Data:

We will use satellite imagery to help our prediction. Specifically, we will be using RGB daytime images and greyscale nightlight images. The RGB images are pulled using the Google Static Maps API and the nightlight data is available through The National Oceanic and Atmospheric Administration (NOAA). The RGB images are 640 pixels by 640 pixels in size, with each pixel have a resolution of about 2.63 meters by 2.63 meters, while the nightlight image is 43201 pixels by 16801 pixels and covers the entire Earth from -180 to 180 degrees longitude and -65 to 75 degrees latitude which puts each pixel at about 240 meters by 240 meters resolution.

The reference data we are using is Malawi’s Integrated Household Surveys. Malawi has run its Integrated Household Survey (IHS) in 2010-2011, 2013-2014, and most recently in 2016-2017. Households surveyed are designed to be a nationally representative sample and are diverse in terms of spatial location and month of interview time. From the surveys we are able to obtain variables of interest: household wealth and assets, food security measures, and geospatial data. We are interested in predicating the average roof types and asset index in this study. However, in order to keep the confidentiality of each respondent, the IHS does not give the exact coordinates of the respondents. Instead, they first break the dataset into clusters (about the size of a village) and assign every household in the cluster the average coordinates of everyone in the cluster, then assign a randomized offset from this center point. In urban areas, this offset is between 0-2km and in rural areas this offset is between 0-5km (with 1% of rural clusters receiving another 0-10km offset). The implication of this spatial distortion suggests that the lowest spatial aggregation level of analysis would be at the clusters with a 2-5 km buffer.

Method:

We will use a convolutional neural network on the RGB images to predict the average amount of assets and the consumption budget at the cluster level. This network will be trained through supervised learning with the reference being Malawi’s IHS data. To avoid possible measurement error, similar to Jean et al (2016), we will train the convolutional neural network on ImageNet and then on the nightlight data as a proxy using a transfer learning approach. The Nightlight data has been shown to be a good proxy for economic activity, so by using it through transfer learning it should help negate some of the problems associated with the lack of data in Malawi’s IHS. The final step is to use the trained model directly onto the wealth and asset measures in the IHS survey from the RGB images. After the model has been trained, its prediction will then be combined with soil modules, crop modules, and market access modules to build the food security prediction.