Predicting socioeconomic indicators using transfer learning on imagery data: an application in Brazil

Diego A. Castro · Mauricio A. Álvarez

Abstract Censuses and other surveys responsible for gathering socioeconomic data are expensive and time consuming. For this reason, in poor and developing countries there often is a long gap between these surveys, which hinders the appropriate formulation of public policies as well as the development of researches. One possible approach to overcome this challenge for some socioeconomic indicators is to use satellite imagery to estimate these variables, although it is not possible to replace demographic census surveys completely due to its territorial coverage, level of disaggregation of information and large set of information. Even though using orbital images properly requires, at least, a basic remote sensing knowledge level, these images have the advantage of being commonly free and easy to access. In this paper, we use daytime and nighttime satellite imagery and apply a transfer learning technique to estimate average income, GDP per capita and a constructed water index at the city level in two Brazilian states, Bahia and Rio Grande do Sul. The transfer learning approach could explain up to 64% of the variation in city-level variables depending on the state and variable. Although data from different countries may be considerably different, results are consistent with the literature and encouraging as it is a first analysis of its kind for Brazil.

Keywords Socioeconomic indicators · Satellite imagery · Transfer learning · Machine learning

Introduction

Declining quality of life and reduced income level are critical problems around the world, especially in developing countries. According to estimates of Jolliffe et al. (2018), in 2015, approximately 736 million people (10% of the world’s population) lived in an extreme poverty condition, that is, with less than $1.90 a day. If a higher-value poverty line of $5.50 per person per day is considered, about 46% of the world’s population lived in poor conditions in 2015. In Brazil, the proportions of poor people are smaller, but still significant. Approximately 13.7 million people (6.5% of the Brazilian population) lived with less than $1.90 a day in 2018 and about 53.2 million people (25.3%) lived with less than $5.50 per day (Brazilian Institute of Geography and Statistics 2019).

Dealing with these issues is crucial and requires accurate and updated data for research and effective public policy design. However, as stated by Zhao et al. (2019), Pandey et al. (2018), Engstrom et al. (2017a) and Jean et al. (2016), censuses and other ground level surveys are expensive, time consuming...
and, consequently, carried out with significant time gaps between surveys. Besides, they cannot be conducted when there are health and safety concerns in a region, like wars and pandemics. To illustrate this, according to Brazilian Institute of Geography and Statistics (2010a), the 2010 Brazilian Census had a budget of R$ 1.677 billion (about $ 980 million using the exchange rate of 2010), which corresponded to approximately 0.04% of the Brazilian gross domestic product (GDP) for the same year. Also, although the Brazilian Census is decennial, data collection that would have happened in 2020 was postponed due to the Covid-19 pandemic.

As presented by Patino and Duque (2013), using the hypothesis that the interaction of human beings with the environment changes the surface appearance of a region and, consequently, can present some of their social and cultural behaviour, since the 1950s researchers have been trying to assess the relation between remotely sensed data and social features of cities. With the emergence of more modern satellites and statistical techniques, some authors have suggested the use of machine learning techniques and remote images to predict socioeconomic indicators and, consequently, to facilitate decision-making processes of those responsible for increasing the quality of life of a population.

Therefore, this work aims to combine deep learning techniques and satellite imagery to predict socioeconomic indicators in Brazil. More specifically, it uses diurnal and nocturnal satellite imagery to predict average income, GDP per capita and a constructed water index, applying a transfer learning method. The approach applied is similar to the Xie et al. (2016) and Jean et al. (2016) papers, but focusing on two Brazilian states, Bahia and Rio Grande do Sul.

It is important to mention that the criterion used to choose the Brazilian states of Bahia and Rio Grande do Sul was based in three aspects: number of cities, location and Gini coefficient. Since the predictive abilities of a machine learning model increase with the amount of data, we selected states with a minimum number of 400 cities for the analysis. Besides, as the focus of this work is to assess the capacity of using satellite images to predict some socioeconomic indicators, we decided to work with one state from the north and one from the south of Brazil due to geographic differences. Finally, from the states that fitted the previous requirements, one state with low Gini coefficient (Rio Grande do Sul) and one state with high Gini coefficient (Bahia) were chosen in order to verify whether or not there are differences in the quality of predictions for states with different inequality levels.

To our knowledge, this is the first work assessing the ability of transfer learning techniques to predict these indicators in Brazil and, thus, it has an important role of verifying the applicability of this method for this country. In order to have this verification done, transfer learning results are compared to simpler approaches, that is, using only nighttime lights (from now on, NTL) features, using only daytime features, and using features extracted directly from a pre-trained model to verify if the transfer learning technique outperforms these other methods. Also, a final model that join features from all approaches is tested to assess if it can improve the predictive performance.

The paper is organised as follows. “Relation between satellite imagery and socioeconomic indicators” section presents previous works about the relationship between satellite imagery and socioeconomic variables. In “Study area and dataset” section, the study area and the data used in this work are presented. “Methodology” section explains the details of the methods employed to predict socioeconomic indicators for Brazil using satellite images. Finally, the results are shown in “Results and discussion” section and conclusions are presented in “Conclusions” section.

Relation between satellite imagery and socioeconomic indicators

Satellite imagery has been extensively used for years to monitor weather, land cover and topography. Satellites from Landsat Program, Defense Meteorological Satellite Program (DMSP) and National Aeronautics and Space Administration (NASA)/National Oceanic and Atmospheric Administration (NOAA) Suomi National Polar-orbiting Partnership (Suomi-NPP) have provided images from the Earth since 1970s. However, during that time these images were not as

\[ \text{Gini coefficient is an index that ranges from 0 to 1, commonly used to measure inequality. Low values of the index express equality, whilst high values express inequality.} \]
easily accessible and machine learning techniques were not yet well developed.

New fields could be explored after satellite imagery became widely available, machine learning/statistics evolved and computational capacity increased. From the end of the 1990s and beginning of 2000s, researches have emerged in various areas, specially combining these images with socioeconomic variables. The link between images from space and socioeconomic indicators exists because, as stated by Patino and Duque (2013), remote sensing can be a measure of how people interact with the environment and land surface.

Satellite imagery, income and poverty

Many studies have analysed the interaction between satellite imagery and income and poverty data. Approaches differ regarding the type of satellite imagery: only nighttime lights (Elvidge et al. 2009; Yu et al. 2015); only daytime images (Engstrom et al. 2017a, b; Pandey et al. 2018; Bai et al. 2020); a combination of both (Xie et al. 2016; Jean et al. 2016; Perez et al. 2017; Zhao et al. 2019).

Using DMSP-OLS NTL, Elvidge et al. (2009) constructed a global poverty map. They did this by creating a poverty index (Landscan population count divided by the average visible band digital number from the lights) for each of the 233 countries in the sample and then regressing such index on the percentage of people living with $2 per day or less. They noticed that there is a high correlation between this index and poverty measures at national and subnational levels. More recently, Yu et al. (2015) used a similar approach to evaluate the relation between NASA/NOAA’s Suomi-NPP satellite with the Visible Infrared Imaging Radiometer Suite (VIIRS) radiance and poverty in China. Lights were represented by the radiance average for administrative regions. Poverty was measured as an index formed by the weighted average of 10 socioeconomic variables related to fields such as economic development, health and education. Lights and poverty index revealed a good correlation (coefficient of determination $R^2$—of 0.86), as well as lights and official national poor counties data.

Using only daytime imagery, two studies investigated how poverty in Sri Lanka can be explained by high spatial-resolution features. Engstrom et al. (2017a) extracted features such as number of cars, number and size of buildings, type of farmland, type of roofs, share of shadow pixels, road extent and road material from images captured by the sensors Worldview 2, GeoEye 1 and Quickbird 2. It was done by applying a combination of deep learning-based convolution neural networks (CNN) and classification of spectral and textural characteristics. After that, poverty and log welfare were estimated by a two-step approach, which first estimates a lasso model (Tibshirani 1996) over the full set of coefficients and then applies an ordinary least squares (OLS) model over the set of non-zero coefficients from the lasso step. The results show that daytime satellite features are highly correlated to poverty. The second study, developed by Engstrom et al. (2017b), extracted features from Quickbird images and applied a hybrid methodology, which used partial least squares, ridge and lasso methods for variable selection and OLS to fit the model. The linear regression results indicate that poverty can be partially explained by spatial and spectral features, since the $R^2$ for each estimated poverty level was slightly higher than 0.5. Focusing on the predictive power of daytime imagery regarding poverty in India, Pandey et al. (2018) obtained $1920 \times 1920$ sized images from Google Static Maps API. According to the authors, this was the first study to report deep learning experiments on images of such size. They employed a multi-task learning method which consists of applying a multi-task fully convolutional model to predict source of drinking water, source of lighting and material of roof and, then, used the output parameters of the first step to predict poverty (income levels). This method was able to estimate poverty with accuracy varying from 75 to 100%, depending on the model specification. Finally, Bai et al. (2020) estimated per-capita income and household income across different parts of New York City. They developed a Siamese-like Convolutional Neural Network, integrating Ridge Regression and Gaussian Process Regression, for predicting income which used house prices, daytime satellite images, the street view and spatial location informations as inputs of the model. The model achieved $R^2$ higher than 0.72, that is, it outperformed other state-of-the-art income estimation regression models.

Combining $400 \times 400$ pixels daytime images (downloaded from Google Static Maps API) and nighttime satellite images (DMSP-OLS) from Uganda, Xie et al. (2016) applied a transfer
learning method to extract poverty indicators from high-resolution satellite imagery. This method relies on the fact that training data is scarce and a CNN technique in general needs vast amount of data to generate accurate predictions. To deal with this issue, they first trained a CNN model to predict NTL intensity from daytime features and then extracted new features from the trained model and used them to map poverty applying logistic lasso regression. The results show that the transfer learning model could predict poverty well (accuracy of 0.72) and that it outperformed the other models tested, which used only ImageNet, only NTL or ImageNet combined with NTL. Extending this investigation, Jean et al. (2016) employed the same method, but this time for five African countries (Nigeria, Tanzania, Uganda, Malawi, and Rwanda) and using a ridge regression model instead of logistic regression for classification. Two poverty proxies were tested: consumption expenditure and asset wealth. The cross-validated coefficients of determination indicate that transfer learning features can explain from 37 to 55% of variations in consumption and 55 to 75% of variation in asset wealth. Also, when a model trained in one country is used to estimate consumption or assets in another country, predictive power declines only slightly, indicating its capacity of generalisation. Perez et al. (2017) and Zhao et al. (2019) also used transfer learning to predict poverty. The first work applied this method for the African continent, using DMSP-OLS and Landsat 7 imagery and several different models such as ridge regression and gradient-boosted trees. The results are aligned with Jean et al. (2016), with $R^2$ between 0.63 and 0.66. The second work employed transfer learning to NPP-VIIRS and Google Static Map images of Bangladesh. They also used features from other data sources, such as Land Cover Map and Open Street Map Road Map. The method consisted on putting all features together and using an algorithm to select the most important variables. After that, a random forest regression was trained to predict poverty. The $R^2$ between predicted and actual wealth index (poverty proxy) was 0.70. They also predicted poverty in Nepal using the model trained with data from Bangladesh. The $R^2$ in this case was 0.61.

Satellite imagery and general socioeconomic variables (GDP, population, urbanisation, quality of life)

Besides poverty and income, there are several works evaluating the relation between satellite imagery and other socioeconomic variables, such as GDP (Elvidge et al. 1997; Doll et al. 2006; Sutton et al. 2007; Tilottama et al. 2010; Henderson et al. 2012; Chen and Nordhaus 2019), population (Sandborn and Engstrom 2016; Engstrom et al. 2019), urbanisation (Gao et al. 2015; Zhou et al. 2015), quality of life (Elvidge et al. 2012), and other social variables, as educational attainment and child growth failure (Zimmerman et al. 2018; Graetz et al. 2018).

One intuitive approach is to assess the relation between nighttime lights intensity and GDP, once it is expected that regions with higher economic production radiate lights with more intensity due to a greater electricity consumption, for example. One of the first works that evaluated this relation was Elvidge et al. (1997), which used NTL data from the Defense Meteorological Program Operational Line-Scan System (DMSP-OLS) and GDP values for 21 countries. After using a regression analysis, they found a strong correlation between area lit and GDP. Henderson et al. (2012), in their turn, developed a statistical framework to show how variations in light intensity can be combined with GDP growth to generate an improved estimate of true income growth. They used countries with different developing levels and concluded that lights data can be a crucial variable in analysing economic growth.

Doll et al. (2006), Sutton et al. (2007) and Tilottama et al. (2010), for instance, used DMSP-OLS data to estimate how NTL is related to GDP at subnational level. Doll et al. (2006) observed strong correlations between total NTL and gross regional product (GRP) for United States and some European countries. Sutton et al. (2007) assessed the capacity of changes in light intensities to predict changes in GDP of India, China, Turkey, and the United States. They used two different approaches: one based only on the GDP data and another one that estimated an approximation of urban population for each state and used it as a proxy of GDP. The results showed that the method using urban population had higher $R^2$ for every country, which varied from 0.72 to 0.96. Finally, Tilottama et al. (2010) evaluated the relation between NTL
and estimated total (formal plus informal) economic activity for countries and states of the world. They used regression models to calibrate NTL values to economic measures at subnational level for China, India, Mexico, and the United States and at national level for other countries. The parameters obtained were used to estimate economic activities and the results showed that this approach provides an alternative means for measuring global economic activity.

As explained by Bennett and Smith (2017), DMSP-OLS data is only available until 2013 and, because of that, studies have arisen using the NASA/NOAA’s Suomi-NPP satellite with the Visible Infrared Imaging Radiometer Suite sensor. They showed that this satellite also provides data that can be used to accurately predict economic activity and other socioeconomic variables. An example is the article of Chen and Nordhaus (2019) that, using linear regression to evaluate the relationship between VIIRS light data and GDP of US states and metropolitan statistical areas, concluded that the nighttime lights-GDP correlation is stronger at the metropolitan areas and that cross-section predictions are more accurate than time-series predictions.

Regarding population, Sandborn and Engstrom (2016) used QuickBird-2 multispectral images from Ghana to assess their capacity of predicting census variables. They extracted five different spatial features (line support regions, PanTex, histograms of oriented gradients, local binary patterns, and Fourier transform) and developed a correlation coefficient matrix to evaluate their relationships with demographic data at the neighbourhood level. The variable that returned the strongest correlation with census variables, including population density, was the normalized difference vegetation index. Moreover, Engstrom et al. (2019) studied how feasible it is to combine satellite images and household surveys to track changes in Sri Lankan local population density. They used different satellite data sources, such as VIIRS night lights, Global Forest Change data, Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER)’s elevation and slope data, the Global Urban Footprint (GUF), Global Urban Footprint plus (GUF+), Global Human Settlement Layer’s (GHSL) and Facebook’s High-Resolution Settlement Layer (HRSL). The method employed was a cross-validated two-step regression, in which the first step consists on conducting a lasso regression for variable selection and the second step is a Poisson regression to estimate population density. The findings show that satellite indicators have exceptionally strong predictive power in predicting village level population density, and also that accuracies are similar when using a random forest model and when density estimates are expressed in terms of population counts.

Urbanisation, which has a close relation to population density, but focusing on urban population, was the object of study of many papers using lights and Asian countries’ data. Gao et al. (2015) analysed to which extent NTL can be used to infer urbanisation in China. They compared variations in a night light index to variations in proportion of the non-agricultural population and proportion of built-up area. The main finding was that changes in these variables were consistent for some regions, but not consistent for others, meaning that the rate of change of the relationship between urbanisation and NTL can differ across cities within a single country. With similar findings, but for different countries in Asia, Zhou et al. (2015) showed that the slope of the regression of NTL growth rate from 2000 to 2010 against logged population in 2000 differs considerably among countries. However, when there are not measurements of economic activity at the sub-national level, NTL data provide a useful proxy to analyse spatial patterns of urban growth.

Finally, satellite imagery was also used to predict other socioeconomic variables. Elvidge et al. (2012) developed from DMSP-OLS lights and population density a Night Light Development Index (NLDI) and tested its relationship with several variables for different countries around the world. On one hand, they did not find correlation between NLDI and Gini coefficient and found poor correlation between NLDI and the percentage of the population living in urban areas. On the other hand, NLDI presented strong relations with electrification rates, Human Development Index (HDI) and International Poverty Rate. Focusing on Africa, Graetz et al. (2018) estimated the average educational attainment by age and sex at subnational levels. By using Bayesian geostatistical methods in geolocated datasets, they demonstrated that attainment has generally improved for women of reproductive age in Africa since 2000 and, also, that there were relatively stable gaps over time between sexes. Zimmerman et al. (2018), also focusing on African countries and using Bayesian geostatistical methods,
aimed to map child growth failure. By 2015, nearly all locations showed improvements on this indicator compared to 2000, although some countries performed poorly in all child growth failure indicators.

Satellite images applied to Brazil

Looking specifically at Brazil, satellite images have been used mainly to investigate if they are able to detect slums (Hofmann et al. 2008; Nadalin and Mation 2018), estimate population (Amaral et al. 2005, 2006; Tomás et al. 2016; Neves et al. 2017; Maroko et al. 2019; Campos et al. 2020) and map Human Development Index (Charris et al. 2019).

Regarding detection of slums, Hofmann et al. (2008) developed an approach to obtain information about informal settlements from Quickbird satellite images. Basically, they employed a rule-based approach to generate a hierarchy of classes focusing on relevant properties according to the informal settlements ontology. Examples of classes are red roofs, small shadows/dark objects, bright small roofs/objects and vegetation, which act as indicators of informal settlements. The accuracy obtained using this method was 47% before applying an iterative process of a knowledge-based object enhancement and (re-)classification, and 68% after using the iterative approach. Nadalin and Mation (2018) provided estimates of slums locations in Brazil and their socio-economic characteristics using satellite images. From regression analysis, they found that there is no clear association between income and distance to the city centre for slums in Brazil and that steep areas near to rivers are more likely to have slums.

Amaral et al. (2005), using DMSP-OLS nighttime imagery, analysed lights-population and lights-power consumption correlation in the Brazilian Amazonia region. Their findings are consistent with the literature, since the results show that there are linear relations between lights and urban population ($R^2 = 0.79$) and between lights and electrical power consumption ($R^2 = 0.80$). Also using DMSP-OLS nighttime imagery and focusing on Brazilian Amazonia, Amaral et al. (2006) integrated NTL from 1995 and 1999 to population data from Brazilian censuses of 1996 and 2000 and obtained coefficients of determination higher than 0.8 from linear regressions. Besides, they estimated the population for 2003 using NTL data from 2002 and compared with the Brazilian Institute of Geography and Statistics (IBGE) population projection. The linear model used overestimated the population. Both works showed that DMSP-OLS nighttime light data can be used as an indicator of human activity in Amazonia.

Instead of using nighttime lights imagery, Tomás et al. (2016) presented an approach to estimate urban population based on the volume of single houses and high-rise buildings obtained from an IKONOS-2 ortho-image and light detection and ranging (lidar) data. The idea was to construct a 3D city model to quantitate urban population. Applying a linear regression to associate the estimated population to the population data released by IBGE for the study area (city of Uberlândia), the authors found a systematic underestimation of population. Neves et al. (2017) applied a dasymetric mapping approach, which is used to estimate the spatial distribution of a population, to estimate the variation of the distribution in the population in the Jacarepaguá Watershed. They used the land use and land cover map, obtained by a supervised classification of RapidEye sensor images, and population data from the 2010 Brazilian Census to create dasymetric zones and, then, calculate the estimated population of the dasymetric areas using two different methods. Also using a dasymetric approach, but in this case a 3D dasymetric mapping technique, Maroko et al. (2019) estimated population exposure to three-dimensional phenomena (e.g., air pollution), using a combination of cadastral data, building footprint data, and building height data. The idea was to disaggregate the population into individual buildings in residential areas and then distribute it vertically based on building height. This method was tested in São Paulo and New York City and the results showed that it can be useful for purposes such as improving emergency management operations or selecting optimal sites for emergency shelter locations. Using orbital images from Landsat ETM+ and census data, Campos et al. (2020) evaluated the capacity of these images to estimate population for the municipality of Contagem in 2000, 2010 and 2015. Regression models were constructed, in which population is the dependent variable and the reflectance of the bands of the Landsat images and some additional variables (e.g., whether the area is urban, rural or slums) are the explanatory variables. The estimates for 2010 calculated based on 2000 data were compared with the population listed in the 2010 census and estimates
for 2015, using data from 2000 and 2010, were compared with other post-census estimates. The results show that, although the models tested presented good estimates, in general, the errors generated by these models were larger than the ones obtained by demographic estimates.

Finally, Charris et al. (2019) assessed the relationship between DMSP-OLS nighttime lights imagery and the Human Development Index (HDI) across Brazilian municipalities among 1991 and 2010. From a panel data regression with fixed effects, the authors encountered the expected relation between those variables, that is, municipalities with higher nighttime lights intensity have better HDI. They also investigated the relationship between NTL and variables associated with HDI as schooling outcomes, infant mortality and number of plants in the municipalities and found that more illuminated areas are related to better results for the variables examined. All these findings suggest that there is evidence that NTL may be a good proxy for socioeconomic indicators in Brazil.

Study area and dataset

In this section we present how and which Brazilian states were chosen to be assessed in this paper and the sources of data used to generate the predictions.

Brazilian states overview

The selection of Brazilian states was based on the number of cities in the state (as the predictive abilities of a machine learning model increase with the amount of data), on their Gini coefficients (to verify the model ability to predict the chosen socioeconomic indicators for states with different income inequality levels) and on the state location inside Brazil (states from the north and the south of Brazil, in general, have different characteristics regarding, for example, vegetation and climate). In practical terms, only four states have more than 400 cities: Minas Gerais, São Paulo, Rio Grande do Sul and Bahia. From these states, Bahia is the one with the highest Gini coefficient and Rio Grande do Sul is the one with the lowest. Besides, both states are in different Brazilian regions, Bahia in the Northeast and Rio Grande do Sul in the South (Fig. 1). Therefore, as Bahia and Rio Grande do Sul have more than 400 cities and are considerably distinct both regarding income inequality levels and geographic aspects, they were chosen to be examined in this paper.

According to the Brazilian Institute of Geography and Statistics (2018a), Bahia has an area of 564,722 km$^2$ (the 5th biggest Brazilian state out of 26 states and one federal district) made of 417 cities with a total population of 14 million people in 2010. It is located at the North-eastern Region and its territory is characterised by highlands with caatinga.2 The average per capita household income in 2018 was R$ 841.00 per month, which is 12% lower than the minimum monthly wage in Brazil for the same year. The 2017 Gini index of Bahia was 0.599, the 3rd highest considering all Brazilian states and the Federal District.

Rio Grande do Sul has an area of 281,707 km$^2$ (the 9th biggest Brazilian state) formed by 497 cities (during the last census in 2010, the number of cities was 496) with total population of 10.7 million people in 2010. It is located at the Southern Region and its territory is characterised by a diversified geographical relief, with highlands, depressions and plains. Besides, differently from Bahia, the vegetation in Rio Grande do Sul is more dense and tolerant to lower temperatures. The average per capita household income in 2018 was R$ 1705.00 per month, which is 79% greater than the minimum monthly wage in Brazil for the same year. The 2017 Gini index of Rio Grande do Sul was 0.492, the 6th lowest considering all Brazilian states and the Federal District (Brazilian Institute of Geography and Statistics 2018b).

Shapefiles

Shapefile is a file format that stores georeferenced information. More specifically, it provides geometric locations and attributes of geographic features as, for example, points, lines and polygons.

This data type is required in this work because nighttime lights are available for broad areas (as explained in the next subsection), but we only need

---

2 Caatinga is a type of desert vegetation often found in northeastern Brazil, where the climate is hot with long dry seasons. In this biome, it is common to find leafless trees, as they need to resist droughts.

data for Brazil. So, the shapefiles are used to filter NTL data inside Brazil, also identifying the points related to the states of Bahia and Rio Grande do Sul. Three shapefiles covering different regions were necessary: Brazil, to get all the lights inside the country to define ranges of radiance and classify areas into low, medium and high radiance; Bahia and Rio Grande do Sul to get all the lights inside the states, allocating them into cities.

The shapefiles used are provided by Brazilian Institute of Geography and Statistics (2010b) in versions for different years. This report uses 2010 version, since it corresponds to the last Brazilian census year. The city of Pinto Bandeira in Rio Grande do Sul was created after 2010 and, therefore, it is not included in the analysis.

Satellite imagery

*VIIRS day/night band nighttime lights*

As explained in the previous section, one of the most common NTL data used to estimate socioeconomic variables was the DMSP-OLS. However, this satellite was discontinued in 2013 and, consequently, other sources started being used.

Even some years after 2013, articles still used DMSP-OLS, claiming that the data used as a natural
substitute, the NPP-VIIRS Day/Night Band (which was launched in 2011), was not yet aggregated annually or filtered to screen out temporal lights from aurora, fires, boats, and other transient sources, making it noisier and potentially biased (Jean et al. 2016). Nowadays VIIRS Day/Night Band (DNB) data is available on an annual basis and any data impacted by stray light is excluded. Also, it is possible to obtain cloud-free average radiance values, which passed through an outlier removal process to filter out fires and other ephemeral lights (National Oceanic and Atmospheric Administration 2019). VIIRS DNB data is freely available on NOAA’s website and, in contrast to the 1 km spatial resolution of DMSP-OLS (Ghosh et al. 2013), the spatial resolution of VIIRS DNB data is about 750 m. The radiance unit of this satellite band is nanoWatts/(cm²/sr) and the more recent year for the annual NTL data is 2016. The products are made available in geotiff format as a set of six tiles.

Following Jean et al. (2016) and Xie et al. (2016), we use 100 night light points for each city. To choose these points for a specific city, first we identified all NTL points inside its boundaries. Second, we calculated the Euclidean distance of each NTL point to the city centre. Finally, we filtered the 100 NTL points with the smallest Euclidean distances, that is, the night lights which are the closest points to the city centre.

To sum up, this study uses annual VIIRS DNB data from 2016, filtered from impacts of stray light, clouds, fires and other ephemeral lights. Figure 2 presents an example of part of a tile that covers part of South America and Africa.

Daytime imagery

Besides nighttime images, high resolution daytime satellite images were also used. Google Maps Static API³ was used to download 400 × 400 pixel images at zoom level 16, which correspond to approximately 740–810 m spatial resolution depending on the latitude of the image. As daytime images should be related to nighttime images (see “Methodology” section), their resolutions are similar enough for this task. For the same reason, the images downloaded correspond exactly to the same geographic coordinates of NTL points selected, that is, for each city of Bahia and Rio Grande do Sul, 100 daytime images were downloaded. As there are 417 cities in Bahia and, during the 2010 Census, 496 cities in Rio Grande do Sul, this leads to a total of 91,300 daytime images.

Google Maps Static API does not allow the user to choose the year in which the satellite picture was taken. However, there is a mark on every image that shows such year (Fig. 3). The images used in these paper are from 2019. Figure 3 shows examples of daytime images. From left to right images are sorted according to their night light intensity. Images on the top are from Rio Grande do Sul and images on the bottom are from Bahia.

Census and survey data

This subsection presents the three city-level socioeconomic variables used in this paper: average income, GDP per capita and water index.

³ The API documentation can be found at https://developers.google.com/maps/documentation/maps-static/overview.
Regarding the average income of the residents of a city, in the last Brazilian Census (2010) published by the IBGE, there is one indicator that is similar to this variable: the average monthly nominal income of people that are 10 years old or older per city (in Brazilian Real). Considering all cities in Brazil, the median value for average monthly nominal income was R$ 537.16, the average value was R$ 566.15 and the standard deviation was R$ 265.77; including only cities from Bahia, the median was R$ 320.85, the average was R$ 350.13 and the standard deviation was R$ 115.80; for Rio Grande do Sul, the median was R$ 795.00, the average was R$ 826.11 and the standard deviation was R$ 223.70.

It is important to highlight that, although the average income variable is from 2010 and the NTL data and daytime images are from 2016 and 2019 respectively, there is evidence that this indicator for a city is stable over time relatively to other cities. FGV Social (2020) estimated the income per inhabitant for each Brazilian city based on the 2018 earning declarations in the income tax registries. The correlation between these estimates and the average income from the 2010 Census is 86%. Besides, the 2019 Continuous National Household Sample Survey from IBGE presents the average monthly earnings of people that are 14 years old or older from 2012 to 2019 by state. The correlations between the average earnings in 2012 and the average earnings in 2016 and 2019 are, respectively, 97% and 95%. Even though there was a severe economic crisis in Brazil in 2015 and 2016, the data shows that it did not affect significantly the relative income of cities and states which, therefore, supports the use of orbital images from 2016 and 2019 to estimate the average income obtained in the 2010 Census.

Another socioeconomic variable, for which there is data from 2016, was assessed: GDP per capita. The source of this data is the IBGE in partnership with the State Statistical Agencies and State Secretariats. Considering all Brazilian cities, the median, the average and the standard deviation for GDP per capita (in thousands and, by definition, annual) were R$ 15,869.60, R$ 21,126.18 and R$
20,332.65 respectively. Considering only cities from Bahia, the median was R$ 8,280.00, the average was R$ 11,144.00 and the standard deviation was R$ 15,970.67. In the case of Rio Grande do Sul, median, average and standard deviation were R$ 28,607.00, R$ 33,725.22 and R$ 21,074.26 respectively.

Finally, the third variable is a water index,\(^4\) which is a simple average of three different indicators from 2017 extracted on the Sistema Nacional de Informações sobre Saneamento (SNIS)\(^5\) (Ministry of Regional Development of Brazil 2019): percentage of the city’s population that has access to water services; percentage of water that is not wasted during the distribution; percentage of water samples analysed that fall inside the normal pattern. The SNIS provides indicators regarding water and sanitation, but many of them have a high quantity of missing values and, for this reason, they were not selected. Furthermore, approximately 10% of the cities of Bahia and Rio Grande do Sul do not have data about one of the indicators of the water index and were dropped, leading to a total of 395 cities for Bahia and 453 for Rio Grande do Sul.

Regarding the water index values, which range from 0 to 1, the median for Brazil was 0.67, the average was 0.66 and the standard deviation was 0.10. For Bahia, the median, average and standard deviation were 0.64, 0.64 and 0.09 respectively. Finally, the median, average and standard deviation for Rio Grande do Sul were 0.65, 0.64 and 0.11.

**Methodology**

The main aim of this work is to predict socioeconomic indicators in Brazil applying a transfer learning approach on satellite images. To verify the effectiveness of this technique, it will be compared to other approaches that use different type of features: only nighttime lights features; only basic daytime features; and daytime features extracted from a VGG16 model (this model is presented in “VGG16 features” section). A final approach that includes features types in the same model is also compared to transfer learning to investigate whether it can improve prediction performance or not. This section presents the complete methodology used in this article.

**Features extraction**

**Nighttime lights features**

As mentioned above, different types of features can be extracted from satellite images. The first option assessed in this work is the use of NTL radiance. The idea is that nighttime lights are correlated with economic activities, as industrial activities for example, and urban areas, with more modern infrastructure than in rural areas, tend to register more intense light.

In the literature, authors often use the average NTL radiance inside a region as a feature to predict socioeconomic variables (Elvidge et al. 2009; Yu et al. 2015), as well as the sum of the lights intensities (Ghosh et al. 2013). Here, five NTL features are extracted for each city: average radiance, median radiance, minimum radiance, maximum radiance and standard deviation of the radiance. These features are used directly as covariates in one of the regressions specifications to predict the socioeconomic variables.

**Daytime basic features**

The second type of features are based only on daytime images. Each image can be represented by a 3-dimensional matrix in which the vertical and horizontal axes denote the pixel position in the image and each plane representing the depth of the matrix is a colour of the RGB colour model,\(^6\) that is, red, green and blue. Five features for each colour in the RGB model (15 features in total) are extracted: average colour value, median colour value, minimum colour value, maximum colour value and standard deviation of the colour values. As well as the nighttime lights features, the daytime basic features are the covariates in one of the regressions specifications.

---

\(^4\) Other authors, as Yu et al. (2015) and Jean et al. (2016) for example, used indices to represent poverty.

\(^5\) This is a system from the Brazilian Ministry of Regional Development which provides data about water, sewage and solid waste.

\(^6\) The RGB colour model is a scheme used to represent colours by adding together red, green and blue lights to reproduce many different colours.
VGG16 features

VGG16 features are obtained from a convolutional neural network. Convolutional neural networks have been used for image classification since the end of 1980s and were proposed by Lecun (1989). This approach fits well for tasks related to images because it considers one of their key properties: the correlation of pixels is inversely proportional to the distance between them (Bishop 2006). Basically, the idea is that local features, which can be extracted from subregions of the images and are able to identify visual features such as edges and corners, can be merged and processed to detect higher-order features and, in the end, to classify the whole image (LeCun et al. 1999). In general, as stated by Rawat and Wang (2017), CNN architectures are formed by convolutional and subsampling (pooling) layers which feed fully connected layers responsible for classifying the input. One important aspect of convolutional layers is that they are formed by feature maps in which every unit receives the same weight. However, different feature maps have different weights, enabling the extraction of different features at each location. The outputs of these layers are sent to subsampling layers through a nonlinear activation function. The role of pooling layers is to reduce the spatial precision of features extracted, decreasing the importance of the feature location (and distortions) within an image in the classification process.

The VGG is a CNN architecture developed by Simonyan and Zisserman (2015) to participate on the ImageNet Large Scale Visual Recognition Challenge 2014, which got the first and the second places in the localisation and classification tracks respectively. This model increases the depth of traditional convolutional neural networks using an architecture with $3 \times 3$ convolutional layers.

The VGG model used in this article is the VGG16, which has 16 weight layers. Figure 4 shows the architecture of the model. It is formed by five blocks of convolutional layers followed by a pooling layer (in this case, max pooling\(^7\)) and one block of fully connected layers with a soft-max function\(^8\) in the end to classify the input in one of the classes.

VGG models can be directly applied in Python and R by using the library keras. VGG16 pre-trained on ImageNet is available in this package. In this work, we use the VGG model to extract features from images and then we use these features in the regression used to predict the socioeconomic indicators. Feature extraction is done by dropping the last

---

\(^7\) The role of pooling layers is to reduce the spatial precision of features extracted. Regarding max pooling, it samples the output of the previous layer by calculating the maximum value in each patch of each feature map, which highlight the most present feature in the patch.

\(^8\) Soft-max is a function that normalises the values of the output from the previous layer into a probability distribution with a probability for each possible class in the analysis.
block of the VGG16 model (fully connected layers—Fig. 4), in a way that the output is the pattern (features) identified by the convolutional layers. As there are 100 images per city, features obtained for all images of one city are averaged to generate an average feature vector representing this city.

Transfer learning

CNN models are generally more efficient in supervised learning with abundant labeled training data (Jean et al. 2016). In general this is not the case when predicting socioeconomic variables, where data is scarce. Transfer learning is a possible method for dealing with this issue and it was used before by Xie et al. (2016), Jean et al. (2016), Perez et al. (2017) and Pandey et al. (2018) to predict socioeconomic indicators.

The idea of transfer learning is that pre-trained models can be used to help extracting image features which can be the input for a different CNN classification task. Normally, a CNN model that was trained on an image database (such as ImageNet) to predict dogs and cats, for example, and has learned how to identify low-level image features is used to fine-tune a second CNN model that aims to classify another type of images, such as cars and motorcycles.

In this paper, the pre-trained CNN used is the VGG16. So, after extracting images features from this model, they are used to train another CNN which predicts NTL radiance class based on daytime images features. Some authors (Jean et al. 2016; Perez et al. 2017; Zhao et al. 2019) fine-tuned a CNN model named VGG-F (Chatfield et al. 2014). Nonetheless, this work tested a similar, but simpler, architecture (Fig. 5), which is formed by two blocks with a convolutional layer followed by a dropout layer, one block with a convolutional layer and an average pooling layer, and the final block with a fully connected layer that uses a soft-max activation function to classify the input into one radiance class.

Regarding the NTL classes, three classes of radiance were defined: low, medium and high. The upper and lower bounds of each class were determined by fitting a Gaussian mixture model (GMM) to the relative frequencies of the NTL intensities across all Brazilian cities. As shown before, in this work we use for each city the 100 geographic points closest to the city centre. The same criterion was applied to choose the NTL points used in the Gaussian mixtures model.

The final classes ranges obtained were: less than or equal to 0.87 nW/(cm²/sr), for low intensity; greater than 0.87 and less than or equal to 10.37 nW/(cm²/sr), for medium intensity; and greater than 10.37 nW/(cm²/sr), for high intensity.
sr), for medium intensity; and greater than 10.37 nW/(cm²/sr), for high intensity.

After extracting images features from the pre-trained VGG16 CNN and fitting a second CNN model to classify NTL into low, medium and high radiance intensity by using the features extracted from VGG16, the transfer learning model is built. This process is possible because each nighttime light pixel corresponds to the same geographic location and similar area of the daytime satellite image. Therefore, the transfer learning model can be used to extract features from images in the same way mentioned in the VGG16 subsection, that is, by dropping the last layer of the network. As there are 100 images representing each city, there will also be 100 feature vectors for each city. To have only one set of features for each city, we use the average feature vector, that is, the set of features generated by taking the average value of each element of all the 100 feature vectors. Figure 6 presents the whole transfer learning process.

Regression analysis

After extracting features using the methods explained above, a data set with images features and their respective socioeconomic indicators values is built. Regression methods are applied in order to fit a model that can estimate the socioeconomic variables.

As there are a relatively high number of covariates in comparison to observations, in order to avoid overfitting, we employed a regularised linear regression technique in this article. Basically, this is done by inserting a term R(β) to the objective function of a linear regression (Murphy 2013), in the following way:

\[
\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2 + \lambda R(\beta)
\]

\[
= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_k x_{ik})^2 + \lambda \left( \alpha \sum_{j=1}^{k} |\beta_j| + (1 - \alpha) \sum_{j=1}^{k} \beta_j^2 \right),
\]

where \(y_i\) is the dependent variable, \(x_{ik}\) are the independent variables and \(\beta_k\) are the coefficients. \(\lambda\) is the penalty parameter, which means that when \(\lambda = 0\), the term \(R(\beta)\) has no effect, and when \(\lambda\) increases, the impact of \(R(\beta)\) also increases. \(\alpha\) is the parameter that defines which type of regularisation is used: ridge regression, when \(\alpha = 0\); lasso regression, when \(\alpha = 1\); and elastic net (a combination of both regularisers), when \(0 < \alpha < 1\).

A nested cross-validation was used to first define the value of \(\lambda\) and \(\alpha\) that yield the lowest Root Mean Square Error (RMSE) and then, with the chosen \(\lambda\) and \(\alpha\), fit the model and evaluate it. The values of \(\lambda\) tested were 0.001, 0.01, 0.1, 1, 10 and 100; the values of \(\alpha\) were 0, 0.5 and 1. Moreover, the inner cross-validation (the one responsible for defining the regularisation parameters) is a five-fold cross-validation. The outer cross-validation, which is responsible for fitting the model and evaluating predictions, is a 10-fold cross-validation.
All specifications used standardised predictors, that is, observations of each feature are subtracted by the feature’s mean and divided by its standard deviation. Also, the dependent variable (socioeconomic indicator) was examined in natural logarithm form.

Results and discussion

The following subsections present the results for each socioeconomic indicator for Bahia and Rio Grande do Sul. As explained in the previous section, although the main technique to extract features used in the regressions is the transfer learning approach, three other type of features were assessed. A specification gathering all features as covariates of the regressions is also evaluated. In Table 1, which exhibit the evaluation metrics for all specifications, lights refers to the features explained in “Nighttime lights features” section; images basic features to the features presented in “Daytime basic features” section; images VGG to the features described in “VGG16 features” section; transfer Learning to the features extracted by using the technique showed in “Transfer learning” section; and all features to the specification mentioned above that gathers all features as covariates.

Table 1  Evaluation metrics for Bahia and Rio Grande do Sul considering the five different types of features

| Features                  | Average income | GDP per capita | Water Index |
|---------------------------|----------------|---------------|-------------|
|                           | R²  | RMSE | MAE | R²  | RMSE | MAE | R²  | RMSE | MAE |
| **Bahia**                 |     |      |     |     |      |     |     |      |     |
| All features              | 0.650 | 68.44 | 48.43 | 0.290 | 10.08 | 3.72 | 0.108 | 0.07 | 0.06 |
| (0.151)                   | (18.14) | (7.55) | (0.233) | (12.24) | (2.26) | (0.094) | (0.01) | (0.01) |
| Transfer learning         | 0.643 | 71.08 | 48.48 | **0.291** | 10.19 | **3.69** | 0.103 | 0.07 | 0.06 |
| (0.117)                   | (19.04) | (8.14) | (0.186) | (12.21) | (2.29) | (0.100) | (0.01) | (0.01) |
| Images VGG                | 0.616 | 74.61 | 52.04 | 0.265 | 10.24 | 3.71 | 0.101 | 0.07 | 0.06 |
| (0.185)                   | (15.39) | (6.56) | (0.224) | (12.19) | (2.16) | (0.081) | (0.01) | (0.01) |
| Images basic features     | 0.334 | 94.08 | 64.50 | 0.143 | 10.72 | 4.15 | 0.047 | 0.07 | 0.06 |
| (0.203)                   | (21.06) | (10.12) | (0.138) | (12.14) | (2.26) | (0.039) | (0.01) | (0.01) |
| Lights                    | 0.564 | 78.65 | 55.33 | 0.231 | 10.88 | 4.11 | 0.090 | 0.07 | 0.06 |
| (0.207)                   | (17.70) | (11.11) | (0.222) | (12.10) | (2.24) | (0.077) | (0.01) | (0.01) |
| **Rio Grande do Sul**     |     |      |     |     |      |     |     |      |     |
| All features              | 0.409 | 174.63 | 129.51 | **0.248** | 17.22 | 9.73 | **0.114** | 0.10 | 0.08 |
| (0.103)                   | (35.03) | (18.39) | (0.185) | (10.45) | (2.80) | (0.099) | (0.01) | (0.01) |
| Transfer learning         | 0.335 | 183.93 | 140.10 | 0.242 | **17.16** | **9.67** | 0.100 | 0.10 | 0.08 |
| (0.095)                   | (32.92) | (17.32) | (0.184) | (10.54) | (2.83) | (0.122) | (0.02) | (0.01) |
| Images VGG                | 0.404 | 174.32 | 129.00 | 0.225 | 17.35 | 9.95 | **0.137** | 0.10 | 0.08 |
| (0.108)                   | (37.44) | (19.98) | (0.176) | (10.51) | (2.91) | (0.125) | (0.02) | (0.01) |
| Images basic features     | 0.165 | 207.15 | 161.78 | 0.045 | 18.86 | 11.61 | 0.065 | 0.10 | 0.08 |
| (0.100)                   | (32.92) | (21.09) | (0.049) | (10.04) | (3.14) | (0.073) | (0.01) | (0.01) |
| Lights                    | 0.140 | 210.56 | 169.00 | 0.025 | 19.02 | 11.81 | 0.067 | 0.10 | 0.08 |
| (0.105)                   | (32.51) | (22.95) | (0.029) | (10.11) | (3.00) | (0.072) | (0.01) | (0.01) |

The presented metrics are calculated by taking the average of the metrics for each outer fold and the values between parentheses are the standard deviation of these metrics. All metrics are presented in the actual scale, although the predictions were made by using the logarithmic scale of the socioeconomic indicators.

The values in bold correspond to the features that generated the highest R² or the smallest RMSE/MAE for each state.

All specifications used standardised predictors, that is, observations of each feature are subtracted by the feature’s mean and divided by its standard deviation. Also, the dependent variable (socioeconomic indicator) was examined in natural logarithm form.
Before presenting the final results, it is important to show how daytime image features can predict NTL, which is one step of the transfer learning process. For Bahia, the fine tune process ran 45 epochs\textsuperscript{12} and stopped for not being able to improve the accuracy after 10 epochs\textsuperscript{13}. The new model predicts nighttime lights intensities in the validation set with an accuracy of 0.871. For Rio Grande do Sul, the fine tune process ran 20 epochs and stopped after 10 epochs. The new model generated an accuracy metric in the validation set of 0.883 when predicting NTL.

Average monthly income

Table 1 presents the $R^2$, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for Bahia and Rio Grande do Sul for all types of features and socioeconomic indicator used. Regarding the average monthly income, results show that there is an improvement on the results when we compare the models that used features extracted by the transfer learning approach and by the VGG16 model with the simplest feature extraction methods (basic image features and light features).

For Bahia, the set of features that generated the highest $R^2$ (0.643), the smallest RMSE (71.08) and the smallest MAE (48.48) was the one extracted by the transfer learning model. The Mean Absolute Error of R$48.48$ achieved by this model is equivalent to 9.5% of the 2010 Brazilian minimum salary.\textsuperscript{14} As the nested cross-validation method uses random samples of the data to calculate these metrics, it is possible to calculate confidence intervals for their values. Considering a 95% confidence interval, it is not possible to assert that the transfer learning results differ from the ones generated by the VGG16 and lights features. Likewise, it is not possible to affirm that the results reached by combining the features extracted by all methods overcame significantly the other methods, except for the images basic features.

In the case of Rio Grande do Sul, the best results were achieved by the VGG16 features. The $R^2$ in this case was 0.404, the RMSE was 174.32 and the MAE was 129.00. Considering a 95% confidence interval, the $R^2$ for the VGG16 features and for the transfer learning features are not significantly different, but they are different from images basic features and light features. Once more, using all features combined could not overcome significantly all other methods.

In order to visualise spatially the quality of the prediction, Fig. 7 presents predictions by city for Bahia and Rio Grande do Sul. The predictions presented in the maps are the ones produced by the transfer learning model. Regarding the average monthly income, the maps show that estimated values are quite consistent with actual values, that is, cities with the highest incomes are in the coast (east) and south of Bahia and in the north of Rio Grande do Sul.

GDP per capita

Focusing on $R^2$ values, Table 1 shows that, regardless the feature type, $R^2$ is lower for GDP per capita than for average monthly income. Still, transfer leaning features reached an $R^2$ of 0.291 for Bahia and 0.242 for Rio Grande do Sul. Although transfer learning is the best model in this case, the results obtained by it are not significantly different from the results of the other features types for Bahia, but are significantly different from the results generated by the images basic features and light features for Rio Grande do Sul (considering a significance level of 95%). Besides, the transfer learning was also the best model if we use RMSE and MAE as evaluation metrics. However, the difference between its results and the results from the other models are not statistically significant. Likewise, the specification that combined features from all models could not overcome significantly the transfer learning approach.

Regarding the alignment of predicted and actual values distribution, Fig. 7 shows that the transfer learning approach could not identify clearly the richest areas on west of Bahia. However, the poor areas on the north and on the south were identified. In the case of Rio Grande do Sul, although it is possible to observe that the model estimated richer areas on the north, the magnitude of predictions is quite different.

\textsuperscript{12} Epoch means the number of times the algorithm passes through the entire training set.

\textsuperscript{13} Such early stopping technique is used to avoid overfitting.

\textsuperscript{14} In 2010, the official minimum salary in Brazil was R$510.00.
from the actual values. A matching of poor areas
between actual and estimated values can be seen on the east and on the south.

Water Index

For both states, $R^2$ for the best models (transfer learning and VGG16) are not significantly different from the $R^2$ obtained using simple features (lights and images basic features). On one hand, looking at this metric, the models did not fit well, since the best $R^2$ for Bahia was around 11% and for Rio Grande do Sul around 14%. On the other hand, the predictions were accurate when we consider RMSE and MAE. For Bahia, which the minimum, median and maximum values are 0.44, 0.80 and 1.00 respectively, the RMSE was 0.07 and MAE was 0.06. For Rio Grande do Sul, which minimum, median and maximum values are 0.41, 0.79 and 1.00 respectively, the RMSE was 0.10 and MAE was 0.08.

Regarding the distribution of the predictions, from Fig. 7 it is possible to observe that the transfer learning model is estimating similar values for water index. It basically estimates values near the average with small shifts. The grey areas in the map are the cities with missing values for this index.

Results using transfer learning model trained on a different state

Besides comparing the prediction performance of transfer learning with other methods, this work also analyses the generalisation ability of this approach. This generalisation ability will assess if we can use only daytime images from one state to predict the socioeconomic indicators by training the whole model using images and data from another state. For example, suppose that Bahia cannot collect survey data in one year and need to estimate the indicators in its cities. We can train the CNN model using images of Rio Grande do Sul and then extract their features. These features are used in both, the inner cross-validation to select the best regularised regression model and the process of fitting the model. Finally, the socioeconomic indicator in Bahia is estimated by applying the fitted model to Bahia images features that are extracted from the CNN model trained on Rio Grande do Sul.

Figure 8 presents results for all socioeconomic indicators and evaluation metrics. In general, the models trained and tested on the same state predict more accurately. For $R^2$, only for the water index the differences between the model trained and evaluated on the same state and the one trained on one state and evaluated on the other are not statistically significant. For the other proxies, $R^2$ is greater when using data of the same state. RMSE values are significantly lower when using in-state model only for average monthly income, considering a significance level of 95%. The other proxies do not present statistically significant differences. Finally, results using MAE are similar to those using $R^2$, that is, we can affirm that predictions were more precise for average income and GDP per capita, but we cannot affirm it for water index.

Discussion

For the three variables tested, average monthly income was the variable for which estimates had the best evaluation metrics for both states when we compare the $R^2$ values. This result is interesting, since the average monthly income data is from 2010 and the nighttime and daytime images from 2016 and 2019, respectively. A possible reason for that is that the difference of average income between cities did not change significantly, as showed in “Census and survey data” section.

The best results for average income are aligned to the ones achieved by other works that tested different types of models to predict poverty in countries of Africa and Asia. For example, Jean et al. (2016) found $R^2$ values between 44 and 75% depending on the African country and variable used, and cross-border results (using a model from one country to estimate poverty on another) varying from 19 to 71%; Perez et al. (2017) achieved an $R^2$ of 66% for multiple countries in Africa using the Asset Wealth Index; Zhao et al. (2019) found that the $R^2$ for Wealth Index for Bangladesh was 70% and, when using the Bangladesh model to estimate values for Nepal, such metric was 61%.

In general, it seems that models could generate more accurate predictions for Bahia than to Rio Grande do Sul, which means that, in this case, models fit better in a highly unequal poor state than in a moderately unequal rich state. More studies including other Brazilian states in the analysis should be done to verify whether such relation holds or not.
Also, for both states and all metrics, the transfer learning approach was one of the best models considering a 95% confidence interval. Therefore, there is evidence to believe that the use of such model, which combines daytime and nighttime images, can improve estimation effectiveness when compared to simpler approaches. Results also show that including all features in the same model did not improve predictions accuracy significantly.

Regarding generalisation, considering average monthly income and GDP per capita, all evaluation metrics values, except for GDP per capita using RMSE, showed that in-state models overcame out-of-state models. The differences for water index are not significant. Theses results go against the possibility of using only the images from one state to predict these socioeconomic indicators using a model trained on another state. However, we believe that such performance should be tested using other states, since Bahia and Rio Grande do Sul are quite different in terms of income level, inequality index and geographic features.

There are some unexplored aspects in this article that can be addressed in future works. First, in the transfer learning approach, other pre-trained models can be tested as well as different architectures for the

**Fig. 8** Evaluation metrics of average monthly income, GDP per capita and water index predictions for transfer learning model trained on one state and tested on the other. Notes: Plots are distributed as following: the first column presents results for average monthly income; the second column, results for GDP per capita; the third column, for water index. Likewise, plots on the first row are related to $R^2$; on the second row, to RMSE; and on the third row, to MAE. The main diagonal shows the cases in which the model was trained on one state and evaluated on the other. The antidiagonal presents the cases in which the model was trained and evaluated on the same state. BA is the abbreviation of Bahia and RS the abbreviation of Rio Grande do Sul.
second convolutional neural network. Likewise, different types of methods such as multi-task fully convolutional model employed by Pandey et al. (2018) and the random forest regression model employed by Zhao et al. (2019) may be applied to Brazilian data. Moreover, the study area can be expanded, in a way that more states or all Brazilian cities are included in the study. Finally, when the next Brazilian census data is available, the socioeconomic indicators can be tested without the issue of having a gap between satellite imagery year and socioeconomic variable year.

Conclusions

This paper assessed the effectiveness of employing satellite images to predict socioeconomic indicators in Brazil. More specifically, it aimed at verifying how useful transfer learning approach based on a model previously trained on the ImageNet database is when compared to other kinds of features extraction techniques. Such type of transfer learning method was chosen because, due to the lack of training data to apply convolutional neural networks models to directly predict these indicators, it can use pre-trained models to extract daytime images features that are used subsequently to fit a new CNN. This new CNN classifies nighttime lights intensity and, then, extracts features that are employed in the regression process.

Two Brazilian states with more than 400 cities and with different characteristics were the study area of this work. Also, nighttime lights intensities, daytime images and socioeconomic variable (average monthly income, GDP per capita and water index) were the data used to investigate the relation between satellite imagery and socioeconomic indicators in the city level.

Regarding the results, although there is a considerable gap between the average monthly income year and the satellite imagery years, the average income was the socioeconomic indicator that could be predicted more similarly to the actual values in general. Also, Bahia presented better evaluation metrics than Rio Grande do Sul overall. As Bahia is poorer and more economically unequal than Rio Grande do Sul, these aspects may be some of the reasons for the higher performance, even though specific studies focusing on this issue should be done to verify this hypothesis.

The regression analyses showed that the transfer learning model was one of the best specifications for all indicators and states. Also, \( R^2 \) obtained in this article were consistent with the literature, reaching the higher metric of 64% for average income in Bahia. Moreover, observing the cross-state transfer learning evaluation metrics, this model was not capable of generating accurate predictions similarly to in-state approaches. One possible reason for that are the differences in terms of income level, inequality index and geographic features between Bahia and Rio Grande do Sul.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Amaral, S., Câmara, G., Monteiro, A. M. V., Quintanilha, J. A., & Elvidge, C. D. (2005). Estimating population and energy consumption in Brazilian Amazonia using DMSP nighttime satellite data. Computers, Environment and Urban Systems, 29(2), 179–195. https://doi.org/10.1016/j.compenvurbsys.2003.09.004.

Amaral, S., Monteiro, A. M. V., Camara, G., & Quintanilha, J. A. (2006). DMSP/OLS night-time light imagery for urban population estimates in the Brazilian Amazon. International Journal of Remote Sensing, 27(5), 855–870. https://doi.org/10.1080/01431160500181861.

Bai, R., Lam, J. C. K., & Li, V. O. K. (2020). Siamese-like convolutional neural network for fine-grained income estimation of developed economies. IEEE Access, 8, 162533–162547. https://doi.org/10.1109/ACCESS.2020.3019239.

Bennett, M. M., & Smith, L. C. (2017). Advances in using multitemporal night-time lights satellite imagery to detect, estimate, and monitor socioeconomic dynamics. Remote Sensing of Environment, 192, 176–197. https://doi.org/10.1016/j.rse.2017.01.005.

Bishop, C. M. (2006). Pattern recognition and machine learning. Berlin: Springer.

Brazilian Institute of Geography and Statistics. (2010a). Censo 2010. Retrieved August 23, 2019 from https://censo2010.ibge.gov.br/materiais/guia-do-censo/operacao-censitaria.html.

Brazilian Institute of Geography and Statistics. (2010b). Portal de mapas. Retrieved August 23, 2019 from https://portaldemapas.ibge.gov.br/portal.php#/homepage.
Brazilian Institute of Geography and Statistics. (2018a). Panorama Bahia. Retrieved August 24, 2019 from https://cidades.ibge.gov.br/brasil/ba/panorama.

Brazilian Institute of Geography and Statistics. (2018b). Panorama Rio Grande do Sul. Retrieved August 24, 2019 from https://cidades.ibge.gov.br/brasil/rs/panorama.

Brazilian Institute of Geography and Statistics. (2019). Síntese de indicadores sociais: uma análise das condições de vida da população brasileira. Estudos e pesquisas, informação demográfica e socioeconômica, n. 40. Retrieved December 03, 2019 from https://biblioteca.ibge.gov.br/visualizacao/livros/liv101678.pdf.

Campos, J., Rigotti, J. I. R., Baptista, E. A., Monteiro, A. M. V., & Reis, I. A. (2020). Population estimates from orbital data of medium spatial resolution: Applications for a Brazilian municipality. Sustainability. https://doi.org/10.3390/su12093565.

Charris, C., Veilhia, R., & Chaves, L. (2019). Mapping the human development index using nighttime lights inside Brazil. In XVII ENABER—Encontro Nacional da Associação Brasileira de Estudos Regionais e Urbanos. https://brsa.org.br/wp-content/uploads/wpcf7-submissions/990/manuscript_Iden.pdf.

Chatfield, K., Simonyan, K., Vedaldi, A., & Zisserman, A. (2014) Return of the devil in the details: Delving deep into convolutional nets. In Proceedings of the British machine vision conference, BMVA Press. https://doi.org/10.5244/C.28.6.

Chen, X., & Nordhaus, W. D. (2019). VIIRS nighttime lights in the estimation of cross-sectional and time-series GDP. Remote Sensing, 11(9), 1057.

Doll, C. N., Muller, J. P., & Morley, J. G. (2006). Mapping regional economic activity from night-time light satellite imagery. Ecological Economics, 57(1), 75–92. https://doi.org/10.1016/j.ecolecon.2005.03.007.

Elvidge, C. D., Baugh, K. E., Anderson, S. J., Sutton, P. C., & Ghosh, T. (2012). The Night Light Development Index (NLDI): A spatially explicit measure of human development from satellite data. Social Geography, 7(1), 23–35. https://doi.org/10.5194/s-7-23-2012.

Elvidge, C. D., Baugh, K. E., Kihn, E. A., Kroepl, H. W., Davis, E. R., & Davis, C. W. (1997). Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. International Journal of Remote Sensing, 18(6), 1373–1379. https://doi.org/10.1080/014311697218485.

Elvidge, C. D., Sutton, P. C., Ghosh, T., Tuttle, B. T., Baugh, K. E., Bhaduri, B., & Bright, E. (2009). A global poverty map derived from satellite data. Computers & Geosciences, 35(8), 1652–1660. https://doi.org/10.1016/j.cageo.2009.01.009.

Engstrom, R., Hersh, J., & Newhouse, D. (2017). Poverty from space: Using high-resolution satellite imagery for estimating economic well-being. The World Bank. https://doi.org/10.1596/1813-9450-8284.

Engstrom, R., Newhouse, D., & Soundrarajan, V. (2019). Estimating small area population density using survey data and satellite imagery: An application to Sri Lanka. Poverty and equity global practice working paper, no. 194, The World Bank, Washington, D.C.

Engstrom, R., Newhouse, D., Haldavanekar, V., Copenhaver, A., & Hersh, J. (2017b). Evaluating the relationship between spatial and spectral features derived from high spatial resolution satellite data and urban poverty in Colombo, Sri Lanka. In 2017 Joint urban remote sensing event (JURSE), pp. 1–4. https://doi.org/10.1109/JURSE.2017.7924590.

FGV Social. (2020). Where are the “rich” in Brazil? FGV social: Social Policy Centre, Rio de Janeiro, Brazil.

Gao, B., Huang, Q., He, C., & Ma, Q. (2015). Dynamics of urbanization levels in China from 1992 to 2012: Perspective from DMSP/OLS nighttime light data. Remote Sensing, 7(2), 1721–1735. https://doi.org/10.3390/rs7021721.

Ghosh, T., Anderson, S., Elvidge, C., & Sutton, P. (2013). Using nighttime satellite imagery as a proxy measure of human well-being. Sustainability, 5(12), 4988–5019. https://doi.org/10.3390/su5124988.

Graetz, N., Friedman, J., Zimmerman, A., Burstein, R., Biehl, M., Shields, C., et al. (2018). Mapping local variation in educational attainment across Africa. Nature, 555, 48–53. https://doi.org/10.1038/nature25761.

Henderson, J. V., Storeygard, A., & Weil, D. N. (2012). Measuring economic growth from outer space. American Economic Review, 102(2), 994–1028. https://doi.org/10.1257/aer.102.2.994.

Hofmann, P., Strobl, J., Blaschke, T., & Kux, H. (2008). Detecting informal settlements from QuickBird data in Rio de Janeiro using an object based approach. In Object-based image analysis, Springer, Lecture notes in geoinformation and cartography, pp. 531–553. https://doi.org/10.1007/978-3-540-77058-9_29.

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.

Jolliffe, D., Negre, M., & Schmidt, M. (2018). Poverty and shared prosperity 2018: Piecing together the poverty puzzle. International Bank for Reconstruction and Development/The World Bank, Washington DC. https://doi.org/10.1596/978-1-4648-1330-6.

Lecun, Y. (1989). Generalization and network design strategies. In R. Pfeifer, Z. Schreter, F. Fogelman & L. Steels (Eds), Proceedings of the international conference connectionism in perspective, Elsevier, University of Zürich.

LeCun, Y., Haffner, P., Bottou, L., & Bengio, Y. (1999) Object recognition with gradient-based learning. In Shape, contour and grouping in computer vision, Springer, London, UK, pp. 319–345. https://doi.org/10.1007/3-540-46805-6_19.

Maroko, A., Maantay, J., Machado, R. P. P., & Barrozo, L. V. (2019). Improving population mapping and exposure assessment: Three-dimensional dasymetric disaggregation in New York City and São Paulo. Brazil. Papers in Applied Geography, 51–52, 45–57.

Ministry of Regional Development of Brazil. (2019). Sistema Nacional de Informações sobre Saneamento. Retrieved August 25, 2019 from http://app4.cidades.gov.br/serieHistorica/.

Murphy, K. P. (2013). Machine learning: A probabilistic perspective. Cambridge, MA: MIT Press.
Nadalini, V. G., & Mation, L. F. (2018). Localização intraurbana das favelas brasileiras: o papel dos fatores geográficos. Texto para discussão 2390, Instituto de Pesquisa Econômica Aplicada. http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=33697.

National Oceanic and Atmospheric Administration. (2019). Version 1 VIIRS Day/Night Band nighttime lights. Retrieved August 23, 2019 from https://ngdc.noaa.gov/eog/viirs/download_dnb_composites.html.

Neves, O. M. S., Strauch, J. C. M., & Ajara, C. (2017). Dasymetric methods applied to Jacarepaguá Watershed. Bulletin of Geodetic Sciences, 23(4), 606–622. https://doi.org/10.1590/S1982-21702017000400040.

Pandey, S., Agarwal, T., & Krishnan, N. C. (2018). Multi-task deep learning for predicting poverty from satellite images. In Proceedings of the thirtieth AAAI conference on innovative applications of artificial intelligence, AAAI Press, IAAI-18, pp. 7793–7798. https://aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16441.

Patino, J. E., & Duque, J. C. (2013). A review of regional science applications of satellite remote sensing in urban settings. Computers, Environment and Urban Systems, 37, 1–17. https://doi.org/10.1016/j.compenvurbsys.2012.06.003.

Perez, A., Yeh, C., Azzari, G., Burke, M., Lobell, D., & Ermon, S. (2017). Poverty prediction with public Landsat 7 satellite imagery and machine learning. In 31st Conference on neural information processing systems, NIPS 2017. arXiv: 1711.03654.

Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. Neural Computation, 29(9), 2352–2449. https://doi.org/10.1162/neco_a_00990.

Sandborn, A., & Engstrom, R. N. (2016). Determining the relationship between census data and spatial features derived from high-resolution imagery in Accra, Ghana. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 9(5), 1970–1977. https://doi.org/10.1109/JSTARS.2016.2519843.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. In International conference on learning representations. https://www.robots.ox.ac.uk/~vgg/publications/2015/Simonyan15/.

Sutton, P. C., Elvidge, C. D., & Ghosh, T. (2007). Estimation of gross domestic product at sub-national scales using nighttime satellite imagery. International Journal of Ecological Economics and Statistics, 8, 5–21.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society (Series B), 58, 267–288.

Tilottama, G., Powell, R., Elvidge, C., Baugh, K., Sutton, P., & Anderson, S. (2010). Shedding light on the global distribution of economic activity. The Open Geography Journal, 3, 147–160. https://doi.org/10.2174/1874923201003010147.

Tomás, L., Fonseca, L., Almeida, C., Leonardi, F., & Pereira, M. (2016). Urban population estimation based on residential buildings volume using IKONOS-2 images and lidar data. International Journal of Remote Sensing, 37(sup1), 1–28. https://doi.org/10.1080/01431161.2015.1121301.

Xie, M., Jean, N., Burke, M., Lobell, D., & Ermon, S. (2016). Transfer learning from deep features for remote sensing and poverty mapping. In Proceedings of the thirtieth AAAI conference on artificial intelligence, AAAI Press, AAAI-16, pp. 3929–3935. arXiv:1510.00098.

Yu, B., Shi, K., Hu, Y., Huang, C., Chen, Z., & Wu, J. (2015). Poverty evaluation using NPP-VIIRS nighttime light composite data at the county level in China. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(3), 1217–1229. https://doi.org/10.1109/JSTARS.2015.2399416.

Zhao, X., Yu, B., Liu, Y., Chen, Z., Li, Q., Wang, C., & Wu, J. (2019). Estimation of poverty using random forest regression with multi-source data: A case study in Bangladesh. Remote Sensing, 11(4), 375. https://doi.org/10.3390/rs1104375.

Zhou, N., Hubacek, K., & Roberts, M. (2015). Analysis of spatial patterns of urban growth across South Asia using DMSP-OLS nighttime lights data. Applied Geography, 63, 292–303. https://doi.org/10.1016/j.apgeog.2015.06.016.

Zimmerman, A., Millear, I., Stubbs, R., Shields, C., Pickering, B., Earl, L., et al. (2018). Mapping child growth failure in Africa between 2000 and 2015. Nature, 555, 41–47. https://doi.org/10.1038/nature25760.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.