A Critical Paper Review

Henning Heyen

11 November 2022

Abstract

This essay analyses the paper "Predicting Livelihood Indicators from Community-Generated Street-Level Imagery" by Lee et al. [4]. The focus of this essay is to critically assess the approach in [4] with respect to the 17 Sustainable Development Goals (SDGs) proposed by the United Nations in 2015 and its impact on sustainability in general.

1 Description of dataset and baseline

1.1 Chosen dataset and model

The underlying dataset that is used to train the model is the *Mapillary* dataset. It is a "global citizen-driven street-level imagery database" [6]. Anyone can can upload an image using a smartphone for instance. The image needs a timestamp and GPS data, which most device capture automatically. The dataset "doubled in size from 500 million images to 1 billion" [4] in eight months and is expected to continue to grow exponentially.

Using the Mapillary dataset the paper aims to accurately predict livelihood indicators in Kenya and India as an "inexpensive, scalable, and interpretable"[4] alternative to survey data. These indicators include *Poverty, Women's Body Mass Index* (BMI) and *Population.* The label space comes from the most recent *Demographic and Health Survey* (DHS) in 2015-2016 for India and 2014 for Kenya. BMI data was not available for Kenya in 2014, so BMI predictions were only made for India. The DHS data is clustured to preserve privacy. Households within a 5km radius share the same geographic coordinates. The Mapillary images were clustered accordingly (7.117 clusters and 1.121.444 images from India and 1.071 clusters and 156.756 images from Kenya). The metrics are classification accuracy and the Pearson's r^2 for regression. Class labels are generated through splitting by the median.

There are three approaches to the prediction task. First, **Image-wise Learning** maps every single image in a cluster to the label space. The model is based on a ResNet architecture and was used as baseline in [9] before. The second approach, **Cluster-wise Learning**, utilizes object counts from street-level images in a cluster using a segmentation model proposed in [8]. The classifier or regression model is trained on cluster-level object counts to predict the indicators. The third approach is a Graph Convolutional Network (**GCN**) that learns "relationships between images, representing a cluster as a graph, where image-based features serve as nodes connected by edges encoding their spatial distance"[4].

In India, the **GCN** performed fairly well in predicting poverty (accuracy: $81,34\%$, r^2: $0,54$), population (accuracy: $94,71\%$, r^2: $0,89$) and women's BMI (accuracy: $89,56\%$, r^2: $0,57$). Similar performance was achieved by **Cluster-wise Learning** in Kenya. Results for **Image-wise Learning** were similar but less effective in both countries.

1.2 Assessment of documentation

Both the code and the underlying datasets are publicly available. In comparison to similar work such as [2], which is not open source, replicating the results in [4] is much easier. The Github repository[1] is well documented and provides helpful guidelines. However, the paper lacks explanation on how data is labled and collected. Poverty and women's BMI are labled using the DHS survey but details of the data collection methods are completely missing. A more in depth explanation can be found in [11] or on the DHS website[2]. The population indicator is labeled using Facebook's High Resolution Population Density Maps. The paper as well as the github page miss information on labeling techniques and the dataset itself. More information on Facebook's High Resolution Population Density Maps can be found here[3].

[1] https://github.com/sustainlab-group/mapillarygcn
[2] https://dhsprogram.com/Data/Guide-to-DHS-Statistics/index.cfm
[3] https://dataforgood.facebook.com/dfg/tools/high-resolution-population-density-maps

2 Connection to SDGs

2.1 Related goals and targets

As [4] aims to predict the poverty landscape in India and Kenya it directly addresses **SDG 1: No Poverty**. In particular, understanding the wealth distribution in those countries can help "implement nationally appropriate social protection systems and measures" (Target 1.3) and "create sound policy frameworks at the national, regional and international levels, based on pro poor strategies" (Target 1.b). Furthermore predicting women's BMI, which is a key nutritional indicator, directly impacts **SDG 2: Zero Hunger**. More precisely, the predictions help to "enhance agricultural productive capacity in developing countries" (Target 2.a) Since the application is processing a large amount of images using deep convolutional networks it requires large energy consumption and therefore directly impact **SDG 7: Affordable and Clean Energy** and **SDG 13: Climate Action**.

2.2 SDG enabler and inhibitor

Accurately predicting the poorest regions on a frequent basis can help policymakers take appropriate action. Therefore the technology clearly enables SDG 1: No Poverty. In addition, predicting women's BMI, is an effective way to understand both food scarcity and obesity. The technology can help making effective decisions on food provision and therefore enables SDG 2: Zero Hunger. However, assuming the technology would be in use across the entire less developed landscape, while the number of processed images increases exponentially, the computational effort will definitely have a negative impact on SDG 7: Clean and Affordable Energy and SDG 13: Climate Action. Further discussion about this will follow in section 3.2 Environmental sustainability.

3 Impact on sustainability

3.1 Social sustainability

Is it equitable; does it offer more opportunities to certain groups? The spatial coverage in the Mapillary dataset is very uneven[4]. Especially in developed countries the data representaion is low. For example coverage in Africa is very poor as compared to North America, see Figure 1. Consequently, the model might be applied preferably in regions where the data coverage is already high leaving out poorer countries. More generally, [10] states "AI applications are currently biased towards SDG issues that are mainly relevant to those nations where most AI researchers live and work". Obviously, if images are not available, the model cannot make any predictions. As long as there is insufficient image coverage, the solution can never capture the full socioeconomic landscape. Leaving behind those areas that cannot benefit from the predictions might lead to an increase of inequalities (SDG 10).

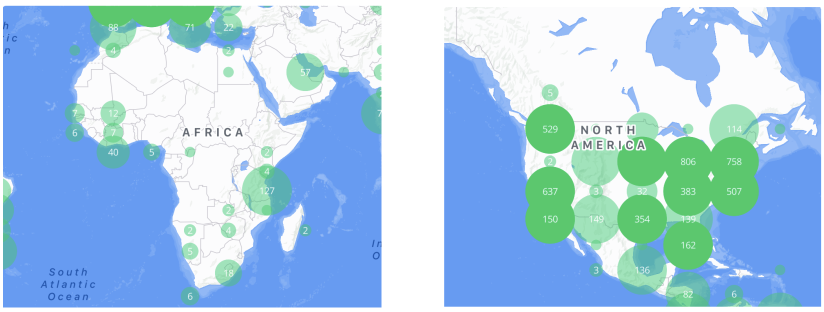

Figure 1: Comparing Mapillary dataset coverage in Africa and North America

Are individuals and communities involved in making decisions about the activity, and is the decision-making process fair and democratic? Due to privacy concerns predictions are made on a cluster level. Therefore the solution cannot identify poverty and health indicators on an individual or household level. Policy makers will not be able to address issues on that finer level which is crucial to ensure equally distributed positive impact. Even though individuals can participate in the dataset by uploading images, the model disregards any involvement of local communities and individuals in the decision making process. "This inherent dilemma of collective vs. individual benefit is relevant in the scope of AI applications but is not one that should be solved by the application of AI itself" [10]. The approach can only be democratic

[4]https://www.mapillary.com/dataset/vistas

if internet and camera access is equally distributed such that the underlying dataset becomes more consistent. Moreover, local communities need to be involved more directly. Further regulatory solutions to this dilemma are discussed in section 6 Speculative solutions.

3.2 Environmental sustainability

Is the application resource intensive? What about the data collection? The paper [4] fails to address the ecological impact of the computation for accurate prediction. The paper claims to be an "inexpensive, scalable, and interpretable approach to predict key livelihood indicators from public crowd-sourced street-level imagery" [4]. However, a critical assessment on the computational effort is missing. Obviously, scaling up the technology will come along with massive computational and therefore energy requirements. Assuming the technology, which uses quite deep convolutional architectures, will be used across the world, while the number of processed images will continue to grow exponentially to ensure accuracy, the computational effort will raise accordingly. As long as energy is reliable on fossil resources this clearly inhibits SDG 7: Affordable and Clean Energy and SDG 13: Climate Action.

Does the application protect us against risks related to climate instability and disasters? As described in a report from the *European Parliamentary Research Service* (EPRS) [5]: "Developing countries are most vulnerable to climate change" and "the poor are likely to suffer most from climate change". Having an accurate poverty map across the developing countries could help equip the poorest areas with infrastructure that make them more resilient against the effects of climate change. For example, as floods and droughts are expected to occur more frequently in poor countries [5], institutions can use the ML poverty predictions such as in [4] to support the most vulnerable areas with appropriate measures (robust buildings, water supply, etc.).

3.3 Economic sustainability

What could be the economic impact of the application? Surveys are expensive, conducted infrequently and "may only capture an extremely small proportion of households" [4]. As depicted in Figure 2 especially less developed countries lack information on economic development. Inference on socioeconomic indicators from street level images can be a lightway and less expensive solution to that problem. Assuming the predictions are accurate and capture the real economic local situation the ML model can be an alternative to expensive surveys.

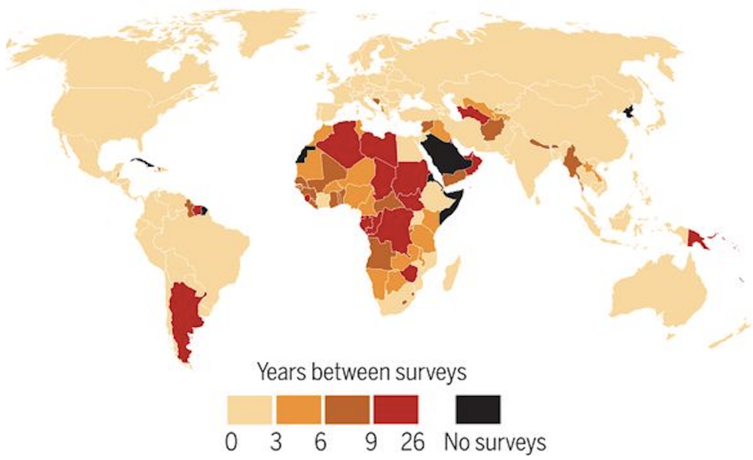

Figure 2: Average interval between economic surveys, 1993 to 2021, borrowed from [1]

Is the model open-source? How could that impact our economy, specially if the solution can be deployed and reutilised in a scalable fashion to other problems? The model uses publicly available data, i.e. the Mapillary dataset and DHS surveys. The source code is also open source[5] which enables researchers to replicate results, improve the model and transfer it to other countries. As mentioned above that may result in remarkable cost savings as expensive surveys may become redundant. However, scaling up the solution requires more AI specialized labor for research and deployment and therefore shifts labor from local institutions in developing countries towards already well off and better educated people and wealthy nations.

[5]https://github.com/sustainlab-group/mapillarygcn

4 Other sustainability factors

Question 1: Is privacy preserved? Since [4] is working with survey and street level image data the question of privacy concerns has to be asked in more detail. [4] addresses that DHS geo tags are shared within a 5 km radius to mitigate privacy issues but fails to address how privacy is preserved in the Mapillary dataset. Street level images capture a wide variety of personal information such as plate numbers and faces. Therefore a more in depth analysis on privacy measures is necessary. In fact, the Mapillary dataset uses an automated blurring mechanism with 99.9% accuracy[6]. Furthermore, users can blur images manually in case the algorithm misses something[7]. [4] should have dedicated privacy concerns more attention.

Question 2: How do you ensure that the street level images capture the current situation? [4] mentions no details on how the images are selected. Many of the images might be outdated and therefore not representing the current situation anymore. One has to ensure that the DHS datapoints and the image timestamps intersect to some extent. As suggested in [11] a possible approach is that an image "must have been captured within 3 years before or after the year of the DHS datapoint"[11].

Question 3: Is the model transferable and generalisable and interpretabe? The model was tested only in two countries, Kenya and India. It is highly questionable how well the model generalises to other developing countries even though efforts were made to increase transferability by using pooled models (i.e. the model is trained on images from both countries and then predicts on each country separately) [4]. Even tough street level images and satellite images are not directly comparable it is worth mentioning that [2] found predicting livelihood indicators using satellite images gerneralises poorly to different countries. One of the biggest challenges is the interpretability of AI models. [4] made efforts to increase interpretability by showing feature importance by indicator. For example, street lights and vehicles were the most important features for predicting poverty. An overview is depicted in Figure 3 which is borrowed directly from [4].

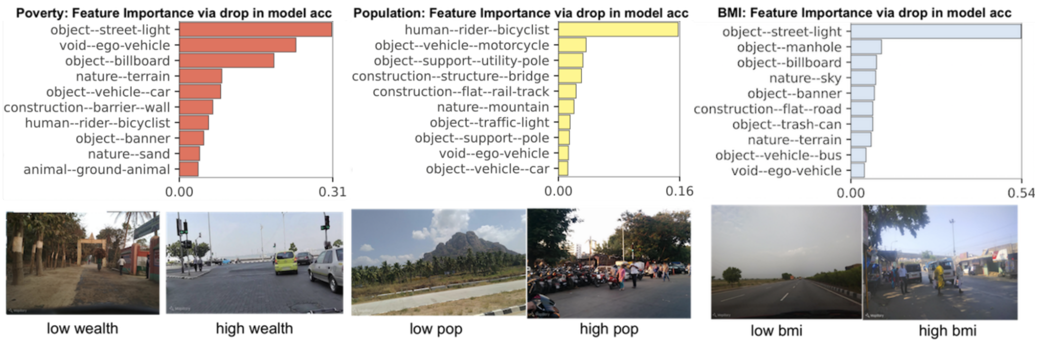

Figure 3: Feature importance by indicator. Note that BMI is specific to India as BMI data was not available in the DHS data in Kenya.

5 Interactions between SDGs

Interaction 1: Predicting poverty using street level images (SDG 1) can indirectly improve health (SDG 3: Good Health and Well Being). For example "there is evidence that child mortality is connected to environmental factors such as housing quality, slum-like conditions, and neighborhood levels of vegetation" [3] which is observable on street level images. So interestingly, street level images might be able to predict indicators that are not directly captured by the image. Policymakers can use insights from the predictions and perform appropriate measures to improve health related issues like child mortality (Target 3.2). First approaches to predict child mortality using satellite images were made in [11]. On the interaction scale in [7] this interaction can be categorised as *reinforcing*.

Interaction 2: Predicting poverty (SDG 1) and women's BMI (SDG 2) can impact SDG 10 (Reduced inequalities) both positively and negatively (Interaction 3). Positively in a sense that the prediction can help identify sources of inequality (e.g. unfair food supply, wealth gaps etc.) Policymaker can use the predictions to "empower and promote the social, economic and political inclusion of all" (Target 10.2). In this case SDG 10 is *enabled* by SDG 1 and 2 according to [7].

[6] https://blog.mapillary.com/update/2018/04/19/accurate-privacy-blurring-at-scale.html
[7] https://help.mapillary.com/hc/en-us/articles/115001663705-Blurring-images

Interaction 3: However, the prediction may also negatively impact SDG 10 and increase inequalities. As discussed in section 3.1 inequalities might arise because of poor coverage in the image dataset. Obviously, less wealthy and political unstable areas tend to have less access to cameras and the internet which results in less representation in the Mapillary dataset. "crowdsourced data can be noisy and inconsistent" [4]. Biases in the data might lead to inaccurate predictions or even no usable prediction at all. Those areas might be left behind resulting in inequalities that also have a negative cascade affect on SDG 8: Economic Growth and Decent Work, and SDG 9: Industry, Innovation and Infrastructure. SDG 8, 9 and 10 are therefore *constrained* by SDG 1 and 2 according to [7].

6 Speculative solutions

6.1 Changes to dataset

To address issues related to transferablity and the lack of street level images a reasonable approach would be include satellite images to the dataset. As suggested in [11] "we encourage researchers to develop new methods that can utilize both satellite imagery and street-level imagery, where available". An example for using satellite images for a similar task is made in [11] and [12].

6.2 Changes to ML model

So far the model operates on a cluster-wise level to preserve privacy. For finer household level predictions it would be helpful to allow the model to predict at a more individual scale which should be possible as privacy measures are very effective in the Mapillary dataset (see section 4). However, prediction accuracy may suffer as the DHS labels are clustered. Anonymization techniques in survery data might enable better predictions. This way individuals and local communities are much better represented and involved.

6.3 Technology governance

To address issues discussed in section 3.1 regarding the exclusion of local communities in the decision making process researchers and policymakers have to ensure that local politics and history is taken into account when deploying the technology. Applying the concerns presented in the YouTube video "Can experts solve poverty"[8] predicting poverty is not only a technical but also a political problem. "If we want to solve poverty, we have to first recognise that it has no technical solution"[9]. Furthermore, as pointed out in [10] "there is the risk that AI-based technologies with potential to achieve certain SDGs may not be prioritized, if their expected economic impact is not high". Funding of AI research should increasingly dependent on its impact on the SDGs rather than commercial interests. Lastly, to address the computational constraints to minimise the carbon footprint and reducing the energy requirement (discussed in section 3.2) legal interventions such as carbon tax can be helpful and nudge researcher to develop less resource intense architectures.

6.4 Speculating on the ideal scenario/dataset/task

In an ideal scenario, every individual and every household is equally represented in the underlying dataset. That would require to equip the less developed world with internet access and smartphones such that everyone can contribute to the dataset. Moreover, general IT skills have to be taught in order to ensure that all social classes can participate. Most importantly, local communities and politicians are involved in the decision making on how to implement appropriate measures to reduce poverty and hunger. To ensure better data coverage local governments could nudge its citizen, e.g. financially, to upload images in less represented regions.

[8] https://www.youtube.com/watch?v=8jqEj8XUPlk
[9] see footnote above

References

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

[2] Andrew Head, Mélanie Manguin, Nhat Tran, and Joshua E Blumenstock. Can human development be measured with satellite imagery? *Ictd*, 17:16–19, 2017.

[3] Marta M Jankowska, Magdalena Benza, and John R Weeks. Estimating spatial inequalities of urban child mortality. *Demographic research*, 28:33, 2013.

[4] Jihyeon Lee, Dylan Grosz, Burak Uzkent, Sicheng Zeng, Marshall Burke, David Lobell, and Stefano Ermon. Predicting livelihood indicators from community-generated street-level imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 268–276, 2021.

[5] Fulco Ludwig, Catharien Terwisscha van Scheltinga, Jan Verhagen, Bart Kruijt, Ekko van Ierland, Rob Dellink, Karianne de Bruin, Kelly de Bruin, et al. Climate change impacts on developing countries-eu accountability. 2007.

[6] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*, pages 4990–4999, 2017.

[7] Måns Nilsson, Dave Griggs, and Martin Visbeck. Policy: map the interactions between sustainable development goals. *Nature*, 534(7607):320–322, 2016.

[8] Lorenzo Porzi, Samuel Rota Bulo, Aleksander Colovic, and Peter Kontschieder. Seamless scene segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8277–8286, 2019.

[9] Esra Suel, John W Polak, James E Bennett, and Majid Ezzati. Measuring social, environmental and health inequalities using deep learning and street imagery. *Scientific reports*, 9(1):1–10, 2019.

[10] Ricardo Vinuesa, Hossein Azizpour, Iolanda Leite, Madeline Balaam, Virginia Dignum, Sami Domisch, Anna Felländer, Simone Daniela Langhans, Max Tegmark, and Francesco Fuso Nerini. The role of artificial intelligence in achieving the sustainable development goals. *Nature communications*, 11(1):1–10, 2020.

[11] Christopher Yeh, Chenlin Meng, Sherrie Wang, Anne Driscoll, Erik Rozi, Patrick Liu, Jihyeon Lee, Marshall Burke, David B Lobell, and Stefano Ermon. Sustainbench: Benchmarks for monitoring the sustainable development goals with machine learning. *arXiv preprint arXiv:2111.04724*, 2021.

[12] Christopher Yeh, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nature communications*, 11(1):1–11, 2020.