Poverty Prediction with Satellite Imagery and Deep Learning

Chiyuan Cheng

August 2020

Data Science Career Track, Springboard

1. Introduction

Mapping the spatial distribution of poverty in developing countries is an important but costly challenge. According to World Banks, although poverty in several areas around the world have recently declined, many countries in sub-Saharan Africa still lag behind in terms of poverty eradication (Figure 1). Moreover, COVID-19 is likely to cause the first increase in global poverty since 1998 [1]. Sub-Saharan Africa is predicted to be the region hit hardest in terms of increased extreme poverty due to COVID-19. However, the major challenge in fighting poverty is the lack of reliable and accurate socioeconomic data, which is time-consuming, labor-intensive, and expensive to collect by the traditional household surveys. It is also difficult to collect up-to-data data to guide the policy agents to allocate the limited resource in the right places.

Recently, major advancement in computer visions research and increasing accessibility of geospatial data have enabled new technology to estimate socioeconomic indicators. Previous work by Jean et al in 2016 has introduced deep learning methods to estimate economic livelihood indicators in five African countries from satellite imagery [2]. The idea was to use a neutral network to train the nighttime light intensity, that is correlated with urban developments and economic activities, and to predict the nighttime light intensity from the daytime satellite imagery. They also revealed that the transfer learning was useful in predicting asset wealth within the poorest segment of the population, when compared to the established method. Since then, several researches have been conducted to extract socioeconomic information from satellite imagery using similar approaches [3,4,5]. In this respect, the geospatial deep learning methods can be useful to improve efficiency and optimization for traditional household survey methods. However, significant works are still needed before the predictive models can fully replace the ground-based surveys. To further tackle the challenges of poverty eradication building on Jeal et al [2], we combine transfer learning and computer vision with the geospatial information as a fast and cost-saving tool to provide poverty prediction.

In this work, we investigate the extent to which geospatial data, including nighttime light luminosity, daytime satellite imagery, and cluster sourced wealth information, can be used to estimate the socioeconomic well-being in a sub-Sahara country, Burundi. The goal of this work is to answer the following questions.

1. Is it possible to extend the deep learning approach to estimate poverty levels in poor countries with limited training data?

2. Can we fine tune the existing pre-trained models or leverage on the easy-to-access variables, such as nighttime light intensity data, to achieve better performance?

3. What is the predictive power of such a model, when tested on the same country but in a different year, or a completely new country?

Specifically, we built and evaluated regression models to predict the wealth index from the nighttime luminosity. We further classified the daytime satellite images to three classes based on their nighttime luminosity by unsupervised learning. Finally we combined the transfer learning and neural networks to predict the poverty levels based on the daytime satellite imagery.

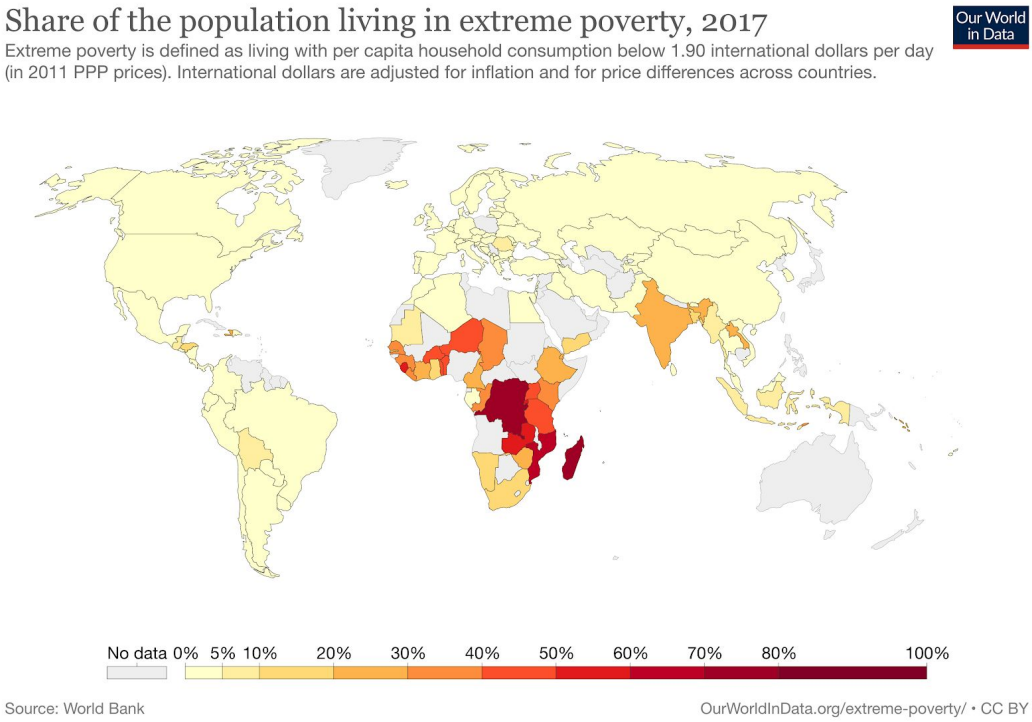

Figure 1. Poverty map (2017)

2. Data Acquisition and Cleaning

2.1 Data Acquisition

2.1.1 Economic Variables

- **Wealth Index:** We used the 2010 Demographic and Health Survey (DHS) in Burundi as a measure of ground truth for the socioeconomic indicators. We focused on the "Wealth Index" or "Wealth Index Factor Score" from the DHS dataset. The primary measure of socioeconomic well-being is the "Wealth Index", which is computed as the first principal component of attributes related to common asset ownership on a per-household level. We transformed the "Wealth Index" to the range [0,1] by the sklearn MinMaxScaler module. We then calculated the median value of wealth index per cluster, as it reported in the DHS dataset.
- **Education Completed:** The DHS captures information on the number of years of education completed by household members over 6 years old. This value was aggregated by the total value across all households per cluster.
- **Access to Electricity.** The DHS dataset contains information on the number of affirmative responses to the survey question related to access to electricity. This value was aggregated by the total value across all households per cluster.
- **Access to water.** The DHS dataset contains information on the total travel time in minutes to get to the water source. If the water source is on-site, time is set to zero. We get the total time to access a water source across all households per cluster.
- **Access to cellphone.** The DHS dataset contains information on the total number of cellphone per household. This value was aggregated by the total number per cluster.
- **HIV blood test result.** The DHS dataset contains information on the number of people who have received HIV blood tests. This value was aggregated by the total number per cluster.

2.1.2 Nighttime Luminosity

The nighttime luminosity (or light intensity) data is taken from the satellite imagery in the National Centers for Environmental Information 2010. The data includes a continuous luminosity level from 0 to 63 for the Burundi, with 0 being the darkest pixel.

2.1.3 Daytime Satellite Imagery

We retrieved satellite images per cluster based on the cluster centroids reported in the DHS dataset, where the location of each cluster is defined by the mean latitude and longitude of the households. We used Google Static Maps API to download a total of 50,000 images with a pixel resolution of approximately 2.5 meter. The size of each image is 400 pixel x 400 pixel, and it matches the area covered by a single pixel of nighttime image data, which covers 0.25 km.

2.2 Data Wrangling and Cleaning

For DHS survey data, we compute the median wealth index for each cluster. The resolution of each pixel in the nighttime satellite image is about 1 km, and we used 10 pixels x 10 pixels to take average on the luminosity of each cluster. We then merged the mean nighttime luminosity and wealth index at the cluster level. For the regression model to wealth index from luminosity, we used power transform on the wealth index prior to the modeling. The inverse transform was performed to obtain the predicted wealth index.

We implemented the satellite based deep learning approach proposed by Jean et al with the assumption that nighttime light luminosity can act as a good proxyl for economic activity. We used a pre-trained VGG16 model (Figure 2) as a filter with the goal of learning features from imagery that are useful for poverty prediction. We also fine-tuned effective pre-trained models (ResNet50 and InceptionV3) to get better performance. For deep learning, we treated the problem as a classification task with 3 night time intensity classes: low, medium, high, which was classified based on the Gaussian mixture model. We used 80% of images for training and 20% of images for testing. The class imbalance was treated by downsampling the majority class (high nighttime intensities), such that each class contains 200 images.

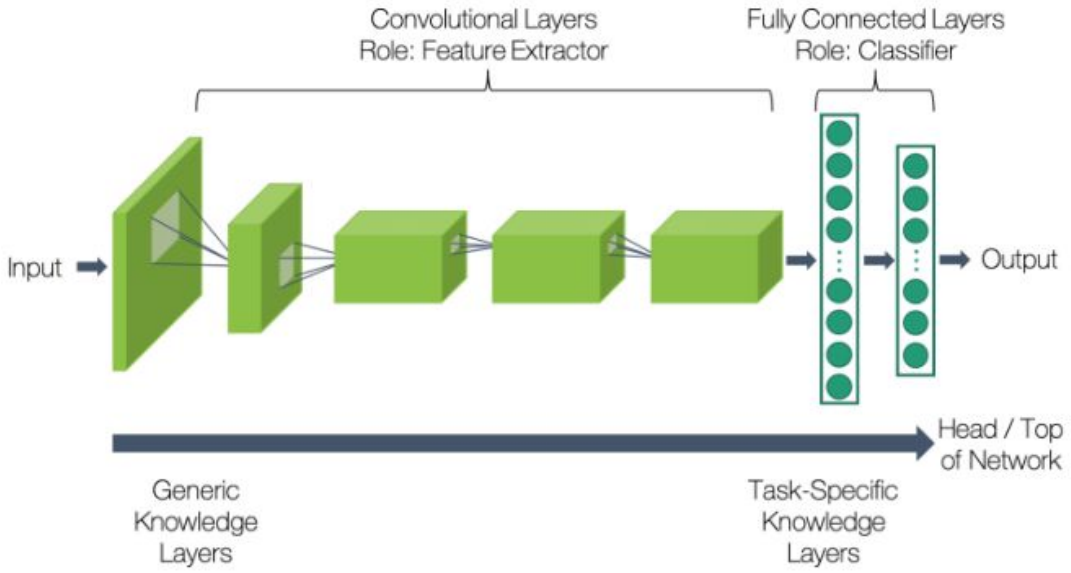

Figure 2. VGG16 architecture.

3. Data Exploration and Machine Learning

3.1 Nighttime Imagery and Wealth Index

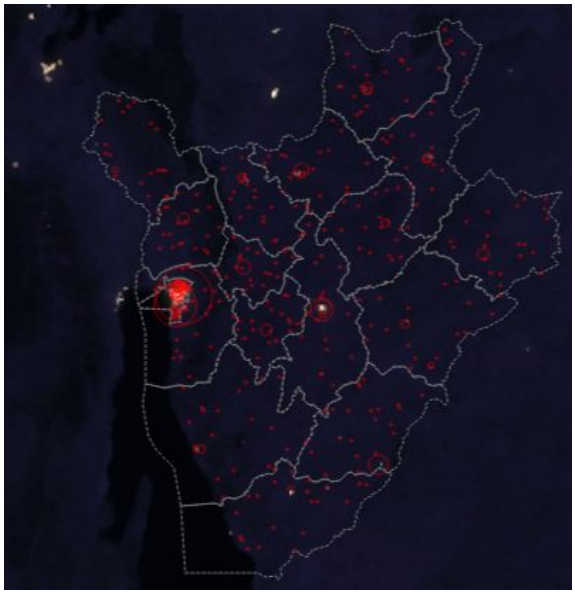

Figure 3. The median wealth index overlaid on the nightlight satellite imagery of Burundi (2010). The size of the red circle represents the wealth index value.

The median value of wealth index for each cluster superimposed with the nighttime satellite imagery of Burundi, shown in Figure 3. The initial observation is that the wealth index visually matches very well with the light luminosity, with bright areas generally having higher values of wealth index. However, the majority of the image is dark. After analyzing the luminosity data from the nighttime image, we found 73% of the area has zero luminosity.

3.2 Distribution of Wealth Index

Next, we look at the distribution of wealth index and compare the value with zero luminosity and when luminosity is greater than zero (Figure 4). We can see that when the areas are completely dark in the nighttime (luminosity is zero), their wealth indexes are close to zero. When the luminosity is greater than zero, the value of wealth index increases. But the distribution of wealth indexes does not follow a Gaussian distribution. Therefore, a power transform (Yeo–Johnson transformation) is applied to the wealth index prior to the machine learning model.

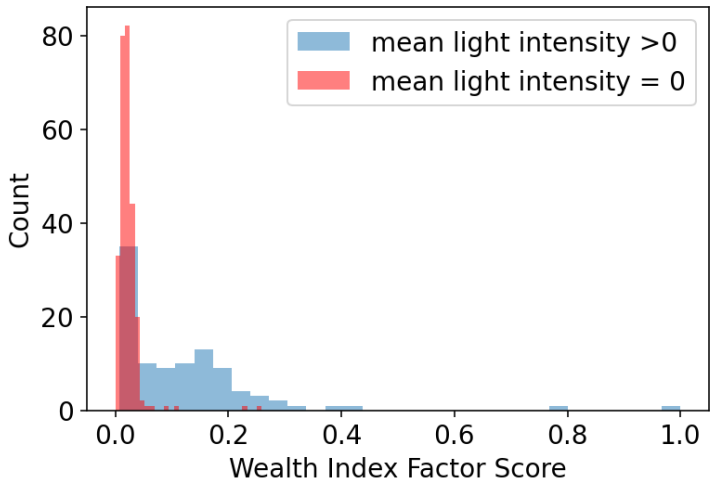

Figure 4. Distribution of wealth index of Burundi (2010)

3.3 Correlation between Nighttime Luminosity and Economic Indices

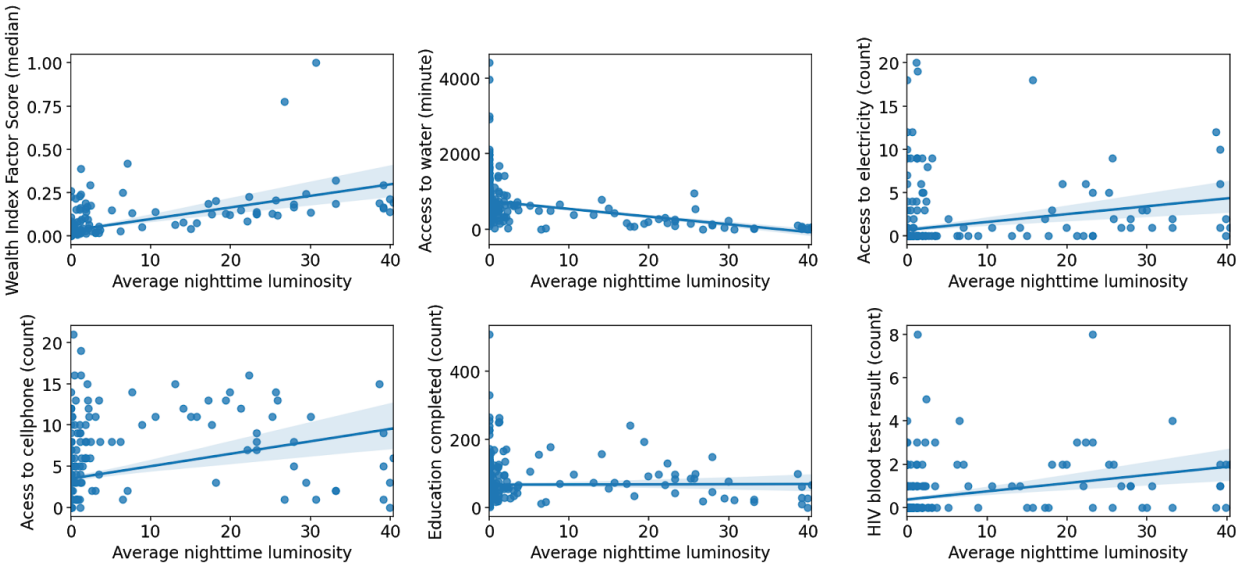

Figure 5. Scattering plots of various economic indicators (wealth index, access to water, access to electricity, access to cellphone, education completed, and HIV blood test result) per cluster from DHS data v.s. average nighttime luminosity of Burundi in 2010.

The relationship between nighttime luminosity and various economic indexes from the DHS dataset aggregated at the cluster level was compared (Figure 5). Figure 5 indicates that all the economic indicators do not have a clear linear relationship with luminosity. But the general trend is that in the area with higher luminosity, where the economic activities are considered to be high, they also have high accessibility of water, electricity, cellphone, education, and HIV tests. However, it was reported that these indicators can not be expected to have a clear relationship with the nighttime when evaluating with machine learning models [6]. One reason is that clouds covered on the sky in the nighttime satellite imagery may cause a bias on the model.

3.4 Regression Model to Predict Wealth Index from Nighttime Luminosity

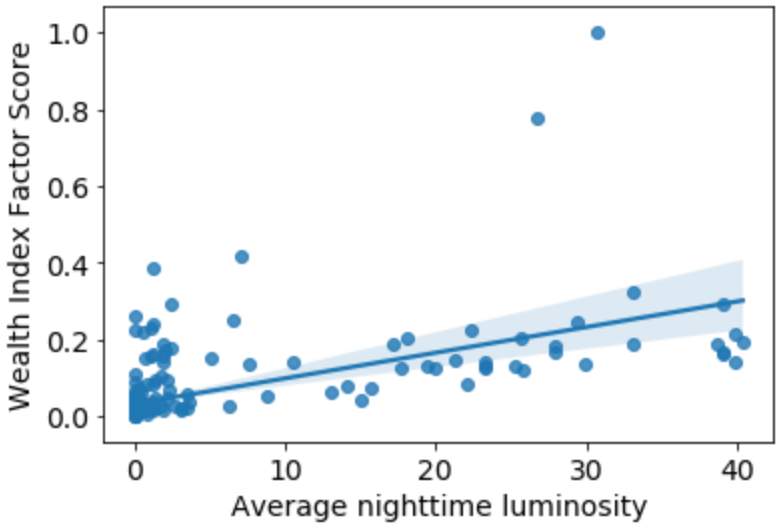

Figure 6. Wealth index vs nighttime luminosity

Table 1. Regression results for wealth index prediction from luminosity.

Regression model	R^2 (test)
Linear regression	0.50
Lasso	0.50
Rigid	0.50
Random forest	0.54

Previous studies have used machine learning methods to predict wealth in sub-Saharan African countries, as well as non-African countries from the nightime satellite imagery [2,3,4]. In general, the predictive models reported in the literature achieved r-squared values ranging from 0.51 to 0.75, depending on the countries and models .

In this work, we compared different regression models (Table 1) to predict the wealth index from nighttime light luminosity in Burundi. Although we used a power transform on the wealth index prior to models, the regression models do not have a good fit in terms of r-squared, with the best score of 0.54 from Random Forest Regressor. We then used the best predicted wealth index to reconstitute the poverty map at cluster level (Figure 7).

Although the predicted performance is not good, we can still virtually see the predicted wealth indices (Figure 7b) have a good match with the ground-truth wealth indices (Figure 7a) on the map.

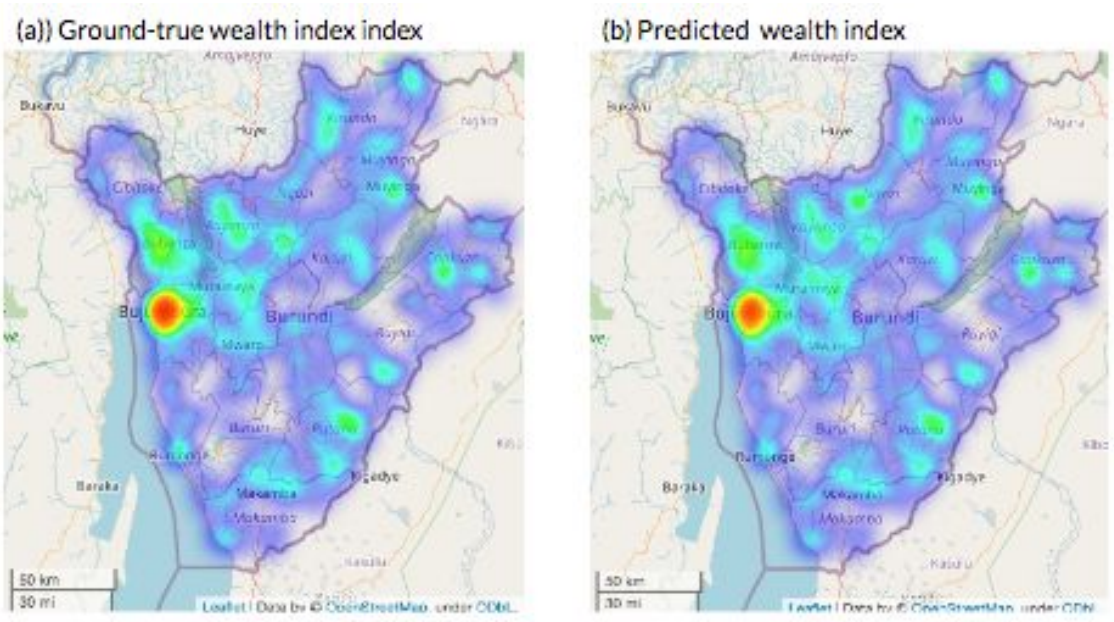

Figure 7. Ground-truth wealth indices and the predicted wealth indices from nighttime satellite imagery using machine learning aggregated to the cluster level.

3.4 Visualization of Daytime Satellite Imagery from Nighttime Luminosity

Although the predicted wealth index has a good visual match with the group truth value on the map, the major problems of this method are 1) the cloud covered on the sky can cause model bias, and 2) 73% of the luminosity values are zero. To improve the method, we further use daytime satellite imagery to see if deep learning can help capture the feature from the image and provide a good predictive performance. A Gaussian Mixture model (GMM) was used to classify the daytime satellite imagery into three classes based on their corresponding nighttime luminosity i.e. (high, medium, low) [7]. The GMM classified zero luminosity to be 'low' class, luminosity between 1 and 9 to be ''medium' class, and luminosity between 10 and 63 to be 'high' class (Figure 8). The result show that images related to building and roads as important for classifying high and median night light intensity, whereas images related to trees and lands are given more weight in the low nighttime intensity class.

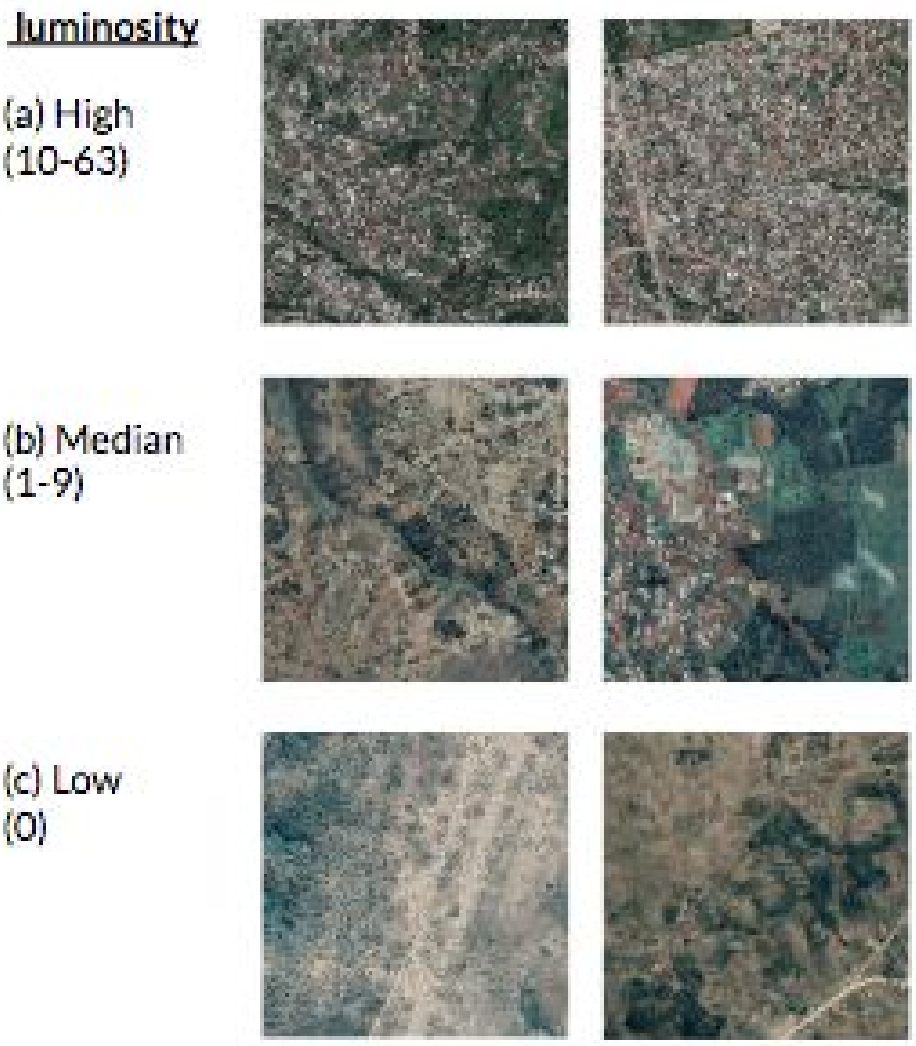

Figure 8 Daytime satellite imagery in the nighttime light classification task

To ensure each class has sufficient difference, we measure the pairwise similarity between two classes using the Bootstrapping sampling method (Figure 9). The result suggests that average similarity between low and median class is around 0.65, while similarity between median and high class is around 0.5.

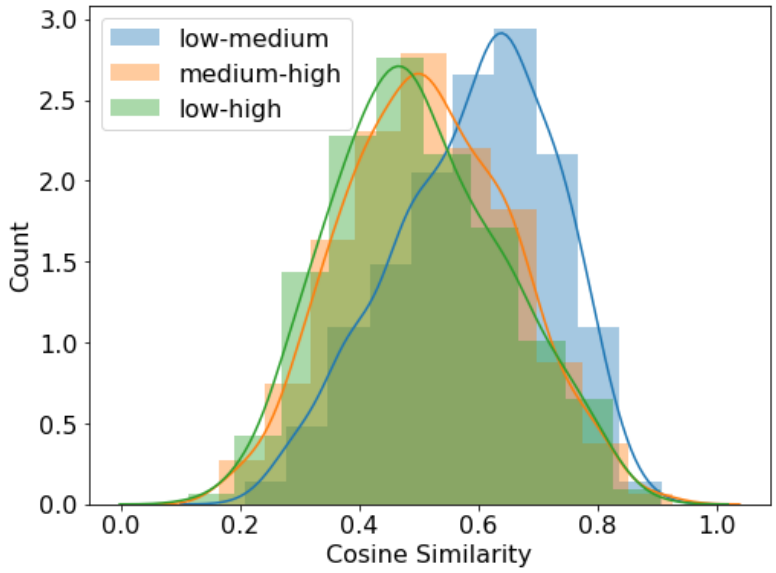

Figure 9. Pairwise cosine similarity between two classes classified by GMM

3.6 Deep Learning to Predict Poverty from Daytime Satellite Imagery

For deep learning, we tried a basic CNN model, as well as pre-trained models, including VGG16, ResNet50, and Inception V6, to predict the poverty classes from daytime satellite imagery.

3.6.1 CNN model

We trained a basic CNN model (with architecture similar to LeNet-5 [ref], which has two Convolutional layers and one max_pooling layer) to predict the nightime light intensity classes from daytime satellite images. A L2 regularization and 30% dropout were included to generalize the model. The model was trained for 100 epochs and plotted the accuracy and loss for both training and test set. The result in Figure 10a shows training accuracy is higher than the test accuracy, suggesting overfitting. The CNN model was generated by Image augmentation on the training set to avoid overfitting. The accuracy of training and test set reaches to approximately 60% after ~80 epochs (Figure 10b).

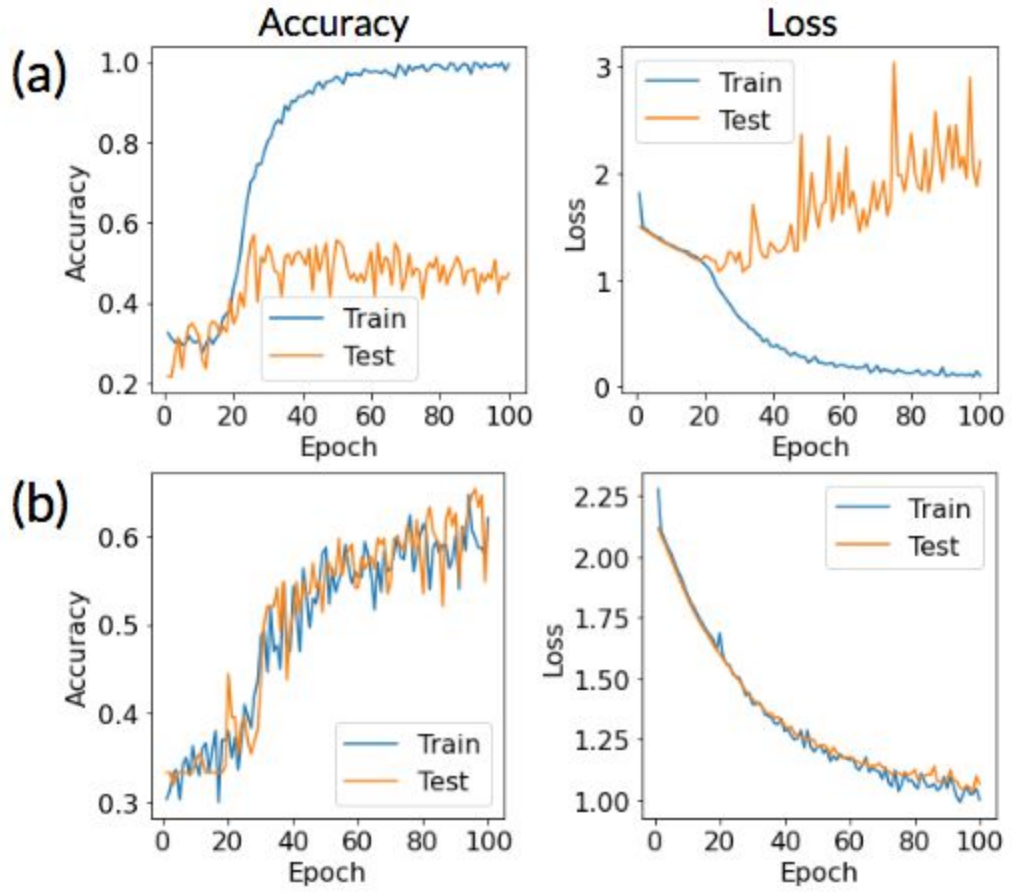

Figure 10. Model performance for CNN model (a) without augmentation and (b) with augmentation.

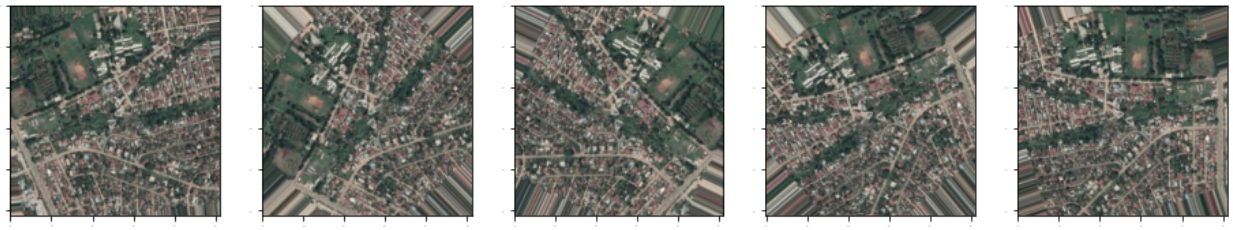

Figure 11. Image Augmentation (rotation_range=45, horizontal flips=True)

3.6.2 Transfer Learning and Feature Extraction (VGG16)

Next, we used a sequence of transfer learning steps and Dense layers to overcome the problems of limited training data. The pre-trained model, VGG16, was leveraged to extract features of satellite images. The high-level representations of the images are

useful for the task of interests, such as extracting the socioeconomic data for poverty mapping. For example, the first few Convolutional layers of VGG16 can learn low-level features, such as edges (Figure 12). The similar strategies have proven quite successful in the past [2,3,4].

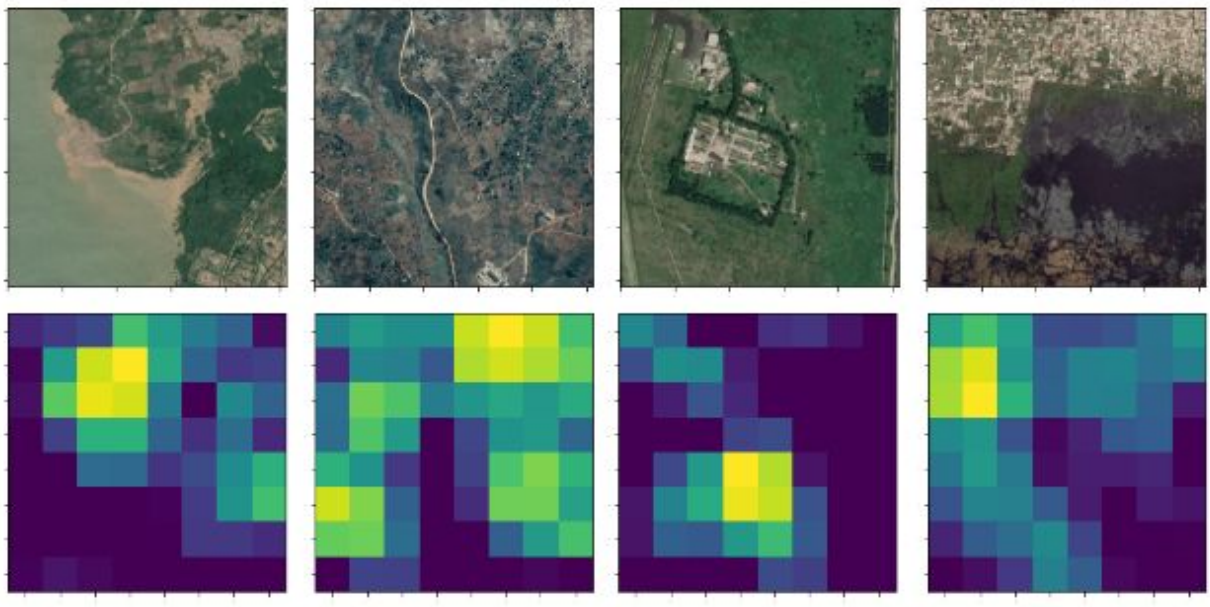

Figure 12. Visualization of features. Four daytime satellite images from Google Static Map (top) with their convolutional filters extracted from VGG16 (bottom). Images by column, from left to right, the features corresponding to lake areas, roads, urban areas, and non-urban areas) in the convolutional neural network model used for feature extraction.

Figure 13 presented the VGG16 model performance. It shows that a significant overfit was observed without image augmentation (Figure 13a). After augmentation of training images and fine-tunes the last two layers of the VGG16 model, it alleviates the overfit and increases the accuracy of training to 70-80% (Figure 13b).

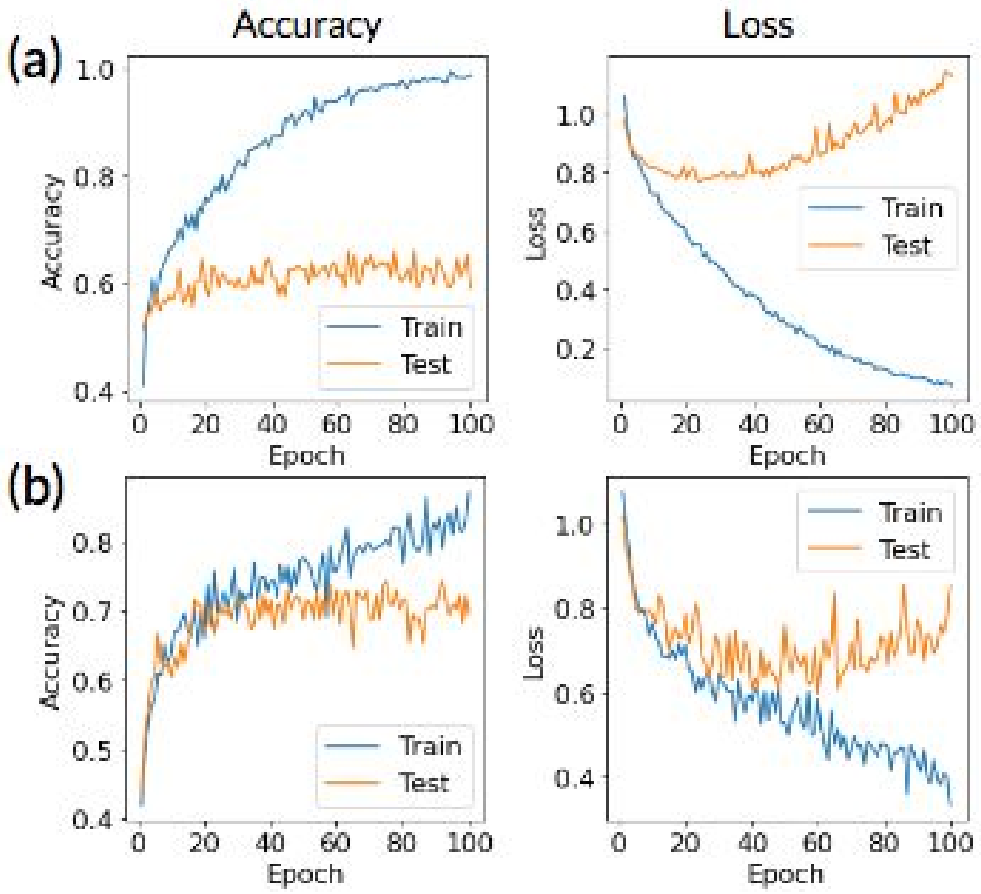

Figure 13. Model performance for VGG model (a) without augmentation and (b) fine-tuned with augmentation.

3.6.3 Transfer Learning and Fine-tune Models (ResNet, Inception)

Two effective pre-trained models (ResNet50 and Inception V3) were further used to fit the satellite image data with image augmentation and model fine-tuning. The results show in Figure 14. The accuracy of Inception V3 can reach to about 80%, while the accuracy of ResNet50 is about 60%.

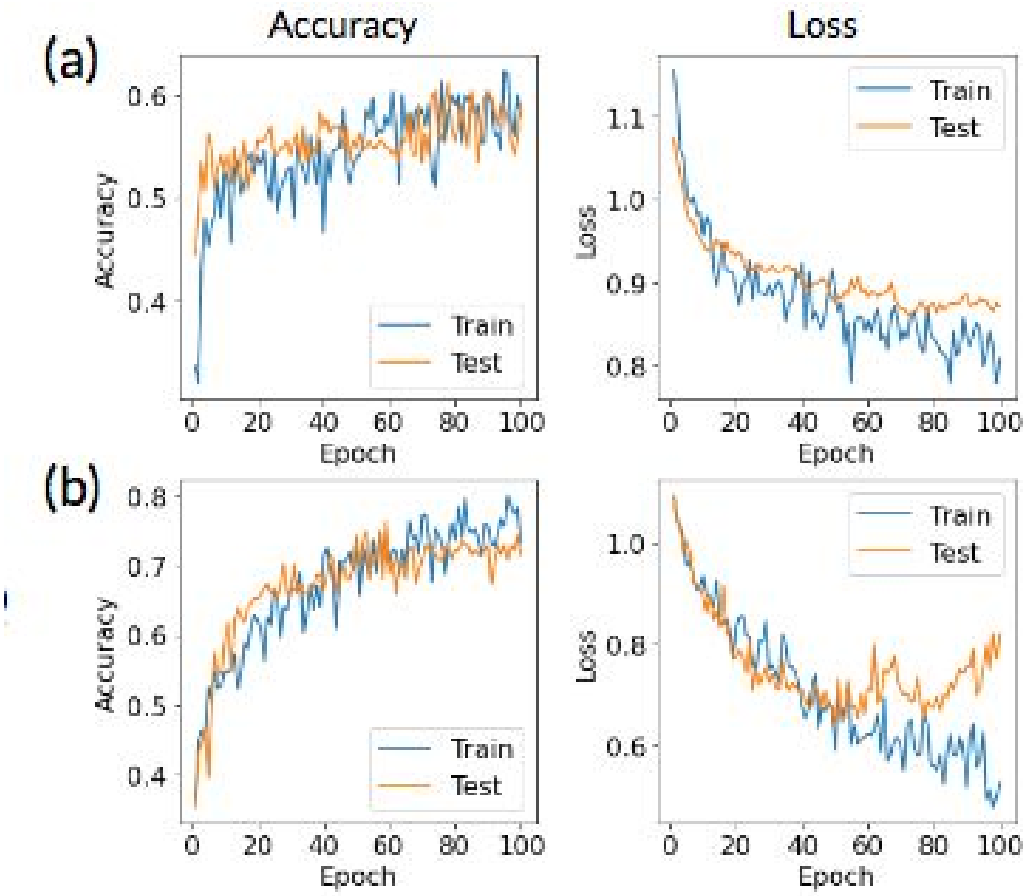

Figure 14. Model performance of (a) ResNet50 and (b) Inception V3 with fine-tuning and augmentation.

3.7 Evaluation of Model Performance

By comparing various models with different hyperparameters and fine-tuning, the Inception V3 with augmentation (Aug) and fine-tuning (FT) generally has a better performance, with the training accuracy of 81% and test accuracy of 72%.

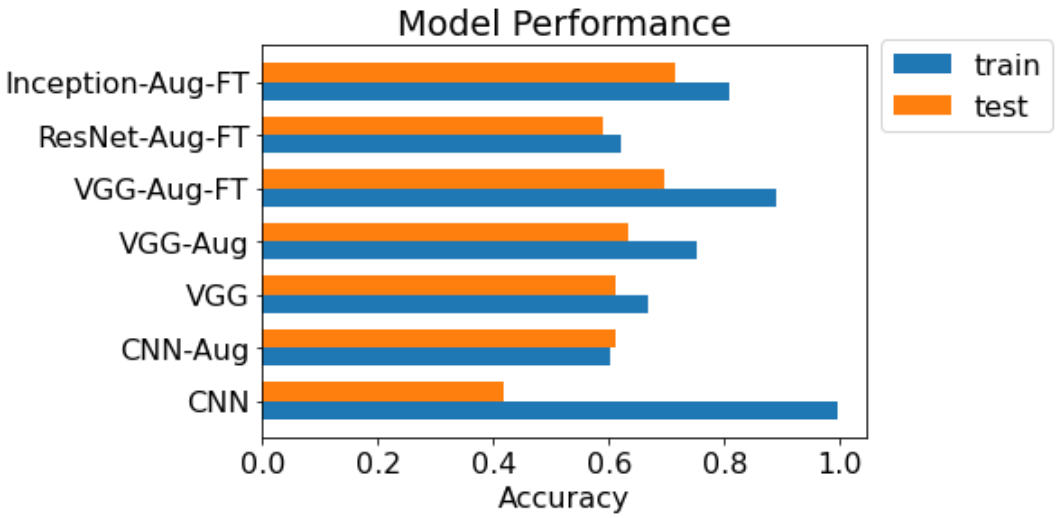

Figure 15. Model performance with various models.

4. Limitation and Future Work

Using the combination of nightlights and predictions from the proposed models may yield further predictive improvements. Here, we only trained models to predict wealth index in a single year of a sub-Saharan African country. We could extend the capabilities by training the model in one year and use the model to do the prediction in different years. We can further apply the model to the test countries where the model has not seen before. This will also open up unique opportunities to obtain economic indicators over time in developing countries, which typically requires expensive household surveys.

5. Conclusion

In this work, we implemented the satellite-based deep learning approach in a sub-Saharan African country, Burundi, in 2010. Our results confirm the applicability of the methodology, with the best regression model achieving an r-squared of 0.54 for estimating wealth index from nighttime luminosity. We also demonstrate that a combination of transfer learning and neural networks can be implemented to capture the features from daytime satellite images to predict poverty, with the best model achieving 80% accuracy.

6. References

1. https://blogs.worldbank.org/opendata/impact-covid-19-coronavirus-global-poverty-why-sub-saharan-africa-might-be-region-hardest
2. Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., Combining satellite imagery and machine learning to predict poverty, Science, 2016, 353(6301), 790-794 .
3. Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., Burke, M., Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. Nature Comm, 2020, 11, 2583
4. Gebru, T., Krause, J., Wang, Y., Chen, D., Deng, J., Lieberman, E.A. and Li, F. Using deep learning and Google Street View to estimate the demographic makeup of neighborhoods across the United States. PNAS, 2017, 114(50) 13108-13113; first
5. Bruederie, A., Hodler, R., Nighttime lights as a proxy for human development at the local level, PloS One, 2018, 13(9), e0202231.
6. Head, A., Manguin, M., Tran, N., Blumenstock, J.E. Can human development be measured with satellite imagery? Proceeding of the Ninth International Conference on Information and Communication Technologies and Development, 2017, 8.
7. Kiana, E., Homayouni, S., Sharifi, M.A., Farid-Rohani, M., Unsupervised change detection in SAR images using Gaussian mixture models. The International Archives of the Remote Sensing and Spatial Information Science., 2015, XL-1/W5. 407.