**Predicting Food Security with Machine Learning**

Yujun Zhou, Kathy Baylis, Erin Lentz, Hope Michelson

September 25, 2019

**Abstract**

Hunger is on the rise throughout Africa, with famine threatening millions across several countries. Identifying food insecurity crises rapidly and accurately can enable humanitarian responses to mitigate casualties from hunger and save lives. We develop a predictive model based on readily available, spatially granular data on prices, geography, and demographics. Using machine learning techniques, we are able to improve the accuracy of predicting those villages that face a potential threat of hunger. As with any rare event, one challenge with predicting food insecurity is the low rate of severe food insecurity in the baseline data. We use several different approaches to address this imbalance to allow us to capture a higher fraction of these rare events. We apply our procedure to three sub-Saharan African countries: Malawi, Tanzania, and Uganda to predict food security in out-of-sample villages. We correctly identify up to 88 percent of the food insecure clusters, which is 70 percent higher than the baseline model using a logistic regression. We further explore which data splits perform best under the spatial-temporal correlations between observations in the panel dataset to reduce overfitting in actual implementation. The amount of correlation… Our result shows that a data-driven model with the help of machine learning methods can significantly improve its performance on capturing the food insecure households despite the imbalance in the data. Our paper demonstrates that this approach could be used in a scalable, automatically updated prediction model that could enhance the current famine early warning systems.

**Keywords:** food insecurity, machine learning, early warning, Sub-Saharan Africa, famine

**Predicting Food Security with Machine Learning**

1. **Introduction**

Hunger crises are increasing in frequency and severity in many parts of the world. Identifying the scale and scope of these crises in a timely and accurate fashion is essential for providing food aid and organizing humanitarian responses to mitigate the long-run effects of food insecurity. Without timely identification to target the vulnerable population, food aid often fails to arrive the areas in time where the assistance is needed the most (Barrett and Headey 2014). One of the many reasons that prevent building a successful early warning system is that data are scarce, and data collection is costly (Hutchinson 1991). The data gap hinders efforts to effectively target the population in need and calls for the use of data and method that are cost-effective and accurate. Currently, governments and NGOs in the sub-Saharan Africa region use the Integrated Food Security Phase Classification System (IPC) as the early warning system. The IPC uses a Delphic system that requires detailed on-the ground data and is updated quarterly for each livelihood zones, making it difficult to identify specific villages that might be at risk of hunger in the near term.

The recent increase in available data on geography, weather, and market price for food staples provides us with the opportunity to predict food security more frequently at a finer geographic level. The use of remotely sensed data to predict socio-economic outcomes is a growing endeavor. Nightlights data (Chen and Nordhaus 2011; Henderson et al. 2012) can serve as a proxy for economic activity but variations in the nightlight intensity are too low in remote rural or better off urban areas to detect any substantial changes in economic outcomes. Mobile phone data (Blumenstock, Cadamuro, and On (2015); Steele et al., 2017) also hold potential for identifying economic outcomes and are more frequent and less expensive than census surveys. However, geocodes are often limited to cell towers, and the biases associated with using relying on cell phone-sourced information to infer population statistics are as of yet, not well understood. Very high-resolution satellite imagery is becoming cheaper but the lack of labeled data in the imageries makes it challenging to extract structured information from the raw images (Engstrom et al., 2017; Donaldson and Storeygard, 2016). Recent studies using a Convolutional Neural Network (CNN) and transfer learning (Jean et al., 2016; Babenko et al. 2017) make promising progress utilizing the information in satellite imageries. These models can explain up to 60% - 75% of the variation at the village level wealth and asset measures in several sub-Saharan Africa countries. However, the reliance on the information in the satellite imagery (specifically, building size, roof type, road conditions) limits its performance for time-varying development indicators. Head et al. (2017) finds that the prediction performance of the Jean et al. method degrades quickly on health and nutrition outcomes to no better than random guessing in some cases. The reliance on nightlight data in this approach also limits the prediction accuracy when applied in countries with different socioeconomic conditions. The external validity and interpretability of this deep-learning-based approach call for a method tailored for food security predictions.

To the best of our knowledge, Lentz et al. (2019) is one of the few papers that combines publicly available spatially and temporally granular data to predict village-level food security status that greatly improves the prediction accuracy without significant cost in data collection and model training. Building on the framework in Lentz et al. (2019), in this paper, we construct a prototype of an early-warning system that is automatically updated, generalizable, scalable, and cost-effective in predicting areas of potential food shortage. Like Lentz et al., we incorporate data from different sources, dimension, and scales into a single predictive model of food security status. We use machine learning models to predict cluster-level food security status for targeting aid in times of food shortage. Variables in the model include the market price of food staples, weather shocks in growing seasons, and geospatial features around village clusters. We combine data techniques such as oversampling and data segmentation with the machine learning models to improve the prediction of food insecure categories. The models correctly capture 69-88 % of food insecurity categories among the three countries for different food security measures. The main contribution of this paper is to improve the prediction of clusters with potential food crisis in an imbalanced data setting with spatial-temporal correlations in the data. We are able to achieve this by choosing an objective function that balances overall accuracy against the ability to capture the food insecure, along with techniques like data sampling, data split, and data segmentation.

Instead of predicting the overall accuracy of food security status in previous studies, this study focuses on correctly detecting the clusters that are food insecure. Similar to a fraud detection problem, severe food insecurity crises are rare but too valuable to miss. The failure of identifying villages with food shortage is more costly than falsely sending food assistance to areas that do not have a shortage of food. Therefore, we focus more on identifying all the insecure clusters to maximize the recall rate in classification, where the recall rate is defined as percent of truly insecure households correctly captured by the model. We care less about minimizing the number of secure clusters misclassified as unsafe. In technical terms, we put a higher weight on the recall rate than the precision rate for classifying the food security categories. Choosing the right criterion to optimize matters for model selection and parameter tuning. Ultimately, we want the model to correctly detect the minority classes that are food insecure without too many falsely positive cases.

Along with choosing the optimization criterion, we also explore the effect of different up and down sampling approaches to identify the food-insecure groups. Various sampling techniques create a balanced dataset that forces the classifiers to learn about the characteristics of the minority class. One such method is oversampling the minority class at the risk of model overfitting. SMOTE creates synthetic new data of the minority class by forming convex combinations of neighboring points, as a way to reduce the overfitting in oversampling.

Data split and data segmentation rules also matter for creating a balanced and representative training and testing data set. We explore several different options in the context of our panel data to determine how best to split the data to train the model. Given spatial and temporal correlation among observations in the panel data, the independence assumption will not hold across different training and testing splits. Without consideration of these correlations, we are inclined to choose more complex models and the testing would indicate we have higher accuracy than we actually do (Robert et al. (2017)). Our paper contributes to the literature on socio-economic predictions using a combination of available data by emphasizing the importance of addressing and correcting for the spatial-temporal correlations among observations to reduce overfitting.

As a natural extension to Lentz et al. (2019), this study expands the study areas to Malawi, Uganda, and Tanzania with more years of data to test the framework with more heterogeneity in geography, environment, and socioeconomic status. For example, Uganda has two growing seasons, and the main food staples are matoke and cassava while most areas in Malawi and Tanzania have only one growing season and rely on maize as the staple food. This means adjustments to the local climate and agricultural markets, such as having the weather variables during local growing seasons, grabbing markets data on the staples that take up a more significant share in the household budget in that specific area. The machine learning algorithms and data techniques used for prediction are the same kinds for the three countries, but the hyperparameters are tuned on the training dataset of each country separately. Using the same procedure and types of data supports the validity of the method as it proves to display similar results across countries despite the heterogeneity. ~~makes the model generalizable for application in other data-scarce countries and areas with some previous household survey data (LSMS or DHS) and frequently updated market price for food staples~~. At the same time, the model remains flexible and adaptable enough to capture the differences between countries such as climate, crops, and different levels of infrastructure. This research also sheds light on the ability to apply the model trained on areas where we have ground-truth data to offer insights on areas of the world where we do not. We compare different methods and protocols of handling the raw data, choosing the right data split and data segmentation, selecting the optimal model, to come up with a standardized data flow that maximizes our chances of making the model generalizable for potentially other areas in the world.

This paper uses a data-driven framework with machine learning techniques to predict the onset of food crises. Combining remote sensing data with household surveys and price data, the models are able to produce the most spatially and temporally granular predictions of food security. With an emphasis on the structure of the prediction error, this paper uses various machine learning techniques to reduce the misclassification of food insecure cluster. The framework developed in this paper has important policy implications for accurately target and aid areas of potential food shortage in data scarce environments.

1. **Data:**

***Food Security measurement***

We predict three measures of food security used by the international humanitarian organizations, including USAID and the World Food Programme (WFP): the reduced coping strategies index (rCSI), the household dietary diversity score (HDDS) and the food consumption score (FCS). The HDDS measures the number of different food categories that a household consumes in past seven days. The FCS gives nutrient-related weighting to the count of food categories to come up with a weighted score of food quality. Higher values of both the FCS and the HDDS indicate more diversity of nutrition intake and higher food security. The rCSI reflects the number of coping strategies a household uses to address possible food shortages with higher values of the rCSI indicating lower food security. The rCSI is believed to capture inadequate quantities of food consumed, which is consistent with acute food insecurity. Governments and international agencies apply cut-offs to categorize food security status rather than use the continuous measure (Vaitla et al., 2017). This is why this paper focus on the categorical prediction for the given cutoffs instead of the continuous measures of socio-economic outcomes in previous works (Jean et al., (2016), Steele et al. (2017), and Lentz et al. (2019), among others). The food security category is close to the actual policy scenarios where policymakers are trying to capture all the insecure households in a potential famine year in the currently used IPC system.

***Explanatory data***

The variables used to predict food security are high-frequency data, including precipitation, temperature, market prices, soil quality, and geographic variables. These data are generally collected remotely and are widely available. We handcrafted weather-related variables such as the first day of rain, length of dry spells, growing degree days and heating degree days from the raw precipitation and temperature data during the previous growing seasons specific for each country. We gather the market prices for main food grains such as maize, rice, and groundnuts for the major markets in each country and align the villages to the prices in their nearest markets. To help forecast future food security status, we use prices with one to three months prior to the household survey time. The tree-based models help us choose from a variety of price variables of different products, lag length and format, with more details in the discussion section. For missing data in the market prices, we construct market thinness measures defined as the number of weeks with price information missing in a given month. Variables regarding wealth status, asset ownership, and demographic characteristics are created using answers from the LSMS surveys. Although our results rely mainly on these variables from the surveys, we have variables that can serve as proxies for the information from the household surveys. Household roof type is used as a crude proxy of poverty that can be accurately captured from satellite imagery. Cellular phones are access to financial resources, market information, and remittance flow (Eagle et al. 2010, Blumenstock et al. 2016) also serve as significant predictors.

1. **Method**

This section explains the primary approach and techniques used in this paper. In summary, we use readily available data to model the food security status of village clusters in Uganda, Malawi, and Tanzania, with various machine learning related techniques.

***Categorical vs. continuous measure***

We focus on the categorical prediction for the given cutoffs of each food security measure as it captures the policy scenario where the policymakers need to target communities that are likely to fall under some average food security threshold. We do care about the overall fit of the prediction on the actual food security measures and we achieve a similar performance of model fit compared to previous studies at around 0.7 R squared. Since this paper focuses on successfully detecting the villages in need of food assistance, we use categorical measures of food security to transform the prediction into a classification problem. In this way, we can

utilize data techniques such as choosing the right result metrics, sampling techniques to improve the chance of detecting insecure villages.

***Result metrics***

Predicting when and where the food security crisis will happen is more important than having an accurate assessment of the food security status of the general population. In technical terms, this study focuses on the recall rate of insecure households, rather than the overall prediction accuracy. Models aiming to maximize the overall accuracy tend to capture characteristics that are rich in the majority of the population and fails to understand the insecure households enough. We want to maximize the recall rate to try and get all the insecure households, without a too low precision rate so that we do not mistakenly categorize all the secure households as insecure. The F-1 score and gmeans, essentially the weighted harmonic and geometric averages of precision and recall, serve as balanced measures of both recall and precision. To summarize, the prediction results will be evaluated based mainly on recall of the insecure category but also in consideration of the performance of other measures.

***Sampling***

Food secure households and villages make up the majority of the data. When we feed the prediction algorithms with training data made up with these proportions, the models naturally better identify the characteristics of the secure households more than the insecure ones. As a result, the models tend to predict villages in the testing dataset to be secure. To force the models to gain as much information as possible about the food insecure households, we apply downsampling and oversampling techniques to create a more balanced training dataset in terms of the outcome variable while the testing set remains intact. Specifically, we downsample the clusters in the food secure category and oversample the observations that are food insecure to artificially create a training set where the insecure households make up half of the observations. These methods are broadly used to deal with imbalanced datasets. The main disadvantage with oversampling is that by making exact copies of the minority class, the models tend to overfit. We also use a method called SMOTE (Synthetic Minority Oversampling Technique) to creates synthetic new data of the minority class by forming convex combinations of neighboring points, as a way to reduce the overfitting in oversampling.

As an alternative to the sampling technique, the cost-sensitive learning approach penalizes misclassifications of the minority class more heavily by having a cost function, which is equal to the inverse of the class proportions (Elkan 2001). This produces extra reward for identifying the minority class over the majority class. This approach penalizes misclassifications of the minority class more heavily than misclassifications of the majority class, in hopes that this increases the true positive rate. However, defining the “right” cost is not always easy, and they might vary from one case to another. The sampling approach that we use in this paper can be seen as a wrapper-based method that can make any learning algorithm cost-sensitive. The sampling effectively imposes non-uniform misclassification costs on the different categories (Elkan, 2001). Weiss, McCarthy and Zabar (2007) show that oversampling maybe preferable for smaller data sets like ours and cost-sensitive learning are more suitable for datasets with over 10,000 observations.

***Data split***

In the case of spatial and temporal correlation, splitting the data set into training and testing sets is harder than one would think, as the assumption of independence between the training and the testing set is not easily satisfied. The classical assumption of a random split in generating the testing data set may not hold since the points randomly assigned to the testing may have strong spatial and temporal correlation with some of the points in the training set. Because of these correlations, models with high prediction accuracy on the training set would appear to have higher than actual accuracy in the testing set as well. Consequently, the training process would prefer models and parameters that are complex enough to understand every bit of details in the training set. These models would tend to overfit in a truly independent testing set as is demonstrated in simulation results in Robert et al. (2017). We use three different training, and testing data split methods to feature the importance of having the right split. Yearly split is using one year as a training set to forecast places where the famine may happen in a future year. The regional division is used to predict rural or more remote areas given current information in more accessible regions. Random split is used as a demonstration of general prediction or the “best guess” on unknown households based on all the information that we have. We show in our results the degree of temporal correlation between data collected in different years around the same regions and between points across different regions but are close in geographic distances. Strong temporal and spatial correlation would mean more suspicions on the validity of the split and hence preference over other methods. The regional and random split method exhibits relatively higher temporal and spatial correlations between the training and testing dataset than our main result using a year split.

***Data segmentation***

Data segmentation is another choice to make in defining our training dataset. Due to the heterogeneity of different countries and regions, including more data in the training data sometimes is adding more noise than information. This leaves us the option of choosing the right subset of data to be trained when predicting certain regions or groups of data in order to improve model performance. We compare the results of models trained on the entire dataset of three countries, with models trained by each country separately. Similarly, urban and rural cluster may face different cost of living and respond differently to weather shocks. In predicting clusters located in urban areas, we make the comparison of models trained on urban clusters only with models trained on the entire country. Lastly, we train a shallow tree in each country based on observables to automatically split the data into several subsets that are very much different from one another, suggested by the training data. By using the same tree splits criteria in the testing data set, we divide the data into relative smaller groups. We compare the results of different data segmentation to find the optimal data segmentation strategy in food security prediction in order to improve the model performance even further.

***Classification algorithms***

For structured data like ours (unlike unstructured data like text or pure images), tree-based methods are popular, as they work well with the nonlinearities and a large number of variables with different scales in the data. We start with a simple machine learning algorithm known as the Decision Trees. The decision tree splits the training data based on several input variables, such as does the village undergo a long dry spell in the last year? If so, is the cluster rural or urban? The splits created by the questions divide the training data into different subset of data known as the “leaves.” In this classification problem, each leaf is associated with a probability of falling into the secure or insecure category and the probability is trained using the training dataset. The main hyperparameter of the classification trees is the depth of the trees. Deeper trees tend to capture more complexity of the training data but may suffer from overfitting when applied to the testing set. Shallow trees may suffer from losing essential splits in the data and underfit the training set. As a result, tree-based ensemble learning methods like the Random Forest, and Gradient Boosting help improve model performances by averaging and sequentially improving the base trees, which is particularly helpful when there is a large number of potential variables in our model. Random Forest is a kind of bagging algorithms, short for “Bootstrap aggregating.” The idea of the algorithm is to create many deep trees based on randomly selected subsamples of the training data with randomly selected variables determined at each split of the trees (the bootstrap aspect) and then using the weighted average of the results from the trees (the aggregating aspect) to reduce the total variance of the deep trees.

Boosting differs from the bagging method in that trees are grown sequentially instead of in parallel. They usual start with a simple, shallow tree for prediction and then use the error in prediction from the previously grown tree to adjust the weights in the next iteration. Errors are corrected by sequentially adjusting the weights in the existing models until no more improvements can be made. Weiss (2014) suggested that boosting methods tend perform well at classifying minority examples because boosting places more emphasis on misclassified instances in the training set and the errors are more likely to come from the minority classes. Based on the performance across twenty-nine datasets, Tischio and Weiss (2019) find that models like decision trees, Adaboost, and gradient boosting perform better than other algorithms in the presence of imbalanced data. The gradient boosted tree improves the initial decision tree model by sequentially adjusting the model in the direction of the negative gradient of the loss function defined as the squared distance of predicted and actual values. The Xgboost (Extreme Gradient boosting) algorithm relies on the same principles as gradient boosting but is more efficient and faster since it adds regularization to the loss function to prevent overfitting. We also tried out algorithms of anomaly detection such as clustering methods, One-class SVMs, and Isolation Forests but they perform less ideal compared to our main results using tree-based models.

***Baseline model***

As a comparator, we use a standard logistic regression as the baseline model. We estimate each country separately using all but the last year of data, and then use the model results to predict the last year, without downsampling or oversampling methods. We use the same variables used in machine learning models such as food prices, market thinness, cellphone ownership, floor/roof material, asset index, length of dry spells, average temperature, and the amount of rain.

1. **Results**

The results section shows a number of improvements in performance compared to the baseline model and suggest how and why certain data techniques or algorithms assist in predicting the when and where the food insecure instance happen.

1. On Measurement. Table 1 presents the differences in model performance, such as overall accuracy, recall rate, F-1 score, and gmean of our baseline model and the three machine learning models for all of the three countries. Higher accuracy for the ML models but similar recall rate and F-1 score.
2. On sampling: Table 2: Baseline vs ML algorithms with down/oversample technique. There is significant increase in recall rate when ML combined with oversampling
3. On data splits: Figure 1 Data Split Comparisons (showing the degree of correlation across year, across border, and between the “random points”) Year split results more reliable since surveys are conducted one or more than one year afterward
4. On Segmentation: Table 4: Data Segmentation Comparisons. Comparison of one dataset/ by country / urban-rural / automatic segmentation
5. **Discussion**

***Parameter tuning and feature importance analysis***

* Discussion on tree depth, max features and how they impact the results
* Discussion on variables that explains the majority of the variations in data, e.g., cellphone ownership

***Error analysis***

* Urban vs. rural
* By food security measure
* Regional error analysis

***Model deploy and update***

* Model generalization issues:
  + what happens when we directly apply models trained in one country to predict another;
* Model update: compare the results of using only one year to predict the third year, with a training set of two years. Thinking of a dynamic process of continually updating the model with new survey data, but the maximum number of years of data we have in each country is 3.

1. **Conclusion**

|  |  |
| --- | --- |
| A close up of a map  Description automatically generated  Malawi FCS | A close up of a map  Description automatically generated  Malawi rCSI |
| Tanzania FCS A close up of a map  Description automatically generated | A close up of a map  Description automatically generated  Tanzania rCSI |
| A close up of a map  Description automatically generated  Uganda FCS | |

**Fig. 1: Baseline vs. ML algorithms using year split**

**Table 2: Baseline vs. ML algorithms with down/oversample technique**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Country | Food Security Measure | Recall Rate Insecure category  (Baseline) | Recall Rate Insecure category  ML + Oversample | Recall Rate Insecure category  ML + SMOTE | Recall Rate Insecure category  ML +   ADASYN |
| Malawi  2010/11, 2013 to predict 2015/16 | FCS | 0.04 / na | 0.93 / na | 0.10 /na | 0.16 /na |
| HDDS\* | 0.49 / na | 0.68 / na | 0.83 / na | 0.71 / na |
| rCSI | 0.23 / 0.00 | 0.80/ 0.00 | 0.16 / 0.00 | 0.13 / na |
| Tanzania  2010/11, 2012/13 to predict 2014/15 | FCS | 0.02 / 0.00 | 0.63/0.00 | 0.49/0.00 | 0.58/0.00 |
| HDDS\* | 0.99 / na | 0.66/na | 0.92/na | 0.85/na |
| rCSI | 0.37 / 0.00 | 0.44/ 0.00 | 0.42 / 0.00 | 0.46 / 0.00 |
| Uganda  2010/11 to predict 2012 | FCS | 0.37/0.11 | 0.24/ 0.00 | 0.44/0.33 | 0.43/0.33 |
| HDDS\* | 1 / 0.00 | 0.89/0.00 | 1 / 0.00 | 1 / 0.00 |

**Table 3: Data Split Comparisons**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Country | Food Security Measure | Year split | | Regional split | | Random split | |
|  |  | Accuracy | Recall | Accuracy | Recall | Accuracy | Recall |
| Malawi | FCS | 0.80 | 0.93 / na | 0.79 | 0.59 / 0.00 | 0.94 | 0.93 / 0.00 |
| HDDS\* | 0.68 | 0.68 / na | 0.90 | 0.89 / na | 0.94 | 0.93/ 0.00 |
| rCSI | 0.71 | 0.80/ 0.00 | 0.80 | 0.64 / na | 0.93 | 0.83 / 0.00 |
| Tanzania | FCS | 0.80 | 0.93 / na | 0.79 | 0.59 / 0.00 | 0.94 | 0.93 / 0.00 |
| HDDS\* | 0.68 | 0.68 / na | 0.90 | 0.89 / na | 0.94 | 0.93/ 0.00 |
| rCSI | 0.71 | 0.80/ 0.00 | 0.80 | 0.64 / na | 0.93 | 0.83 / 0.00 |
| Uganda | FCS | 0.76 | 0.24 / 0.00 | 0.77 | 0.27 / 0.00 | 0.79 | 0.56 / 0.60 |
| HDDS\* | 0.69 | 0.89 / 0.00 | 0.51 | 0.49 / 0.00 | 0.75 | 0.77/ 0.00 |

**Table 4: Data Segmentation Comparisons (keep the same testing set)**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Overall | Food Security Measure | By country  With oversample and ML | One dataset  With oversample and ML | Rural Clusters only  With oversample and ML |
| Malawi  predict 2015/16 | FCS | 0.93 / na | 0.00 / na | 0.01 / na |
| HDDS | 0.68 / na | 0.62 / na | 0.46 / na |
| rCSI | 0.80/ 0.00 | 0.08 / 0.00 | 0.53 / na |
| Tanzania  predict 2014/15 | FCS | 0.63/0.00 | 0.00 / 0.00 | 0.19 / 0.00 |
| HDDS | 0.66/na | 0.96 / na | 0.97 / na |
| rCSI | 0.44/ 0.00 | 0.08 / 0.00 | 0.86 / 0.00 |
| Uganda  predict 2012 | FCS | 0.24/ 0.00 | 0.00 / 0.00 | 0.16 / 0.00 |
| HDDS | 0.89/0.00 | 0.89/ 0.00 | 0.91 / 0.00 |

****

**Fig. 1 Map of FCS**

****

**Fig. 2 Top tree split**

****

**References**

Babenko, Boris, et al. "Poverty Mapping Using Convolutional Neural Networks Trained on High and Medium Resolution Satellite Images, With an Application in Mexico." *arXiv preprint arXiv:1711.06323* (2017).

Barrett, Christopher B., and Derek Headey. "A proposal for measuring resilience in a risky world." (2014).

Beegle, Kathleen, et al. *Poverty in a rising Africa*. The World Bank, 2016.

Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. "Predicting poverty and wealth from mobile phone metadata." *Science* 350.6264 (2015): 1073-1076.

Castelluccio, Marco, et al. "Land use classification in remote sensing images by convolutional neural networks." *arXiv preprint arXiv:1508.00092* (2015).

Chen, Derek. "Temporal Poverty Prediction using Satellite Imagery." (2017)

Chen, Xi, and William D. Nordhaus. "Using luminosity data as a proxy for economic statistics." *Proceedings of the National Academy of Sciences* 108.21 (2011): 8589-8594.

Dang, Hai-Anh, Dean Jolliffe, and Calogero Carletto. "Data gaps, data incomparability, and data imputation: a review of poverty measurement methods for data-scarce environments." (2017).

Donaldson, Dave, and Adam Storeygard. "The view from above: Applications of satellite data in economics." *Journal of Economic Perspectives* 30.4 (2016): 171-98.

Elkan, Charles. "The foundations of cost-sensitive learning." In *International joint conference on artificial intelligence*, vol. 17, no. 1, pp. 973-978. Lawrence Erlbaum Associates Ltd, 2001.

Engstrom, Ryan, Jonathan Hersh, and David Newhouse. "Poverty from space: using high-resolution satellite imagery for estimating economic well-being." (2017).

Head, Andrew, et al. "Can Human Development be Measured with Satellite Imagery?" *Proceedings of the Ninth International Conference on Information and Communication Technologies and Development*. ACM, 2017.

Henderson, J. Vernon, Adam Storeygard, and David N. Weil. "Measuring economic growth from outer space." *American economic review* 102.2 (2012): 994-1028.

Jean, Neal, et al. "Combining satellite imagery and machine learning to predict poverty." *Science* 353.6301 (2016): 790-794.

Kussul, Nataliia, et al. "Deep learning classification of land cover and crop types using remote sensing data." *IEEE Geoscience and Remote Sensing Letters* 14.5 (2017): 778-782.

Pokhriyal, N., & Jacques, D. C. Combining disparate data sources for improved poverty prediction and mapping. *Proceedings of the National Academy of Sciences*, *114*(46) (2017), E9783-E9792.

Steele, Jessica E., et al. "Mapping poverty using mobile phone and satellite data." *Journal of The Royal Society Interface*14.127 (2017): 20160690.

Ray M. Tischio, and Gary M. Weiss. Identifying Classification Algorithms Most Suitable for Imbalanced Data, *Proceedings of the 15th International Conference on Data Science* 2019, Las Vegas, NV.

G.M. Weiss. Mining with Rarity: A Unifying Framework. *SIGKDD Explorations*, 6(1): 7-19, 2004.

Weiss, Gary M., Kate McCarthy, and Bibi Zabar. "Cost-sensitive learning vs. sampling: Which is best for handling unbalanced classes with unequal error costs?." Dmin 7, no. 35-41 (2007): 24.