Using Satellite Imagery to Predict Health

Jeremy Irvin & Dillon Laird

Stanford University

Background/Introduction

- Organizations such as UNICEF and the Bill and Melinda Gates Foundation currently use surveys to assess where to allocate resources for humanitarian efforts in third world countries
- These surveys are costly and often dangerous to collect resulting in large temporal and spatial gaps between surveys
- Satellite images are a much cheaper, safer, and more frequent method to predict different indicators of poor health
- Previous work has predicted poverty and crop yield from remote sensing data indirectly using proxies

Problem Statement

We combine Demographic and Health Survey (DHS) data with satellite images to predict malnutrition and poverty. We use a Convolutional Neural Network per health indicator to perform a multi-class classification (for example, for poverty we predict buckets poorest, poorer, middle, richer, richest). We evaluate with cross-entropy loss and accuracy.

Datasets

The DHS data contains more than 300 surveys in over 90 countries, where each survey contains health information from thousands of households. The households are grouped into clusters and each cluster is assigned a single latitude-longitude pair of its approximate centroid.

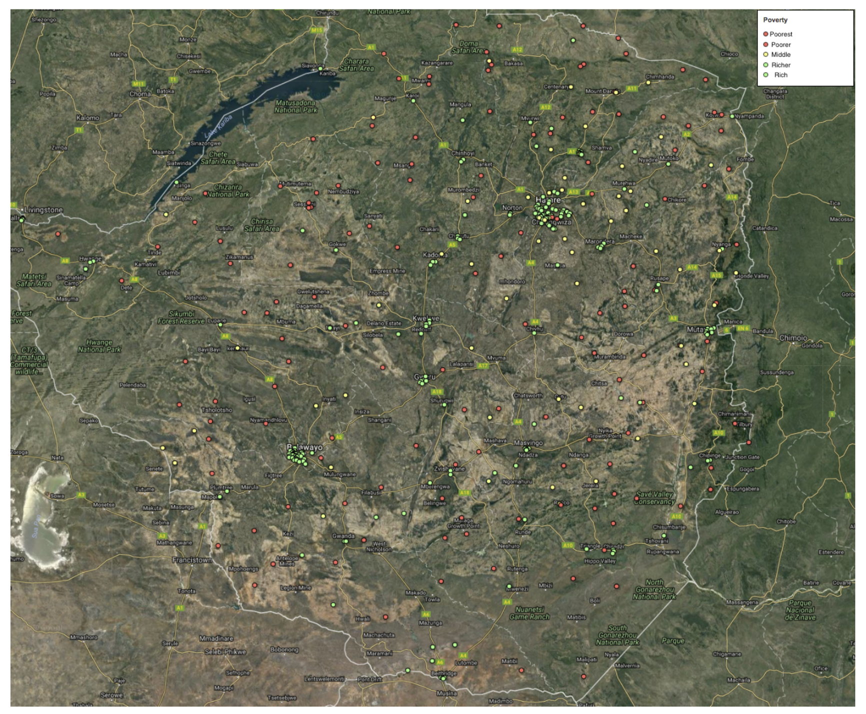

Figure 1: Households from DHS in Zimbabwe.

However, in order to ensure confidentiality, the centroid is displaced: in urban areas, the cluster centroid is uniformly displaced up to 2km, and in rural areas, the cluster centroid is uniformly displaced up to 5km, where 1% are uniformly displaced up to 10km. We label each cluster with the median label of the households within that cluster. We only use surveys from 2010 and later (around 45 countries) yielding 31K clusters.

Figure 2: Example of a rich area (left) and a poor area (right).

Since we must cover a 10km x 10km region to ensure the households are contained, for each latitude-longitude pair, we obtain 12 Google Maps images each of resolution around 5m per pixel:

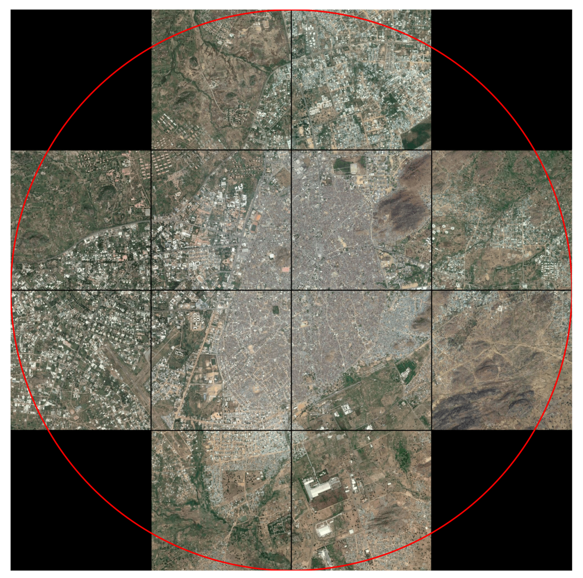

Figure 3: Example of tiling a single cluster region into 12 higher resolution images. The red circle indicates the region within the cluster where the households may lie.

This yields 372K images in the full dataset, split into training (80%), valid (10%), and test (10%).

Model

- Transfer learning using an 18-layer ResNet pretrained on ImageNet
- Train the fully connected layer and last three convolutional layers
- Also test a smaller 3 layer convolutional network with 2 fully connected layers
- Run the model on all 12 images and combine either the spatial or flattened representations, through a mean or max operation

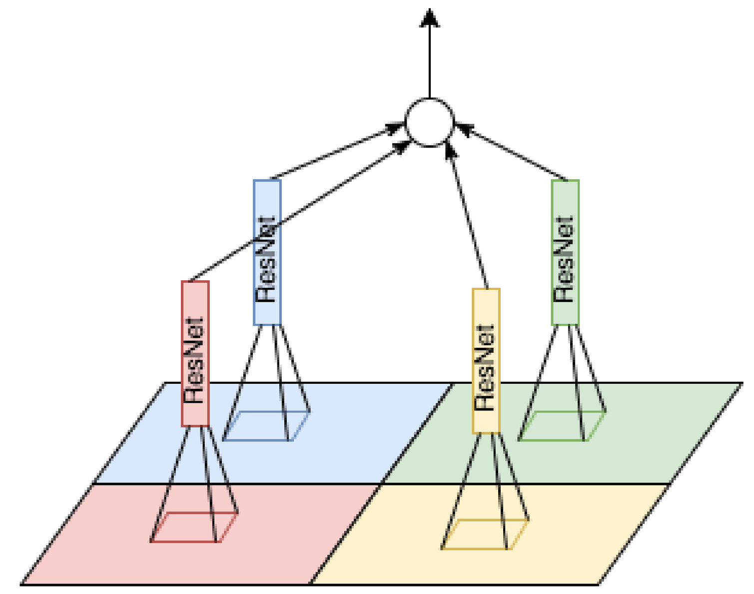

Figure 4: Each of the 12 tiles (only 4 shown here) are fed through the ResNet model and combined using either a mean or max operation.

Experiments and Results

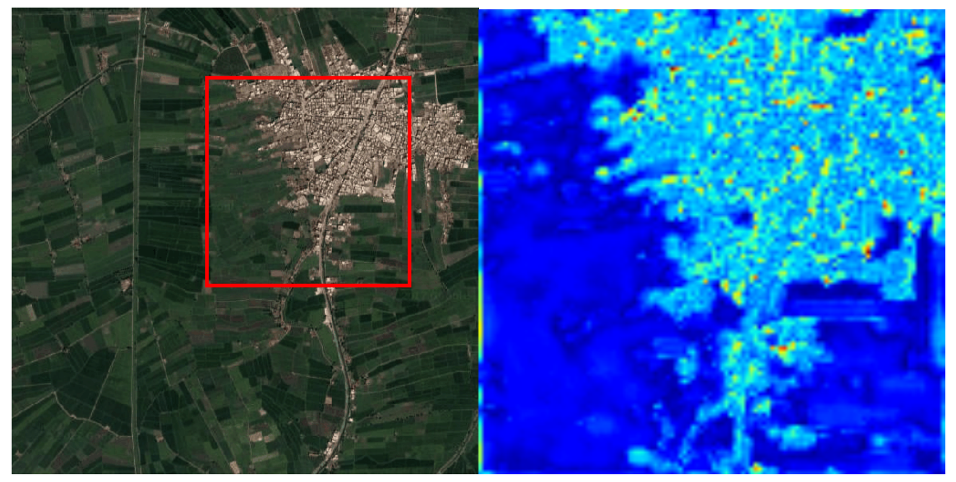

Figure 5: Visualization of activations (right) of the first layer of a pretrained ResNet model tuned on our datset. The input image (left) is randomly cropped to a 224x224 image to feed into the pretrained ResNet model.

We attempted to predict malnutrition under the hypothesis that it is related to poverty but the models performed no better than random.

	Poverty	
Model	Valid	Test
3 Layer Conv, 2 Layer FC	34.541	32.681
ResNet Fully Trained	34.253	32.906
ResNet Pretrained, Mean	**44.131**	**43.489**

Figure 6: Experimental results on the validation and test sets. Human gets 32% accuracy on a sample of 100 images.

We found it difficult to discern poorer vs poorest and richer vs richest images which led us to believe that top 2 accuracy may be a better metric for evaluating the model. For the pretrained ResNet we obtained 72.417 valid and 72.996 test top 2 accuracy scores.

Conclusion and Future Work

- Predict other health outcomes such as malaria, which are more correlated with physical features like wetlands, bodies of water, quality of rooftops, and highly dense populations
- Use higher quality and higher resolution satellite images. Unfortunately high quality satellite images are not cheap and not available for every location
- Test models which are more selective about where to look in the satellite image, rather than randomly sampling a location. This could be accomplished by having an LSTM take a downscaled version of the larger satellite image, output a location which then feeds an image through a CNN and then feed the output back into the LSTM and repeat. This could be trained with reinforcement learning.

Conclusion and Future Work

We'd like to thank Pranav Rajpurkar, Andrew Ng, and Dr. Eran Bendavid for their help with this project.