Satellite Imagery Predicts Socioeconomic Indicators: experiments in Brazil

1st Joao Pedro Silva
University of Sao Paulo, Sao Carlos, SP, Brazil
jp.silva@usp.br

2nd Jose F Rodrigues-Jr
University of Sao Paulo, Sao Carlos, SP, Brazil
junio@usp.br

***Abstract*—The world has a considerable portion of its population living in vulnerable conditions, with about 47% of people living in poverty and approximately 9.3% in extreme poverty. To understand and face this problem, it is essential to have access to updated data on poverty and trace the socioeconomic profile of the population. In this context, Censuses and demographic surveys play a crucial role. However, conducting a Census in certain regions faces significant challenges, both in terms of cost and logistics. It is a complex process that demands significant resources and an extensive data collection effort throughout the national territories. Given the difficulties of collecting data in the traditional way, passively collected data sources, such as satellite imagery, may be an alternative way to measure these results. Therefore, we use computational techniques to combine nighttime and daytime satellite imagery to predict socioeconomic indicators. Our results demonstrate the efficacy of combining these features in predicting the average income of cities across Brazil. In particular, the use of neural networks for inferring socioeconomic indicators has demonstrated highly effective. This technique holds the potential to enhance our comprehension of the socioeconomic landscape in Brazil and provide technical means for the analysis of countries that face difficulties in conducting Censuses. Our method has convincingly demonstrated that satellite images can be used and applied for socioeconomic purposes.**

***Index Terms*—Deep Learning, Socioeconomic indicators, Remote Sensing, Satellite Imagery**

I. Introduction

Recurrently, the world has been confronted with series of global challenges, particularly regarding vulnerable populations. Conflicts, financial crises, and the impact of the COVID-19 pandemic have worsened the socioeconomic conditions of these populations [2], requiring a comprehensive understanding of their needs and the implementation of effective public policies.

Particularly, in Brazil, poverty is a reality that affects a significant portion of the population. Continuously, millions of people live in extreme poverty, facing difficulties in meeting their basic needs [22] [1]. Understanding the magnitude and dynamics of these social problems remains a challenge in many countries, making it difficult to implement socioeconomic measures.

Nationwide census is the standard manner to gather information about a population and its socioeconomic characteristics. These surveys are costly and demand significant effort, especially in countries with a large population like Brazil, which has over 200 million inhabitants. The allocated budget for the 2022 Brazilian census was nearly U$ 400 million [14], which prevents continuous monitoring of the population. With a decennial periodicity between censuses, there is a large gap between social problems and government actions. In Brazil, the COVID-19 pandemic aggravated the situation, especially in hard-to-reach regions, postponing the 2020 census in two years [20].

The delay in obtaining the data negatively impacts the development of efficient public policies. Fortunately, in recent decades, the availability of a large amount of passively generated data has shown potential in overcoming these obstacles - among these data sources, satellite imagery stands out. In the 21st century, numerous satellites have been launched for remote sensing, generating massive public data [18]. At the same time, machine learning (ML) techniques have become increasingly sophisticated, enabling the analysis and mapping of populations in vulnerable situations [4].

This work utilizes Deep Learning models combined with daytime and nighttime satellite images to predict indicators, particularly the average monthly income across the national territory of Brazil, the 5th largest country in the world. The hypothesis is that the patterns observed in aerial images of urban agglomerations can reveal the level of socioeconomic indicators. The proposal uses transfer learning techniques, loading the weights of a model pre-trained with dataset ImageNet [24] into a new model; initially, we train a feedforward network to predict nighttime light (NTL) indices. This first stage provides a fine-tuning of the methodology. Subsequently, the model provides features used to estimate the average income of cities. Through regression analysis, we demonstrate that the features extracted by the model are effective in predicting this socioeconomic indicator.

The rest of the article is organized as follows: Section II describes related works, introducing the background of using deep learning and satellite images for predicting socioeconomic indicators. Section III describes the methodology proposed. Section IV presents the experiments and results. Finally, Section V presents the conclusions.

II. Related work

A popular approach in the analysis of economic activity involves the use of satellite images. Such images, when captured during the night, can detect light emissions. One of the pioneering studies, by Elvidge et al. [8], has demonstrated

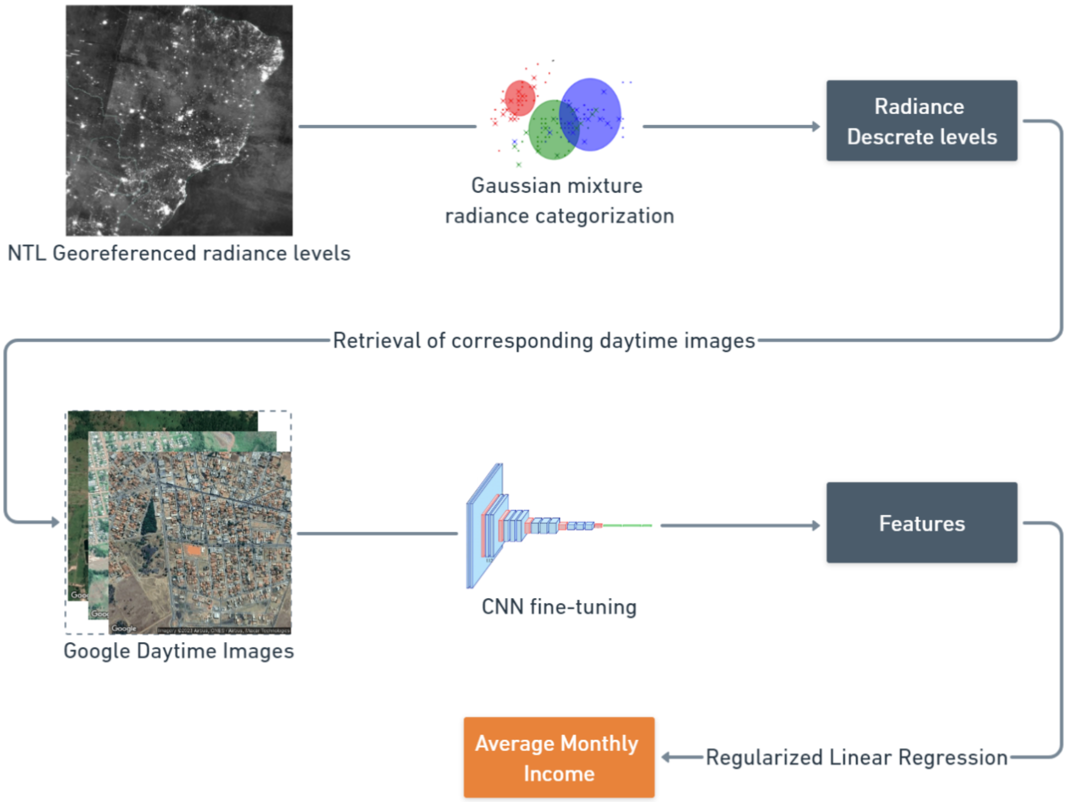

Fig. 1: Methodology Workflow

a high correlation between nighttime lights (NTL) and gross domestic product (GDP), as well as electricity consumption. The study covered 21 countries and utilized nighttime images from the Defense Meteorological Satellite Program (DMSP) [16]. In another work, Elvidge et al. [9] demonstrated that NTL has the potential to indicate human presence and land surface changes, in addition to predicting metrics like annual growth rates. However, according to Jean *et al.* [11], this approach is not as effective to distinguish differences in economic activity over impoverished areas, where light levels are generally low and homogeneous.

Another approach used by various authors involves the utilization of only daytime lights for predicting socioeconomic indicators. The study conducted by Duque *et al.* [7] is based on the premise that the physical appearance of a human settlement reflects its society. The assumption is that people living in urban areas with similar housing conditions have similar social and demographic characteristics. In their study, they used images acquired with DigitalGlobe's Quickbird satellite [6] to extract land cover, urban texture, and urban structure features.

Xie et al. [27] combines nighttime and daytime lights. The authors trained a neural network to predict nighttime light indices from daytime images. From the neural network, they extracted features that relate the images to the socioeconomic indicators of each location. Subsequently, using the features, a second model applies regularized regression methods (ridge and lasso) to predict poverty indices based on the Living Standards Measurement Study (LSMS) conducted in Uganda [13]. Upon this methodology, Jean et al. [11] proposed an expansion of this work to five African countries: Nigeria, Tanzania, Uganda, Malawi, and Rwanda.

In the case of Brazil, the first study proposing a similar methodology to Xie et al. and Jean et al. was conducted by Castro and Alvarez [5], who focused on two Brazilian states: Rio Grande do Sul and Bahia. Following recent works, Castro and Alvarez presented multiple approaches, comparing only nighttime light features, only daytime light features, features extracted from a transfer learning model (pre-trained with ImageNet), and the combination of all these features using regularized regression methods.

III. Methodology

The goal is to extract socioeconomic features from satellite imagery using a fine-tuned Convolutional Neural Network (CNN) [12] enabling us to predict the average monthly income of Brazilian cities. Our hypothesis is that by training a model to predict the radiance levels (class) of Nighttime Lights (NTL), it will produce features suited to identify human-built structures such as roads, buildings, and farms. Such structures, which correlate to the wealth of a population, should provide for predicting the monthly income of the corresponding urban areas by means of a Regularized Regression model.

Our methodology utilizes NTL images – illustrated in Figure 2, from which we gather georeferenced data labeled with radiance indices across the Brazilian territory. From the georeferenced data, we acquire the set of corresponding daytime images considering each NTL point. Next, we use the daytime images as input and the NTL radiance indices of each city as labels; with this setting, we fine-tune a CNN model pre-trained with ImageNet. The model, previously trained on a general images-classification problem, now is refined to predict the radiance levels given the daytime images. Lastly, we use the features extracted by the CNN model as input to a Regularized Linear Regression model whose goal is to predict the average monthly income of the cities.

Figure 1 illustrates the methodology: (I) Gaussian mixture discrete radiance categorization; (ii) retrieval of corresponding daytime images; (iii) fine-tuning of the CNN model; and (iv) regression. The subsequent subsections provide a detailed description of each step.

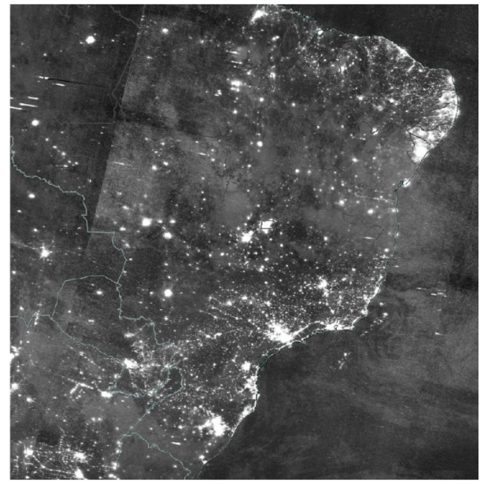

Fig. 2: NTL image from satellite VIIRS/DNB. The image depicts Brazil in 2021.

A. Gaussian mixture radiance categorization

We acquired nighttime satellite images of Brazil from the Visible Infrared Imaging Radiometer Suite (VIIRS) satellite [17], which provides comprehensive global nighttime imagery. The images, dated 2021, contain georeferenced radiance levels representing the intensity of artificial lights in the entire globe. We refer to these georeferenced radiance levels as NTL points.

To associate each NTL point with its corresponding city, we employed a geospatial shapefile containing the boundaries of the Brazilian municipalities. These shapefiles were obtained from the official website of the Brazilian Institute of Geography and Statistics (IBGE) [19]. By matching the coordinates of each NTL point to the boundaries of the municipalities, we assigned the points to their respective city codes.

Next, we classify each point based on its radiance level; we used the technique Gaussian Mixture Model. The categorization was necessary because the next steps of the methodology are based on a classification task, hence, it assumes discrete labels.

According to Bishop [3], a Gaussian Mixture Model is defined as a linear superposition of K Gaussian densities:

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{1}$$

where $\mathbf{x}$ is the observed variable, π_k is the mixing coefficient for the k-th component, and $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ are the mean vector and covariance matrix, respectively, for the k-th component.

We employed K = 3 Gaussian distributions to achieve three categories of NTL intensity: low, medium, and high. By fitting the GMM to the average radiance values of the cities, we estimated the mean, covariance, and weights for each Gaussian distribution. Subsequently, we assigned each city to the category represented by the Gaussian distribution with the highest weight, categorizing them into low, medium, or high NTL intensity.

B. Retrieval of corresponding daytime images

In addition to nighttime images, we employed high-resolution daytime images. For each selected city, we chose the 100 NTL points closest to the city center. For each of these points, we obtained their corresponding daytime image. Complementing the NTL images, the incorporation of daytime images adds contextual information about the urban landscape, infrastructure, and socioeconomic characteristics of the cities.

To acquire the daytime images, we used the Google Maps Static API [23]. This API allows the download of images with 400x400 resolution at zoom level 16 – illustrated in Figure 3. The API ensures consistency as all images are obtained under the same specifications. Accordingly, each daytime image is linked to its corresponding NTL radiance, indicating low, medium, or high intensity.

C. CNN fine-tuning

After labeling the daytime images with their NTL radiance classes, the next step was to extract features from these images using a CNN pre-trained with ImageNet – we experimented with architectures VGG16 [25] and ResNet50 [10]. These images are used as input, and their labels as output, to train the model. That is, the model is trained to predict the corresponding class of each daytime image. Upon completion of the training process, the model will have extracted highly-descriptive features from the daytime images. This process, known as fine-tuning, entailed unfreezing the weights of the convolutional layers and adding a new block of classification.

Upon completion of the training, we considered the last block of layers to extract the output from the final convolutional layer, resulting in a feature vector for each input image. We hypothesize that if these features can accurately predict the radiance class, they are also suited for predicting the monthly income of Brazilian cities. This is due to the correlation between radiance and human-built structures, which in turn reflect the socioeconomic status of a population.

Fig. 3: Daytime images from Google Maps Static API representing **low**, **medium** and **high** NTL radiance classes respectively

We obtained 100 feature vectors for each of the 140 cities, each of which contains 2,048 feature maps with 7x7 size. Each set of 100 vectors was aggregated into a single vector by means of average. This process resulted in 140 vectors, one for each city.

We compared architectures VGG16 and ResNet50 in the preliminary task of predicting the radiance. VGG16 achieved an accuracy of 0.83, while ResNet50 yielded a slightly lower result of 0.82. We present the results in Table I, in the table, one can see that VGG16 presented the best results for all the metrics, but AUC. The performance of architectures VGG16 and ResNet50 differ by nearly 1%, which is not statistically significant. Yet, these results indicate a higher prospect for VGG16, which might surpass the performance of ResNet50 in larger datasets, or for similar problems. For the rest of this work, we experiment with both architectures, as detailed in Section IV.

TABLE I: CNN architecture metrics obtained by training the models to predict the NTL radiance class from daytime images.

Model	Accuracy	Precision	Recall	AUC
VGG16	0,8333	0,8333	0,8333	0,9144
ResNet50	0,8210	0,8213	0,8205	0,915

D. Regularized Linear Regression

In the final step, we employed the features extracted from the daytime images to predict the average monthly income of the cities. This machine learning process considered the projected average monthly income provided by the Brazilian agency Fundacao Getulio Vargas (FGV) for 2020 [26]; this income projection is based on the Individual Income Tax Return and on the preliminary population estimate from the Brazilian Institute of Geography and Statistics.

We used this data to fit a Regularized Linear Regression, which is a technique that introduces a penalty term to the traditional least squares method, helping to prevent overfitting and improving generalization performance. In our case, we used ElasticNet regularization, which combines both L1 (lasso) and L2 (ridge) penalties, as follows:

$$\min_{\mathbf{w}} \frac{1}{2n_{\text{samples}}} \|\mathbf{X}\mathbf{w} - \mathbf{y}\|_2^2 + \alpha\rho\|\mathbf{w}\|_1 + \frac{\alpha(1-\rho}{2})\|\mathbf{w}\|_2^2 \quad (2)$$

where n_{samples} is the number of samples in the dataset, y the target vector, X the feature matrix, and w the coefficient vector to be estimated; α is the regularization parameter that controls the overall strength of regularization – a higher value of α leads to a stronger regularization; ρ is the mixing parameter in the range [0,1] that determines the balance between regularizations L1 and L2 – 1.0 for L1 only and 0.0 for L2 only.

IV. Experiments and Results

We present our experiments and results with respect to the task of predicting the average monthly income of Brazilian municipalities using features extracted from satellite imagery.

A. Dataset

For our experiments, we sampled the georeferenced NTL points to create a representative dataset of the cities and their respective radiance levels. Considering the vast variation of city sizes across Brazil, a different number of NTL points would be required to adequately represent each city. While a small number of NTL points may suffice for smaller cities, the same number would only represent the city center for larger metropolitan areas, failing to capture the socioeconomic reality of the entire municipality. To address this issue, we constructed our database using cities where 100 nighttime light points adequately covered the entire city, thereby excluding very small and very large cities. We followed previous works, by Xie et al. [27], and Jean et al., and Castro and Alvares [5], who selected the 100 nearest NTL points to the urban center of each city.

We derived the list of cities from the Preliminary Population of Municipalities, which is based on the data from the 2022 Brazilian Demographic Census [21]. Specifically, we selected cities based on the population size distribution – see Figure

4, focusing on cities within the 10th and the 75th percentile. This range corresponds to cities with approximately 3,000 to 25,000 inhabitants.

Next, we created a uniform sample comprising 140 cities randomly selected, with approximately six cities per each of the 26 Brazilian states. The sample was designed to ensure an even distribution across the three NTL intensity classes described in the previous subsection, thereby encompassing the full range of city amplitudes. The distribution of cities per NTL radiance class and Brazilian states are presented in Table II. We retrieved 14,000 daytime images, considering the 140 cities of our dataset, or 100 images per city.

For the task of training the CNN model to predict the NTL class of daytime images, we split our dataset of images into training, validation, and testing, with 70% used for training, 15% for validation, and 15% for testing.

TABLE II: Quantity of cities by Brazilian state and NTL radiance class.

State	**Low**	**Medium**	**High**	**Total**
Acre (AC)	1	1	1	3
Alagoas (AL)	2	2	2	6
Amapa (AP)	1	1	1	3
Amazonas (AM)	1	2	2	5
Bahia (BA)	2	2	2	6
Ceara (CE)	2	2	2	6
Espírito Santo (ES)	2	2	2	6
Goias (GO)	2	2	2	6
Maranhão (MA)	2	2	2	6
Mato Grosso (MT)	2	2	2	6
Mato Grosso do Sul (MS)	2	2	2	6
Minas Gerais (MG)	2	1	2	5
Para (PA)	0	2	1	3
Paraíba (PB)	2	2	2	6
Parana (PR)	2	2	2	6
Pernambuco (PE)	2	2	2	6
Piauí (PI)	2	2	2	6
Rio de Janeiro (RJ)	1	2	2	5
Rio Grande do Norte (RN)	2	2	1	5
Rio Grande do Sul (RS)	2	2	2	6
Rondônia (RO)	1	1	1	3
Roraima (RR)	2	2	2	6
Santa Catarina (SC)	2	2	2	6
São Paulo (SP)	2	2	2	6
Sergipe (SE)	2	2	2	6
Tocantins (TO)	2	2	2	6
Total	**45**	**48**	**47**	**140**

B. Experimental setting

We directly compare our methodology with the work proposed by Castro and Alvares [5], our comparison baseline. This work uses a complex workflow and an intricate set of features. It comprises four feature extraction processes, referred to as "Multiple features" that are combined into four groups of features extracted from satellite imagery.

The experiments were performed on a Ryzen(R) 9 CPU, 32GB RAM, NVIDIA RTX3090 24GB RAM GPU. We implemented the deep learning and machine learning models using Keras with TensorFlow on a Linux Pop!OS 22.04 LTS.

C. Experimental comparison

We compare our work with three approaches: a baseline, in-domain pre-trained models, and a variation of our own methodology.

Baseline

The baseline proposed by Castro and Alvarez [5], as introduced in Section II, involves conducting four processes of feature extraction from satellite imagery: extracting features from nighttime images, extracting basic color-related features from daytime images, extracting features from daytime images using a pre-trained VGG16 model with ImageNet, and extracting additional features from daytime images using a CNN model specifically developed by the authors.

We reproduce the entire baseline work and apply their method over our image dataset. This way, we ensure an exact comparison in terms of prediction performance. Similarly to our work, the resulting features are used as input for a Regularized Linear Regression model. The best performance was obtained using L1 regularization.

In-Domain pre-trained models

In addition, we use Transfer Learning without fine-tuning. The goal is to demonstrate the importance of our intermediary fine-tuning step, which is based on radiance prediction. We used a collection with 5 pre-trained ResNet50 models with ImageNet and In-Domain benchmark datasets, namely:

- BigEarthNet
- EuroSat
- RESISC-45
- So2Sat
- UC Merced

These models[1] were proposed by Neumann et al [15]. The 5 models can extract generic representations from remote sensing data. In our experiments, we kept their convolutional layers frozen, adding a new block of classification layers.

Comparing architectures VGG16 and ResNet50

We also compare our methodology configured with architectures VGG16 and ResNet50. Since we verified that VGG16 works only slightly better in the preliminary step of predicting the radiance class (Section III-C), we still experiment with ResNet50.

[1] https://tfhub.dev/google/collections/remote_sensing/1

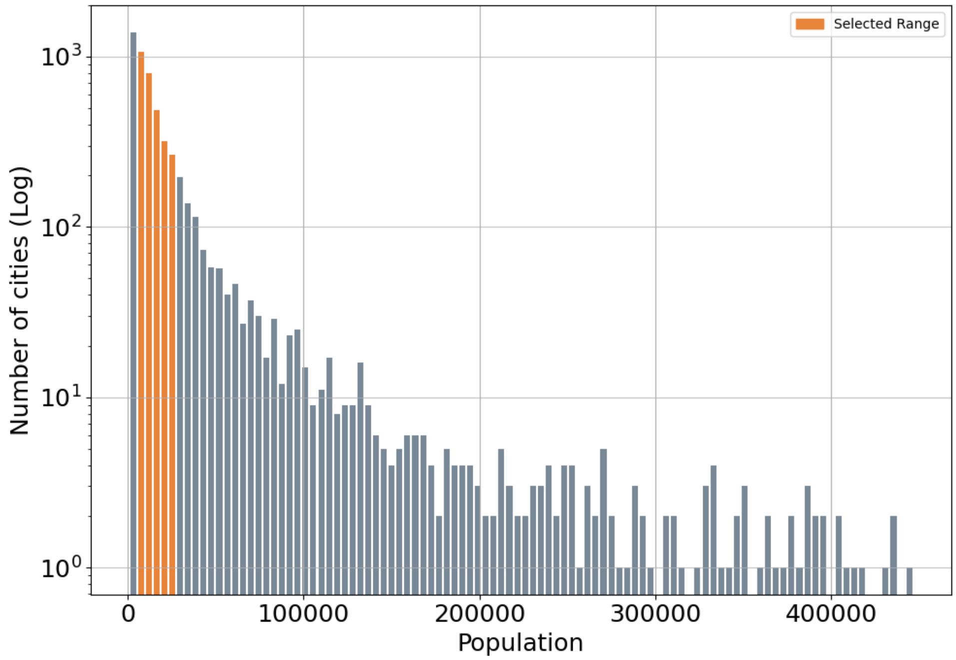

Fig. 4: Distribution of population size in Brazil for cities up to 500,000 inhabitants. Cities above that represent less than 1% of the number of cities and were not plotted for better visualization of the histogram. In orange, the range that contains the selected cities.

TABLE III: Metric comparison to competitor methods Baseline, In-domain, Our Methodology (VGG16) and Our Methodology (ResNet50). Best results in bold.

				Validation			Test		
	Extraction Method	Pre-Training Dataset	Fine-tuning	MAE	RMSE	R^2	MAE	RMSE	R^2
Baseline	Multi-Feature Process	-	-	129,19	182,21	**0,67**			
In-domain	ResNet50	ImageNet, BigEarthNet	×	173.97	231.16	0.057	186.32	277.43	0.36
	ResNet50	ImageNet, EuroSat	×	167.86	224.35	0.13	192.80	282.06	0.34
	ResNet50	ImageNet, RESISC-45	×	167.90	224.30	0.13	192.58	281.70	0.34
	ResNet50	ImageNet, So2Sat	×	160.25	207.93	0.17	222.12	310.58	0.20
	ResNet50	ImageNet, UC Merced	×	163.01	214.89	0.15	180.12	268.08	0.41
Our methodology	VGG16	ImageNet	✓	127.89	183.98	0.44	198.61	303.35	0.24
	ResNet50	ImageNet	✓	**119.80**	**165.72**	0.48	**173.35**	**242.15**	**0.51**

D. Results

We extracted features from the 14,000 images in our dataset. Using VGG16, each image yielded a feature tensor of size 7x7x512, while ResNet50 produced a feature tensor of size 7x7x2048. Since we had 100 images per city, the features were grouped by average to form a feature vector representing each city. With 140 feature vectors, we used ElasticNet regressor, combining penalty indices L1 and L2. The feature vectors worked as inputs to the model; the task was to predict the

TABLE IV: Metrics obtained for each set of ElasticNet hyperparameters. Best results in bold.

Alpha	L1 Ratio	Mean RMSE	Mean MAE	Mean R^2
0.1	0.0	169.43635832	122.93069492	0.45513856
0.1	0.05	**165.72463176**	**119.80087859**	0.48877279
0.1	0.5	174.09002162	123.46320746	0.43509703
0.1	0.7	176.7737528	127.35984324	0.42881993
0.1	0.9	177.63336194	127.28701723	0.39309232
0.1	1.0	171.63107498	126.10964626	0.45020572
1.0	0.0	172.40698062	130.90810401	0.42109313
1.0	0.05	169.77596639	124.92073417	0.46785733
1.0	0.5	171.27794511	124.55082129	0.49029491
1.0	0.7	174.1916442	126.11727016	0.47533054
1.0	0.9	174.81568303	124.58423506	0.47663267
1.0	1.0	186.18446589	134.00251887	0.39630227
10.0	0.0	166.92212153	123.07594144	0.47916583
10.0	0.05	167.78712859	124.32568897	**0.49089806**
10.0	0.5	173.0992524	125.57343898	0.47858027
10.0	0.7	175.85295585	126.95128614	0.46143682
10.0	0.9	177.82726307	126.94112396	0.45366119
10.0	1.0	191.21809159	137.90622246	0.34511397
100.0	0.0	167.86667061	125.93733564	0.4817452
100.0	0.05	170.29064729	126.75586143	0.48456854
100.0	0.5	204.21454958	149.51259664	0.24532264
100.0	0.7	217.30147028	159.93083819	0.13205345
100.0	0.9	228.06798478	169.77264318	0.03324379
100.0	1.0	238.54644352	178.5990335	-0.02512721

average monthly income for each city.

We split the feature vectors into training and testing sets, with 85% of the data used for training and 15% for testing. The testing set was not used during the training process, but only for evaluating the final performance. Fitting used cross-validation with 10 folds, dividing the dataset into 10 equal parts (folds), using 9 folds for training and the remaining fold for validation. The performance metrics corresponded to the average across the 10 folds. We tested multiple hyperparameters for ElasticNet, with values of alpha: 0.1, 1, 10, and 100; and of L1 ratio: 0, 0.05, 0.5, 0.7, 0.9, and 1 – the best ones were alpha 0.1 and L1 ratio 0.05, as presented in Table IV.

Table III presents our main results. We present metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R^2 obtained for each of the comparison settings. In the table, in bold, the best results correspond to our methodology using the fine-tuned ResNet50 architecture.

By implementing the fine-tuning methodology in a ResNet50 architecture, we observed improvements in the MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error) metrics compared to the baseline methodology, the fine-tuned VGG16 model, and all the non-fine-tunned models. Specifically, with ResNet50, we obtained a MAE of 119.80 and an RMSE of 165.72.

MAE measures the average absolute difference between the predicted values and the actual values. A MAE of 119.80 indicates that the predicted values differ from the actual values by approximately R\$ 119.80 ($\sim$U\$ 24.00). A lower MAE signifies a better fit of the regression model, as it suggests that the model's predictions are closer to the actual values.

RMSE, on the other hand, measures the standard deviation of the residuals, which are the differences between the predicted and actual values. An RMSE of 165.72 indicates that the predicted values deviate from the actual values by approximately R\$ 165.72 ($\sim$U\$ 33.00). Similar to MAE, a lower RMSE indicates a better fit of the model, with predictions that are closer to the actual values.

Although ResNet50 had slightly lower results than VGG16 for predicting radiance levels (Section III-C), the features extracted by ResNet50 from the daytime images yielded pronouncedly better results for predicting the average monthly income of cities.

Discussion

Our methodology not only surpasses the performance of the baseline work of Castro and Alvarez [5], but it also introduces a simpler methodology that is less computationally-expensive. While Castro and Alvarez propose a complex methodology, containing 4 feature extraction processes, our work uses a single fine-tuned CNN that is straight to build, train, and employ.

V. Conclusions

This article presents a proposal for predicting the average monthly income of Brazilian cities using features extracted from satellite imagery through Machine Learning. Our dataset consists of 140 small cities distributed throughout Brazil, labeled according to three levels of nighttime light radiance (low, medium, and high). We use this data to fine-tune CNN models that, in turn, became adjusted to extract features that reflect the urban characteristics of satellite imagery.

By employing these features into a Regularized Linear Regression model, we demonstrated that such features have the potential to predict socioeconomic metrics from satellite imagery. This finding is evidence of our hypothesis that such imagery captures human-built structures that, accordingly, correlate to socioeconomic indicators.

Using a ResNet50 model fine-tuned over NTL radiance, we extracted features that supported the prediction of the average monthly income of Brazilian cities with Mean Absolute Error (MAE) of R\$ 119.80 ($\sim$U\$ 24.00), which deviates 65.71% from the mean of the data.

By leveraging the power of machine learning and satellite imagery, we accurately predicted socioeconomic factors, paving the way for informed decision-making and targeted interventions. In particular, our methodology can be employed to regions of the world in which census data is not available, leaving them less sensitive to public politics and philanthropic endeavors. Future directions include medium and large cities, the prediction of other indicators, and the exploration of other Machine Learning models.

ACKNOWLEDGMENT

This work was supported by Brazilian agencies Fapesp (grant 2016/17078-0), CNPq (grant 307946/2021-5), and INCT Instituto Nacional de Inteligência Artificial (406417/2022-9).

REFERENCES

[1] World Bank. *Brazil Poverty and Equity Assessment: Looking Ahead of Two Crises*. World Bank, 2022.

[2] World Bank. *Poverty and shared prosperity 2022: Correcting course*. The World Bank, 2022.

[3] Christopher M Bishop and Nasser M Nasrabadi. *Pattern recognition and machine learning*, volume 4. Springer, 2006.

[4] Marshall Burke, Anne Driscoll, David B Lobell, and Stefano Ermon. Using satellite imagery to understand and promote sustainable development. *Science*, 371(6535):eabe8628, 2021.

[5] Diego A Castro and Mauricio A Álvarez. Predicting socioeconomic indicators using transfer learning on imagery data: an application in brazil. *Geojournal*, 88(1):1081–1102, 2023.

[6] Satellite Imaging Corporation. Quickbird satellite sensor. https://www.satimagingcorp.com/satellite-sensors/quickbird/. Accessed: 2023-06-24.

[7] Juan C Duque, Jorge E Patino, Luis A Ruiz, and Josep E Pardo-Pascual. Measuring intra-urban poverty using land cover and texture metrics derived from remote sensing data. *Landscape and Urban Planning*, 135:11–21, 2015.

[8] Christopher D Elvidge, Kimberley E Baugh, Eric A Kihn, Herbert W Kroehl, Ethan R Davis, and Chris W Davis. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. *International Journal of Remote Sensing*, 18(6):1373–1379, 1997.

[9] Christopher D Elvidge, Jeffrey Safran, Benjamin Tuttle, Paul Sutton, Pierantonio Cinzano, Donald Pettit, John Arvesen, and Christopher Small. Potential for global mapping of development via a nightsat mission. *GeoJournal*, 69:45–53, 2007.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[11] Neal Jean, Marshall Burke, Michael Xie, W Matthew Davis, David B Lobell, and Stefano Ermon. Combining satellite imagery and machine learning to predict poverty. *Science*, 353(6301):790–794, 2016.

[12] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[13] Ben Paul Mungyereza. Uganda national panel survey 2011/2012–wave iii report. *Gates Open Res*, 3(335):335, 2019.

[14] Congresso Nacional. Draft budgetary plan – PLOA 2022 (PL nº 19/2021-CN). https://www2.camara.leg.br/orcamento-da-uniao/estudos/2021/subsidios-a-apreciacao-do-projeto-de-lei-orcamentaria-ploa-para-2022, 2021. Accessed: 2023-06-01.

[15] Maxim Neumann, Andre Susano Pinto, Xiaohua Zhai, and Neil Houlsby. In-domain representation learning for remote sensing. *arXiv preprint arXiv:1911.06721*, 2019.

[16] National Oceanic and Atmospheric Administration (NOAA). Defense meteorological satellite program (dmsp). https://www.ospo.noaa.gov/Operations/DMSP/index.html, 2021. Accessed: 2023-06-13.

[17] National Oceanic and Atmospheric Administration (NOAA). Visible infrared imaging radiometer suite (viirs). https://ngdc.noaa.gov/eog/viirs/, 2021. Accessed: 2023-06-13.

[18] Union of Concerned Scientists. UCS satellite database. https://www.ucsusa.org/resources/satellite-database, 2022. Accessed: 2023-06-01.

[19] Brazilian Institute of Geography and Statistics. Municipal mesh. https://www.ibge.gov.br/en/geosciences/territorial-organization/territorial-meshes/18890-municipal-mesh.html, 2021. Accessed: 2023-06-13.

[20] Brazilian Institute of Geography and Statistics. Postponement of the demographic census. https://www.ibge.gov.br/novo-portal-destaques/30569-adiamento-do-censo-demografico.html, 2021. Accessed: 2023-06-01.

[21] Brazilian Institute of Geography and Statistics. Population census. https://www.ibge.gov.br/en/statistics/social/population/22836-2020-census-censo4.html?edicao=35957&t=resultados, 2022. Accessed: 2023-06-13.

[22] Brazilian Institute of Geography and Statistics. *Synthesis of Social Indicators: An analysis of the living conditions of the Brazilian population*. IBGE, Rio de Janeiro, 2022.

[23] Google Maps Platform. Google static maps api. https://developers.google.com/maps/documentation/maps-static/overview, 2023. Accessed: 2023-06-13.

[24] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252, 2015.

[25] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[26] Fundação Getulio Vargas. Where are located the rich people in brazil? https://www.cps.fgv.br/cps/bd/docs/ranking/TOP_Municipio2020.htm#e, 2020. Accessed: 2023-06-13.

[27] Michael Xie, Neal Jean, Marshall Burke, David Lobell, and Stefano Ermon. Transfer learning from deep features for remote sensing and poverty mapping. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30, 2016.