Experiments- Varun and Zaid

To test my hypothesis that deep NNs can be trained to leverage satellite imagery to accurately identify the poverty level (wealth index) of the region along with several sub-hypotheses, I designed and executed numerous experiments. The experiments are aimed at evaluating the following:

(a) How well can a deep neural network predict the poverty level of a region given an overhead nighttime satellite image with a limited dataset?

(b) How well can a deep neural network predict the poverty level of a region given an overhead daytime satellite image with a limited dataset?

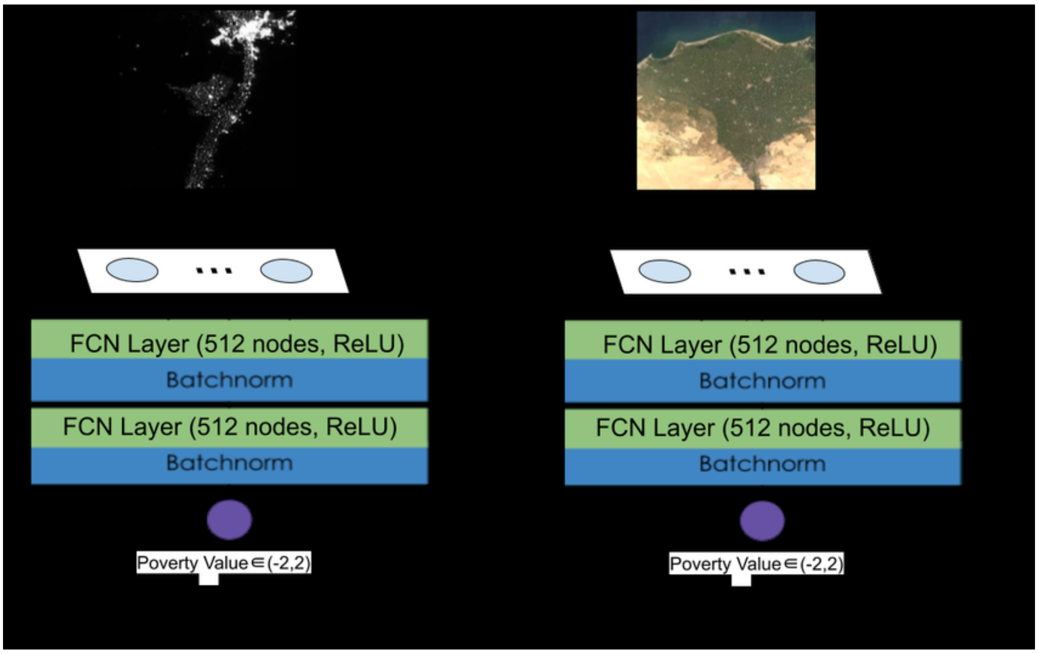
(c) What is the impact of data quantity (i.e. more or less data) on the representational power and overall accuracy of a deep neural network trained to predict the poverty level of a region given an overhead nighttime satellite image?

(d) What is the impact of data quantity (i.e. more or less data) on the representational power and overall accuracy of a deep neural network trained to predict the poverty level of a region given an overhead daytime satellite image?

(e) What is the impact of data augmentation on the representational power and overall accuracy of a deep neural network trained to predict the poverty level of a region given an overhead nighttime satellite image? How does data augmentation compare with a larger dataset?

(f) What is the impact of data augmentation on the representational power and overall accuracy of a deep neural network trained to predict the poverty level of a region given an overhead nighttime daytime image?

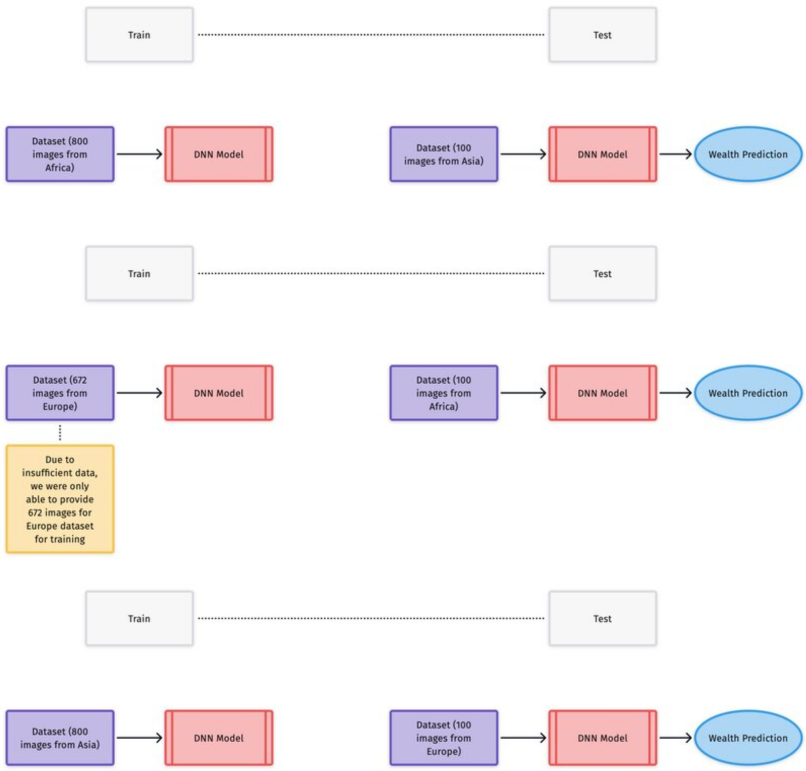
(g) How well can a deep neural network trained on a single continent's data generalize to data from other continents?



Baseline: Using a Limited Dataset of Daytime and Nighttime Satellite Images-

To evaluate my top level hypothesis and establish a baseline for the remaining experiments, I ran experiments evaluating how well a deep neural network can predict the poverty level of a region based on a single daytime or nighttime satellite image, with a limited dataset. I restricted my training dataset to 800 images, and evaluated my network on 100 images, each randomly chosen from their respective splits (see Dataset under Methods).

Data quantity- I trained deep DNNs with the following amount of data: 1. 800 images 2. 1600 images 3. 2400 images
Data Augmentation- The specific augmentations I used were random flipping (180 degree rotation followed by a mirror across the vertical axis through the center of the image), random rotations (90 degrees counterclockwise), and random Gaussian noise with a mean of zero and standard deviation of 2.5 [23].



Generalisation to unseen continents-
I picked a continent at random to train on and a separate continent to evaluate on. Specifically, the three experiments are: 1. Train on data from Africa, evaluate on data from Asia 2. Train on data from Europe, evaluate on data from Africa 3. Train on data from Asia, evaluate on data from Europe.

Results-

Experiment Name	RMSE (95% CI)
Baseline Night	1.222 (1.033-1.423)
Baseline Day	1.611 (1.399-1.832)

Table 2: This table summarizes the results of the model when simple day and night images were used to train and test the model. To evaluate the performance of the model 800 train images and 100 test images were used.

Data quantity and data augmentation- Results for experiments on data quantity and data augmentation are presented in Table 3 and Table 4, respectively. A DNN trained on nighttime images achieves RMSE's of 1.222 (95% CI: 1.033-1.423), 1.132 (95% CI: 0.963-1.354), and 1.106 (95% CI: 0.956-1.283) when trained on 800, 1600, and 2400 images respectively. A DNN trained on daytime images achieves RMSE's of 1.611 (95% CI: 1.399-1.832), 1.459 (95% CI: 1.275-1.633), and 1.477 (95% CI: 1.203-1.894) when trained on 800, 1600, and

2400 images respectively. The same evaluation dataset of 100 images was used across all experiments for each of daytime and nighttime experiments for consistent comparisons.

Experiment Name	RMSE (95% CI)
Night, 800 images	1.222 (1.033-1.423)
Night, 1600 images	1.132 (0.963-1.354)
Night, 2400 images	1.106 (0.956-1.283)
Day, 800 images	1.611 (1.399–1.832)
Day, 1600 images	1.459 (1.275-1.633)
Day, 2400 images	1.477 (1.203-1.894)

Table 3: This table summarizes the results of the model when the amount of data used to train the model was varied.

Experiment Name	RMSE (95% CI)
Baseline Night	1.222 (1.033-1.423)
Night with Augmentation	0.956 (0.872-1.040)
Baseline Day	1.611 (1.399-1.832)
Day with Augmentation	1.463 (1.221-1.705)

Table 4: This table summarizes the results of the model when the data used to train the model was partially augmented.

Distribution shifts between Countries/Continents-

There was a fairly significant data distribution shift across countries and continents. Median wealth indices for each country range from -1.238 in Ethiopia to 1.279 in Armenia. Variance of the wealth index also differed dramatically across countries, ranging from 0.908 in Namibia to 0.0380 in Armenia. Variance in the wealth index is a proxy for wealth inequality within a country where higher variance indicates greater inequality.