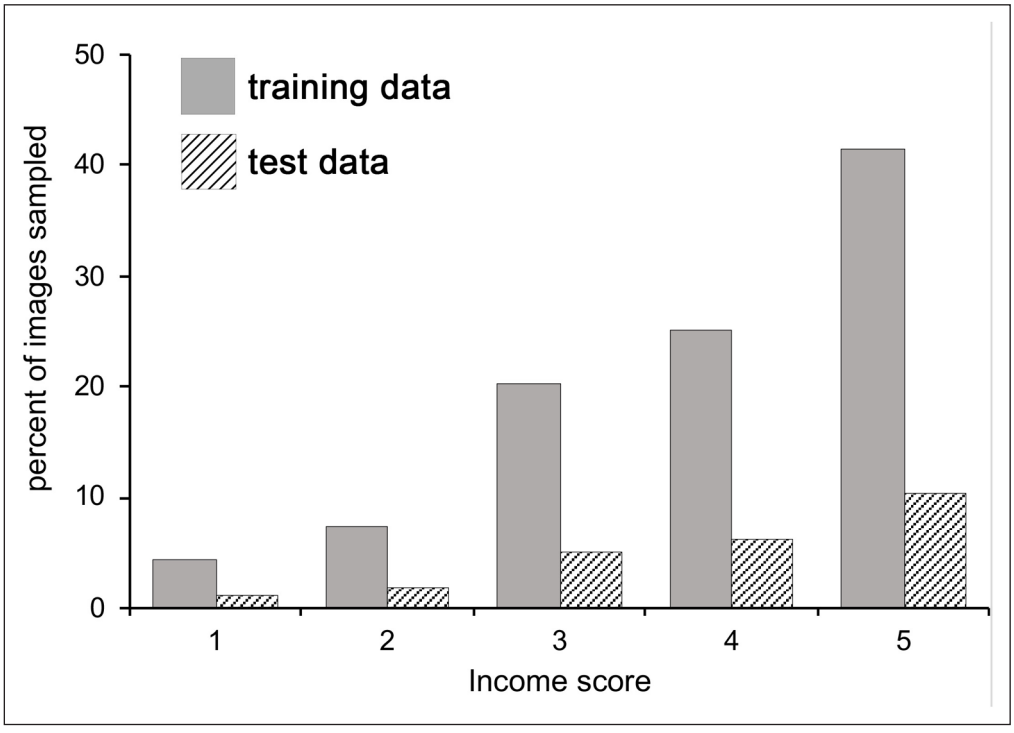

Figure 6 The distribution of images sampled (112,368 in total) for each income score for both the training and the testing datasets.

There was no clear division in income scores between census sectors and municipalities, so we treated the results as a heterogeneous distribution. On visual inspection, a slight division can be made between the given highest income areas in the centre and northeast of Vale do Ribeira, corresponding to the municipalities of the state of São Paulo, while the lowest and middle income scores are mainly located in the southwest of the study area (***Figure 7a***). The spatial representation of the predicted income score shows that some of the lowest income cluster regions were consistent with the observed data (***Figure 7b***). Since the test dataset for each census area often included a range of values, the mode was plotted, and if all values were the same, a random selection was made. Notably there are more predicted instances of the lowest income (score 1) than in the actual.

To complement the visual analysis in ***Figure 7***, we calculated the difference d between the real income score y and the predicted $\hat{y}$ value, defined as follows:

$$d = |y - \hat{y}| \tag{3}$$

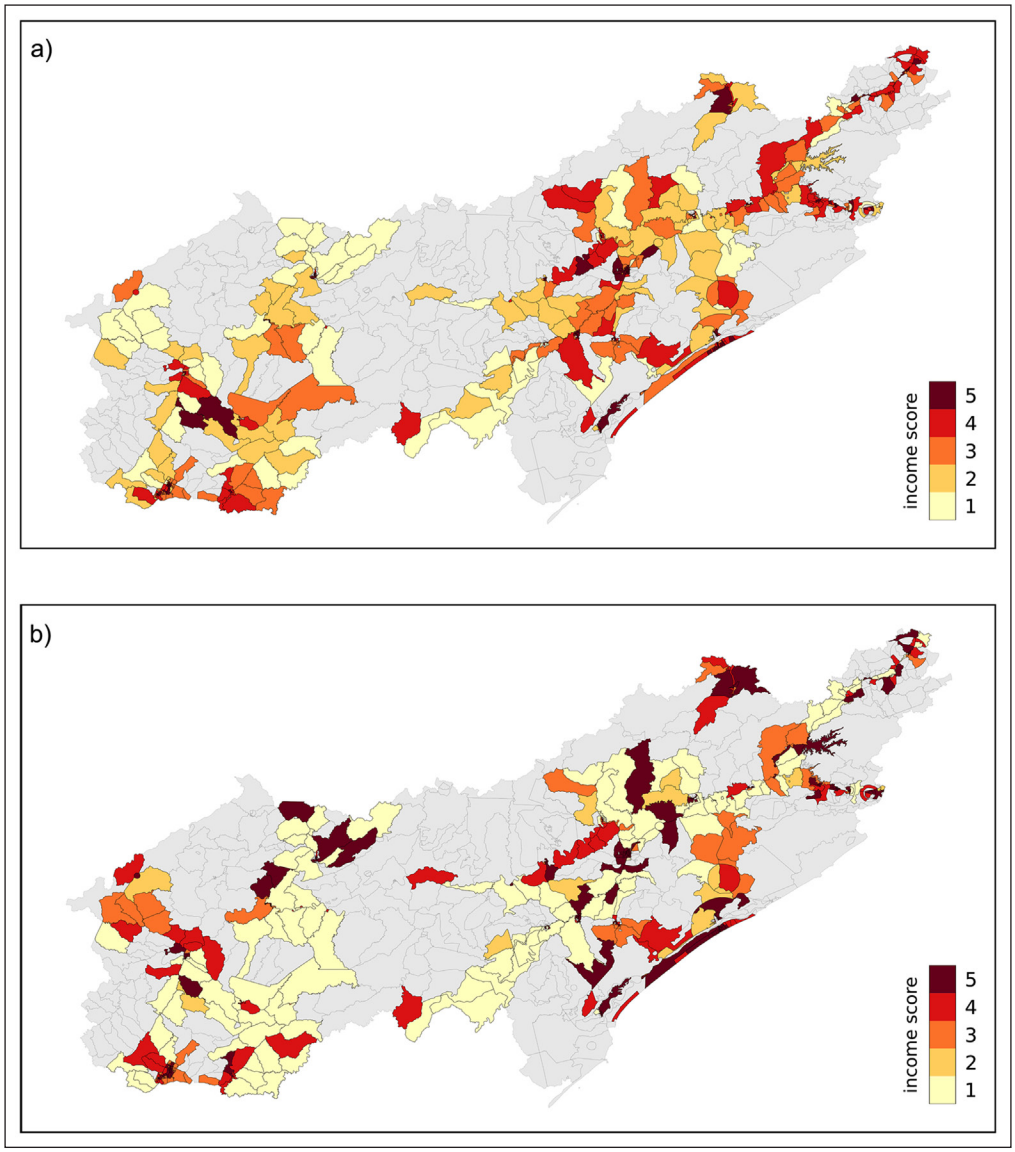

Figure 7 The distribution of income scores for the mode for each census sector is shown for (a) the observed (true) income score, and (b) the predicted income score. The scale ranges from 1 to 5, corresponding to the lowest to highest income score.

To better analyse how well our deep learning model predicts income, we plotted the difference between the actual and predicted incomes (***Figure 8***). To facilitate interpretation, we show the geolocations (red dots) for which a street view image was available (as in ***Figure 4***). It is clear that the census sectors with the worst prediction are correlated with a lack of street view images in these specific census sectors: the more data, the more knowledge the neural network can learn.

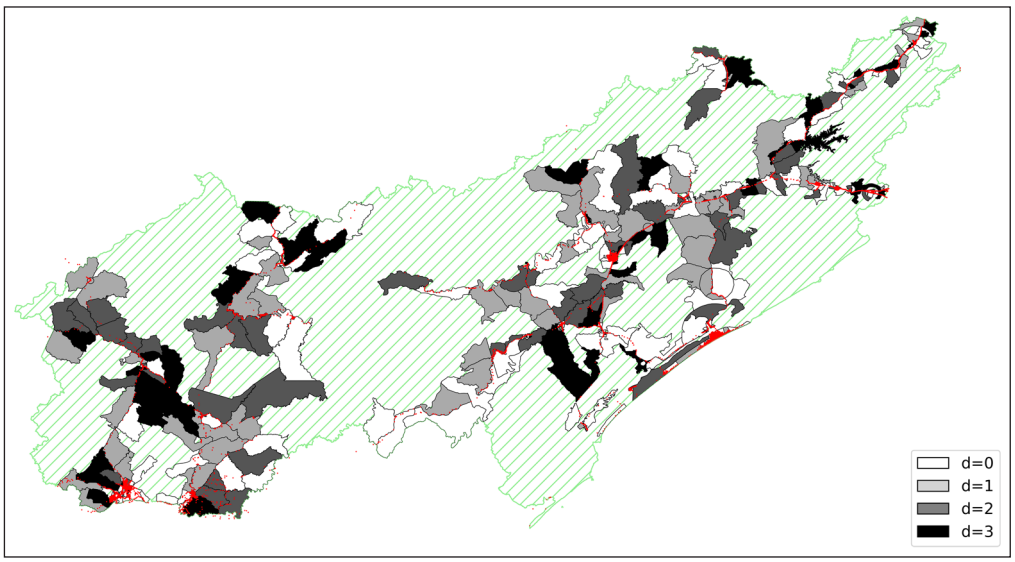

Figure 8 Differences between the real and predicted labels for each census sector of Vale do Ribeira. The grey scale shows the difference between the real and predicted labels (Eq. 3), with d = 3 being a big difference, and d = 0 a perfect match. The red dots represent geolocations for which street images were available. Census sectors for which street imagery was not available are shaded green.

To better understand the relationship between each predicted value and the number of images used to determine the final class we examined the difference between correct predictions (range from 0 to 3) without considering geographic location (***Figure 9***). Each dot represents one analysed census sector out of the 500 available. It is clear from this figure that the best predictions are for the high income level images (income score 5, d = 0). One reason for this result could be that the VGG-16 model decides from a collection of predictions which value to return as the value for a location (census sector) and the more images per location, the more robust the result as long as the information is balanced within a class. In contrast, the areas with lower income levels, with fewer images and a greater concentration on roads, produced poorer results. Another possibility is that the images in the high income class contain more objects that the machine learning protocol can detect, which increases the reliability of the predictions.

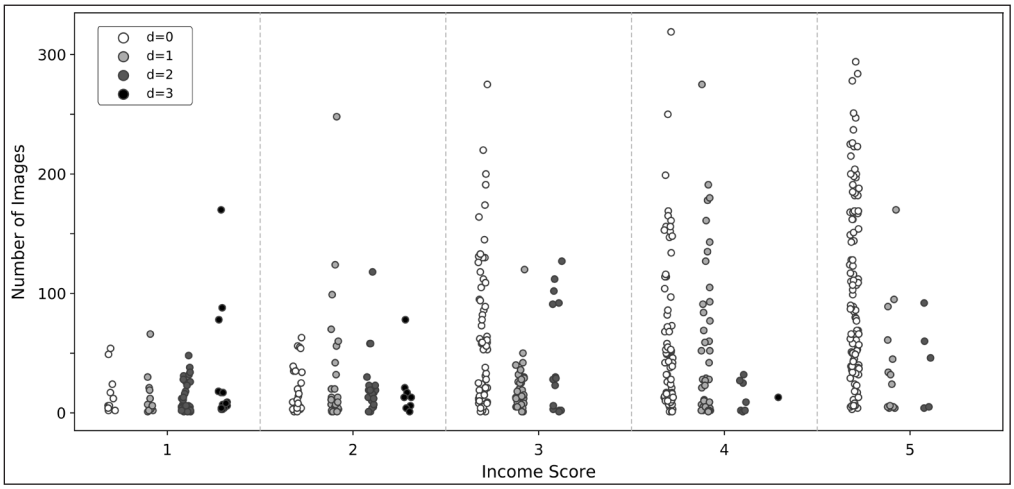

Figure 9 Distribution of predictions by the model indicating the prediction difference of the income score and the number of images per census sector (points). Each income score category includes the 500 census sectors available in Vale do Ribeira. The scale bar shows the difference from 0 (exact match) to 3 (poor match) between the real label and the predicted label (Eq. 3).

4.2 PERFORMANCE

The full performance results for the experiments on the test set for each fold show that the best prediction for income score classification is on average of 55% (with perfect discrimination between the five classes) and 80% (with an error margin of ±1), which was confirmed by the metrics mean absolute error (MAE), Pearson's correlation coefficient (r) and Kendall's Tau rank correlation coefficient (τ). The results were MAE = 0.21, r = 0.71, τ = 0.32 (***Table 3***).

	ACCURACY ERROR MARGIN (%)		MAE	*r*	τ
	±0	±1			
fold 0	53	78	0.21	0.74	0.30
fold 1	55	80	0.21	0.72	0.32
fold 2	56	80	0.20	0.71	0.32
fold 3	57	81	0.23	0.67	0.34
fold 4	56	81	0.22	0.69	0.33
Avg fold.	**55**	**80**	**0.21**	**0.71**	**0.32**

Table 3 Prediction results using the test set for each fold. Each column represents a different metric, from left to right the percentage of correctly predicted classes with an error margin of ± 0, ± 1, mean absolute error (MAE), Pearson's correlation coefficient (*r*), and Kendall's Tau rank correlation coefficient (τ).

The normalized confusion matrix for the classification tasks showed that the best performance was obtained for the highest income score of 5, which yielded 80% of the correct prediction cases (***Figure 10***). A perfect scenario would be represented by a diagonal line with 100% correct prediction. Prediction decreases smoothly as the score decreases. The income score 1 reached 40.2% of the corrected classification. We can observe that the misclassifications tend to decrease when the score values are increased.

4.3 COMPARISON WITH SUEL ET AL. 2019

The main difference between this study and that of Suel et al. (2019) is that the latter is a study of a highly urbanised area (with high and very low levels of income) whereas our work is in a

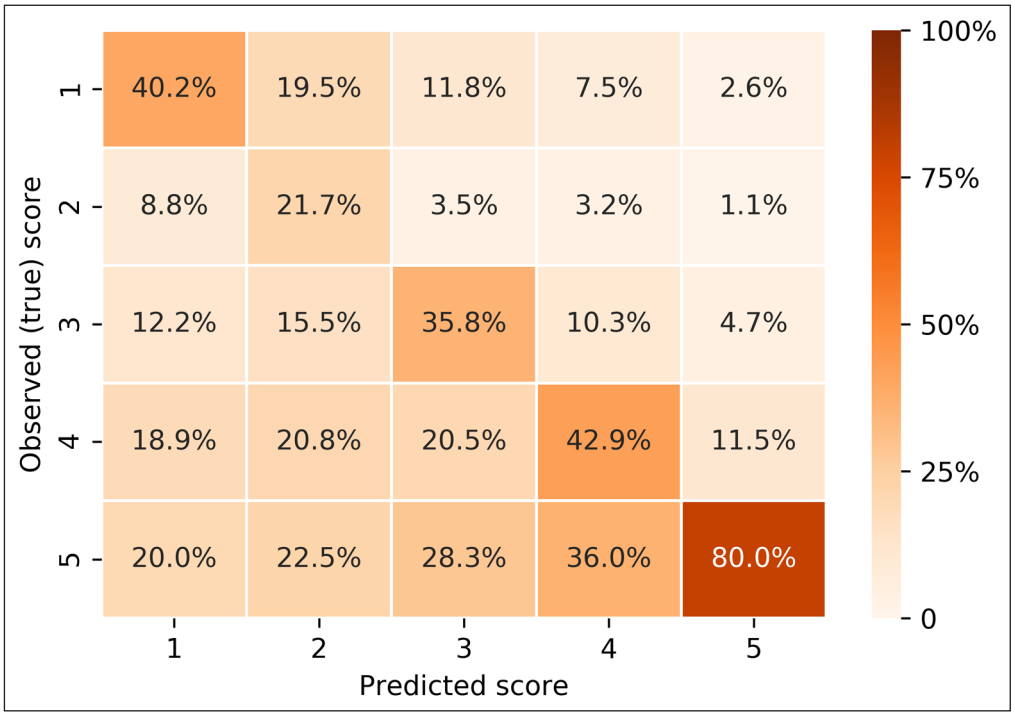

Figure 10 Confusion matrix between observed and predicted income scores. Each cell represents the percentage of the correct predictions. The best case would be a complete diagonal line with perfect accuracy.

semi-rural environment. Suel et al. (2019) examined several socio-economic indicators, and obtained good results for mean income. Comparing their performance using mean income with ours, they reported 32.7% as the percentage of correctly predicted classes while we obtained 55%. With respect to the percentage of correctly predicted classes with ±1 allowed error margin, Suel et al. (2019) reported 71.7%, while we obtained 80%. They reported MAE = 1.10, r = 0.86, τ = 0.72 while we obtained MAE = 0.21, r = 0.71, τ = 0.32. Therefore, our results (***Table 3***) achieved almost as high a performance as Suel et al. (2019), but these comparisons should be regarded with some caution because of the difference in the divisions between indicator classes (Suel et al., 2019 used deciles while we used quintiles).

5 DISCUSSION AND CONCLUSION

We started from the premise that street images contain rich visual information. In principle, the presence of vehicles, highways, infrastructure, houses, buildings or correlated objects in the image are associated with wealth, so a deep learning model should have the ability to learn from the data. Our results have shown that the deep learning framework used here is able to estimate high and low values of the income indicator using images of streets as input. We were successfully able to apply the model supplied by Suel et al. (2019) to the semi-rural situation of Vale do Ribeira, demonstrating the re-usability, reproducibility and replicability of GSV datasets. We paid particular attention to these factors, a common challenge to achieving FAIR outcomes (Wilkinson et al., 2016). The question arises whether the proposed method is generalizable to other types of indicators that may not be as translatable as income, such as longevity and literacy.

GSV is inherently dependent on the presence of roads, which are limited in this study area. This does not mean that the results are not representative of the situation as the road network reflects the location and density of the population. However, the semi-rural nature of the population in this area meant that there was little built infrastructure from which we could derive indicators. This limited our options compared to the wealth of indicators available in dense urban environments such as those studied by Suel et al. (2019) and Diou et al. (2018). An adjustment of the crawler algorithm presented by Suel et al. (2019) was necessary to obtain the images in the systematic manner illustrated in ***Figure 3*** to ensure optimal linkage with the census data.

The ubiquity of GSV worldwide, and its seemingly complete coverage can be deceptive. The images are not systematically collected or published, since it is the policy of the service to provide the most current image. To reduce the impact of this policy on the reproducibility and replicability of our work, we have published the libraries, datasets and images we used. This

results in images of different dates adjoining each other (in one case, an image from 2011 next to one from 2015). The exact information on each date is imprecise because the timestamp is not available through the GSV API. This problem was pointed out by Diou et al. (2018). In addition, the timestamps of the GSV (which ranged from 2011 to 2019) and the census data (2010) rarely matched. This is not an impediment if there have been no major changes in that time but can be misleading if this is the case. We noted that the image dates in this study could not be extracted by digital means, nor was it possible to do so manually given the numbers of images we were interested in. This was also noted by Diou et al. (2018). This potential source of uncertainty had to be accepted, and based on our knowledge of the stability of the communities in Vale do Ribeira during the study period, we were confident that this posed a minor risk to our results. This was an important factor in our ability to compare GSV imagery with census data (and thus train the method).

Although this work was located in a specific area (Vale do Ribeira), the deep learning model has shown great generalizability that could be replicated to other areas of the world (or other areas in Brazil), however the availability of socioeconomic indicators are strongly influenced by the specific situation and the model must be re-trained for each new indicator.

For those who want to repeat this project it worth noting that the acquisition stage is a time-consuming process requiring a stable internet connection and the costs can escalate quickly. Considering that, at the time of this study there was an initial credit of US$300 for use of the API service with the cost per 1,000 images being US$2.00. Above 100,000 requisitions the cost reduced to US$1.60 for every 1,000.

5.1 ETHICS IMPLICATIONS

The feasibility of this work relied on the public availability of street view imagery. However, despite the great utility of GSV imagery, its use can raise serious privacy concerns. GSV provides high-resolution 360° panoramic photos of streets, cities, mountains, and forests. Some images are taken from elevated positions, allowing viewers to see over hedges and walls designed to prevent certain areas from being accessible to the public. The large scale of the GSV approach combined with the extremely easy accessibility of the images increases the potential harmfulness of such information.

To balance the privacy and anonymity issues, GSV now blurs portions of the images that contain licence plates and human faces. However, the current level of blurring does not always prevent a person from being identified; moreover, other non-blurred information can also be indirectly identified. Thus, researchers should be careful with the data they make publicly available.

5.2 NEXT STEPS

Future projects to explore the following strategies would be profitable: (i) use of more complex neural network architectures; (ii) using more heterogeneous datasets in other locations; (iii) testing different training configurations, such as different regional granularities (e.g., municipalities instead of census sectors); (iv) training and validation with images from a range of regions could improve the generalization power of the image classifier; and (v) to update the study with the 2021 census (when it is available), which would improve the consistency with the period of coverage of the images.

Another aspect that could be enhanced by future work is the ability to improve the interpretation of the results of the deep learning model (Li et al. 2021; Abitbol et al. 2020; Dai et al. 2021). There are experiments, like those of Ayush et al. (2020), in generating metrics from the deep learning image interpretation, to correlate with the income indicator estimation. An approach by creating parallel metrics from the model to correlate with main indicator estimation could be a reasonable improvement to increase interpretation level from the model.

6 DATA AND SOFTWARE AVAILABILITY

The curated data (including the images and the income indicator dataset for Vale do Ribeira) and source code is available in a public and academic repository with the following formats and settings to ensure: (1) maximum protection of personal data, (2) compliance with all ethical

requirements of journals, and (3) an optimal level of openness to ensure maximum reuse of the data by potential third parties.

Our source code can be found on the GitHub Digital Repository: *https://github.com/PARSECworld/streetsValeRibeira* and released at *https://doi.org/10.5281/zenodo.4898335*.

The original Census data from the IBGE per census sector can be found at: *ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_do_Universo/Agregados_por_Setores_Censitarios/SP_Exceto_a_Capital_20190207.zip and ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Resultados_do_Universo/Agregados_por_Setores_Censitarios/PR_20171016.zip*.

We used the data from these files:

From files `ResponsavelRenda_PR.xls` and `ResponsavelRenda_SP2.xls`, we used the columns 'V022' named "Total do rendimento nominal mensal das pessoas responsáveis" (Total nominal monthly income for responsible householders).

From files `Basico_PR.xls` and `Basico_SP2.xls` we used the columns 'V002' named "Moradores em domicílios particulares permanentes ou população residente em domicílios particulares permanentes" ('Residents in permanent private households or population residing in permanent private households' in english).

ETHICS AND CONSENT

This paper complies with the Belmont Forum Data and Digital Output Management Plan requirements as stated specifically for the PARSEC project defined in Stall, et al (2020).

ACKNOWLEDGEMENTS

This research is a product of the PARSEC group funded by the Belmont Forum as part of its Collaborative Research Action (CRA) on Science-Driven e-Infrastructures Innovation (SEI) and the synthesis centre CESAB of the French Foundation for Research on Biodiversity. In Brazil the PARSEC project is supported by the grant 2018/24017–3, São Paulo Research Foundation (FAPESP). The authors wish to acknowledge Miguel S. X. Penteado and Nadya M. Deps Miguel from IBGE for their valuable suggestions and support.

J.M. is grateful for the support from FAPESP (grant 2020/03514–9).

R.D. was supported by the EOSC-Life European program (grant agreement No. 824087).

COMPETING INTERESTS

The authors have no competing interests to declare.

AUTHOR AFFILIATIONS

Jeaneth Machicao *orcid.org/0000-0002-1202-0194*
University of São Paulo, BR

Alison Specht *orcid.org/0000-0002-2623-0854*
Terrestrial Ecosystem Research Network, University of Queensland, AU

Danton Vellenich *orcid.org/0000-0002-3223-6996*
University of São Paulo, BR

Leandro Meneguzzi *orcid.org/0000-0002-4845-6758*
University of São Paulo, BR

Romain David *orcid.org/0000-0003-4073-7456*
ERINHA, FR

Shelley Stall *orcid.org/0000-0003-2926-8353*
American Geophysical Union, US

Katia Ferraz *orcid.org/0000-0002-7870-8696*
University of São Paulo, BR

Laurence Mabile orcid.org/0000-0002-7724-1721
University Paul Sabatier Toulouse III, FR
Margaret O'Brien orcid.org/0000-0002-1693-8322
University of California, US
Pedro Corrêa orcid.org/0000-0002-8743-4244
University of São Paulo, BR

REFERENCES

Abitbol, JL and **Karsai, M.** 2020. Interpretable socioeconomic status inference from aerial imagery through urban patterns. *Nat Mach Intell*, 2: 684–692. DOI: *https://doi.org/10.1038/s42256-020-00243-5*

Abreu, MVS, Oliveira, JC, de, Andrade, VDA and **Meira, AD.** 2011. Proposta metodológica para o cálculo e análise espacial do IDH intraurbano de Viçosa–MG. *Revista Brasileira de Estudos de População*, 28: 169–186. DOI: *https://doi.org/10.1590/S0102-30982011000100009*

Ayush, K, Uzkent, B, Burke, M, Lobell, D and **Ermon, S.** 2020. Generating Interpretable Poverty Maps using Object Detection in Satellite Images. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*. DOI: *https://doi.org/10.24963/ijcai.2020/608*

Bishop, CM. 2006. *Pattern recognition and machine learning, Information science and statistics*. New York: Springer. ISBN-10: 0-387-31073-8–ISBN-13: 978-0387-31073-2. DOI: *https://doi.org/10.1126/science.abe8628*

Burke, M, Driscoll, A, Lobell, DB and **Ermon, S.** 2021. Using satellite imagery to understand and promote sustainable development. *Science*, 371: eabe8628. DOI: *https://doi.org/10.1126/science.abe8628*

Bueno, GW, Leonardo, AFG, Machado, LP, Brande, MR, Godoy, EM and **David, FS.** 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. Epub July 06, 2020. DOI: *https://doi.org/10.1590/1678-4162-11389*

Dai, L, Zheng, C, Dong, Z, Yao, Y, Wang, R, Zhang, X, Ren, S, Zhang, J, Song, X and **Guan, Q.** 2021. Analyzing the correlation between visual space and residents' psychology in Wuhan, China using street-view images and deep-learning technique. *City and Environment Interactions*, 11: 100069. DOI: *https://doi.org/10.1016/j.cacint.2021.100069*

Dias, RL and **de Oliveira, RC.** 2015. Caracterização socioeconômica e mapeamento do uso e ocupação da terra do litoral sul do estado de São Paulo. *Sociedade & Natureza*, 27: 111–123. DOI: *https://doi.org/10.1590/1982-451320150108*

Diou, C, Lelekas, P and **Delopoulos, A.** 2018. Image-Based Surrogates of Socio-Economic Status in Urban Neighborhoods Using Deep Multiple Instance Learning. *J. Imaging*, 4: 125. DOI: *https://doi.org/10.3390/jimaging4110125*

Franch-Pardo, I, Napoletano, BM, Rosete-Verges, F and **Billa, L.** 2020. Spatial analysis and GIS in the study of COVID-19. A review. *Science of The Total Environment*, 739. DOI: *https://doi.org/10.1016/j.scitotenv.2020.140033*

Gebru, T, Krause, J, Wang, Y, Chen, D, Deng, J and **Fei-Fei, L.** 2017. *Fine-Grained Car Detection for Visual Census Estimation*. *https://arxiv.org/abs/1709.02480*.

Jean, N, Burke, M, Xie, M, Davis, WM, Lobell, DB and **Ermon, S.** 2016. Combining satellite imagery and machine learning to predict poverty. *Science*, 353: 790–794. DOI: *https://doi.org/10.1126/science.aaf7894*

Johnson, JM and **Khoshgoftaar, TM.** 2019. Survey on deep learning with class imbalance. *J Big Data*, 6: 27. DOI: *https://doi.org/10.1186/s40537-019-0192-5*

Li, Xuhong, Xiong, H, Li, Xingjian, Wu, X, Zhang, X, Liu, J, Bian, J and **Dou, D.** 2021. Interpretable Deep Learning: Interpretation, Interpretability, Trustworthiness, and Beyond. arXiv:2103.10689 [cs].

Mendes, Á, Louvison, MCP, Ianni, AMZ, Leite, MG, Feuerwerker, LCM, Tanaka, OY, Duarte, L, Weiller, JAB, Lara, NCC, Botelho, L de AM and **Almeida, CAL.** 2015. O processo de construção da gestão regional da saúde no estado de São Paulo: subsídios para a análise. *Saúde e Sociedade*, 24: 423–437. DOI: *https://doi.org/10.1590/S0104-12902015000200003*

National Academies of Sciences, Engineering, and Medicine. 2019. *Reproducibility and Replicability in Science*. Washington, DC: The National Academies Press. DOI: *https://doi.org/10.17226/25303*

Russakovsky, O, Deng, J, Su, H, Krause, J, Satheesh, S, Ma, S, Huang, Z, Karpathy, A, Khosla, A, Bernstein, M, Berg, AC and **Fei-Fei, L.** 2015. *ImageNet Large Scale Visual Recognition Challenge*. DOI: *https://doi.org/10.1007/s11263-015-0816-y*

Simonyan, K and **Zisserman, A.** 2015. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. *https://arxiv.org/abs/1409.1556*.

Stall, S, Specht, A, Corrêa, PLP, David, R, Edmunds, R, Mabile, L, ... and **Wyborn, L.** 2020. PARSEC Data and Digital Output Management Plan and Workbook. *Zenodo*. DOI: *http://doi.org/10.5281/zenodo.3891426*

Statistical Office of the European Union. 2017. Final report of the expert group on quality of life indicators: 2017 edition. Publications Office, LU. DOI: *https://doi.org/10.2785/021270*

Stiglitz, JE, Sen, A and **Fitoussi, J-P.** 2009. *Report by the Commission on the Measurement of Economic and Social Progress*. *https://www.economie.gouv.fr/files/finances/presse/dossiers_de_presse/090914mesure_perf_eco_progres_social/synthese_ang.pdf* (accessed 17th April 2021).

Suel, E, Polak, JW, Bennett, JE and **Ezzati, M.** 2019. Measuring social, environmental and health inequalities using deep learning and street imagery. *Sci Rep*, 9: 6229. DOI: *https://doi.org/10.1038/s41598-019-42036-w*

Tetila, EC, Machado, BB, Astolfi, G, de Souza Belete, NA, Amorim, WP, Roel, AR and **Pistori, H.** 2020. Detection and classification of soybean pests using deep learning with UAV images. *Computers and Electronics in Agriculture*, 179: 105836. DOI: *https://doi.org/10.1016/j.compag.2020.105836*

United Nations Department of Economic and Social Affairs. 2015. Sustainable Development Goals, THE 17 GOALS | Sustainable Development (accessed March 2021).

Wilkinson, MD, Dumontier, M, Aalbersberg, IjJ, Appleton, G, Axton, M, Baak, A, Blomberg, N, Boiten, J-W, da Silva Santos, LB, Bourne, PE, Bouwman, J, Brookes, AJ, Clark, T, Crosas, M, Dillo, I, Dumon, O, Edmunds, S, Evelo, CT, Finkers, R, Gonzalez-Beltran, A, Gray, AJG, Groth, P, Goble, C, Grethe, JS, Heringa, J, 't Hoen, PAC, Hooft, R, Kuhn, T, Kok, R, Kok, J, Lusher, SJ, Martone, ME, Mons, A, Packer, AL, Persson, B, Rocca-Serra, P, Roos, M, van Schaik, R, Sansone, S-A, Schultes, E, Sengstag, T, Slater, T, Strawn, G, Swertz, MA, Thompson, M, van der Lei, J, van Mulligen, E, Velterop, J, Waagmeester, A, Wittenburg, P, Wolstencroft, K, Zhao, J and **Mons, B.** 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*, 3: 160018. DOI: *https://doi.org/10.1038/sdata.2016.18*

Xie, M, Jean, N, Burke, M, Lobell, D and **Ermon, S.** 2016. Transfer Learning from Deep Features for Remote Sensing and Poverty Mapping. *https://arxiv.org/abs/1510.00098.*

Yosinski, J, Clune, J, Bengio, Y and **Lipson, H.** 2014. How transferable are features in deep neural networks? *https://arxiv.org/abs/1411.1792.*

Zhuang, F, Qi, Z, Duan, K, Xi, D, Zhu, Y, Zhu, H, Xiong, H and **He, Q.** 2021. A Comprehensive Survey on Transfer Learning. *Proc. IEEE*, 109: 43–76. DOI: *https://doi.org/10.1109/JPROC.2020.3004555*

TO CITE THIS ARTICLE:
Machicao, J, Specht, A, Vellenich, D, Meneguzzi, L, David, R, Stall, S, Ferraz, K, Mabile, L, O'Brien, M and 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: 6, pp. 1–15. DOI: *https://doi.org/10.5334/dsj-2022-006*

Submitted: 22 April 2021
Accepted: 29 January 2022
Published: 11 February 2022

Data Science Journal is a peer-reviewed open access journal published by Ubiquity Press.