Predicting food insecurity

Michael Feffer, Parv Kapoor, Sebastian Dodt
[1]Carnegie Mellon University

Abstract

In its longest drought in four decades, Somalia and other countries in the Horn of Africa are currently facing an unprecedented humanitarian crisis. 40% of the Somali population (six million people) are estimated to be impacted by severe acute food insecurity (WFP 2022c). The United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA) has released additional funds for humanitarian aid in the region based on predictive analytics of its novel Anticipatory Action (AA) Team. Releasing "trigger warnings" when food insecurity levels are predicted to reach critical levels, AA employs data from partners that leverages manual and qualitative methods for food insecurity classification and prediction. The frequency of food insecurity prediction is 2-4 times per year making any subsequent modelling ill-equipped to account for the dynamic nature of natural catastrophes and conflicts. Additionally, the prediction accuracy has been subject to scrutiny. In this paper, we describe an AI-based approach that can automate this system and more accurately forecast levels of hunger in real-time. We hope that our results serve as a foundation for the use of AI in humanitarian relief and lead to more targeted aid for vulnerable populations.

Introduction

Despite the world's steadily increasing per capita food production, food insecurity remains an unsolved problem affecting the lives of more than 700 million people worldwide (Nations 2022; FAO 2022). Facing both anthropogenic climate change and worsening conflict, the Horn of Africa is facing its worst humanitarian crisis in forty years, putting millions at risk of famine and undernourishment.

To make humanitarian assistance faster, more efficient, and more dignified, the United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA) has launched an Anticipatory Actions (AA) team that is attempting to predict food insecurity and reduce its detrimental impact by intervening earlier than common, post-facto humanitarian assistance. This has the potential to improve UN OCHA's response to crises but is limited by several constraints. First, lack of funding has forced the United Nations Secretariate and its agencies to make difficult decisions about the distribution of its limited budget. Second, there is currently little political support for a more quantitative approach to the dissemination of humanitarian aid. As a result, current efforts rely on simplistic statistical decision-making and time-consuming, manual prediction processes.

So far, prediction of food security has relied on target labels derived from these slow and subjective processes, allowing teams like AA to only update their trigger algorithms 2-4 times per year. In contrast, we have designed models using the latest machine learning techniques to predict food insecurity levels more accurately and in real-time. In this paper, we use Somalia as a case study and describe the outcomes of our modeling and results. We argue that predictions from our models have the potential to help AA and other organizations achieve their goals of a more rapid, efficient, and fair distribution of humanitarian assistance.

Related Work

Integrated Food and Security Phase Classification (IPC) Ratings

The Integrated Food and Security Phase Classification (IPC) is a rating scale that categorizes food insecurity and famine risk for a given region (Scicchitano 2019). It is a five-point scale where 1 denotes a secure region and 5 denotes a famine-affected region.

The organization that determines IPC levels for regions does so in a manual and somewhat subjective process. This is evident in a report of the organization's Advanced Technology and Artificial Intelligence (ATARI) initiative stating "that current IPC processes—while well established and developed over the past 17 years—have not caught up to the technological possibilities that currently exist or will exist in the near future" (ATARI Initiative 2021). To that end, the initiative's report goes on to state that capabilities for forecasting and nowcasting (using predictive models to impute data missing in the present) food (in)security and indicators of food (in)security to calculate IPC values would be highly complex but also would have high impact (ATARI Initiative 2021).

UN OCHA Anticipatory Action

The United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA) is one downstream organization that uses inputs like IPC and other indicators of crisis or famine to deliver aid to affected regions. A new framework, anticipatory action (AA), was recently tested in Somalia (Gettliffe 2021). The framework relies on a system of triggers that is activated once food insecurity reaches a certain level, at which point the trigger "releases money to activate the required delivery mechanisms and actions" for aid (Gettliffe 2021). The pilot ran into numerous difficulties, including but not limited to the lack of "frequency and timing of food security projections", but nevertheless, the overall response from the participants was positive and included "a sense that the traditional modes of humanitarian response are outdated, given the predictable and recurrent nature of so many shocks, and the greater efficiency (in financial terms) and human dignity of responding in anticipatory ways" (Gettliffe 2021).

These issues and others from the ATARI report are echoed in Lentz and Maxwell (2022). Lentz and Maxwell describe six challenges of information-driven anticipatory action. All of these challenges demonstrate that while data about food insecurity and relevant indicators exists, it can be difficult to obtain and make sense of this data, let alone actually make meaningful, accurate, and objective predictions about things like war or famine (Lentz and Maxwell 2022).

Existing Predictive Modeling Approaches

Despite all of these difficulties, various teams and researchers over the years have attempted to model and predict food insecurity. The Famine Early Warning Systems Network (FEWS NET) is widely regarded as a leading resource for food insecurity data especially for Sub-Saharan Africa (Backer and Billing 2021). An analysis of their released predictions compared to ground-truth actual numbers over the past decade for African countries revealed that their accuracy was approximately 84% (Backer and Billing 2021).

FEWS NET's process is manual, but other approaches employ statistical modeling and machine learning with the goal of eventually automating parts of such processes and making forecasts more readily available. One such project uses autocorrelation to make "one-ahead forecasts" given previous levels of food insecurity and relevant indicators (Wang et al. 2020). Minor modifications to their model to incorporate Bayesian priors allowed them to take expert opinion into account when calculating most likely model parameters.

Andree et al. (2020) provide multi-level forecasting and prediction. The sub-national prediction focuses on monthly predictions and evaluates associated costs with different weighting strategies for false positives to false negatives. The second level of prediction is focused on individual countries and focuses on the percentage of population impacted in the crisis affected districts of the country. They also use food insecurity data from FEWS NET dataset. On top of food crisis outbreak prediction, the authors also predict other food insecurity correlated indicators like violent conflicts, environmental factors, food price inflation et cetera. Their output variable is a binary variable with value 1 indicating food crisis (IPC 3,4,5) or 0 if not observed. They use the random forests technique for prediction and outperform the baseline methods which are conservative and have a high false negative ratio. Their method also allows them to differentiate impact of specific intervention techniques and provide long horizon future predictions (12 months) which increase the lead time for preventative action.

Zhou et al. (2021) compares tree-based methods (gradient-boosted trees and random forests) to logistic regression results on two binary prediction problems for outcomes 1 month ahead at the village level. The first involved predicting whether 20% of the village was food insecure (roughly corresponding to IPC levels of 3 or greater), and the second involved predicting whether the average household was food insecure (roughly corresponding to IPC level 5). Testing results on data from three African countries (Malawi, Tanzania, and Uganda) revealed that the tree-based approaches outperformed logistic regression (even more so when using upsampling techniques for the second prediction problem) but noted in their discussion and conclusion that their methods are brittle in the face of "conflict and unanticipated disasters" (Zhou et al. 2021).

Westerveld et al. (2021) also uses data from a variety of different sources with gradient-boosted trees to predict IPC for areas of Ethiopia. They found that their model worked better for longer time horizons (7 months) than shorter ones (3 months) but generally outperformed baseline classifiers).

Lastly, the "Hunger Map" by the World Food Programme (WFP 2022b) is a project to map food insecurity in real time. Most of the data displayed in the interactive map is obtained from phone interviews, and they impute missing data via "nowcasting" based on outputs from gradient-boosted tree models (WFP 2022b).

Methods

Data

Choosing Somalia as our case study, we have faced two challenges in the collection of data. First, panel data for the 74 subregions of Somalia is scarce given its limited digital infrastructure. Second, food insecurity cannot be solely attributed to one source but is an interaction between economic, natural, and conflict factors (Appendix, Table 4).

In our supervised models, our categorical target variable was the IPC level of the 74 subregions of Somalia, classified retrospectively by FEWS NET quarterly from July 2009 to October 2020 (NET 2022). Each classification was associated with a near-term and medium-term prediction of the IPC level in the respective region at the time. These predictions helped us calculate FEWS NET prediction accuracy in Somalia as a baseline for our models.

For the features of our model, we used conflict, demographic, socio-economic, geographic, and weather data. Conflict data was obtained from the non-governmental organization ACLED (The Armed Conflict Location & Event Data Project) that monitors conflicts in every country in the world. ACLED's accuracy and reliability of their data has been confirmed in numerous studies and peer review (ACLED 2017). Conflict events (e.g., violence against civilians, protests, battles) were aggregated for each quarter and subregion of Somalia and merged with the IPC classification of the respective subregion at the end of the time period. All values were divided by the number of inhabitants in the region to account for varying sizes of regions.

We added weather data that we received from our collaboration with UN OCHA's Anticipatory Actions team. The variables were aggregated to show the monthly absolute precipitation as well as relative rainfall compared to historical seasonal average.

From the World Food Programme, we obtained the average price of corn in dozens of monitored markets across Somalia. We feature engineered the data and added the average price of corn at the two closest markets to each observation (WFP 2022a).

As demographic and socio-economic predictors, we used data from the World Bank. The data included 1,443 variables with yearly observations for all of Somalia. We added the most relevant features to our existing dataset: Unemployment and child labour rates, exports, urban growth, rural water access, and public debt (World Bank 2022). Additionally, we obtained real-time currency exchange information of the Somali shilling versus US Dollar. In a feature engineering step, we used this data to construct the following variables: average exchange rate, minimum/maximum exchange rate, change within quarter, and volatility (difference between maximum and minimum) (Investing 2022).

To control for regional and temporal fixed effects, we included geographical variables and a time variable as well as the last IPC classification for the respective region.

Models

We trained and evaluated several different classification models in their ability to predict ground-truth IPC classification three months in the future for a region of Somalia. While one can apply time series models to this problem (as others like Wang et al. (2020) have done), we did not consider them in our experiments. Analyzing the data prior to modeling to produce plots like Figure 1 reveals that at least in the case of Somalia, IPC typically does not have a seasonal pattern that time series models can utilize to generate useful predictions. On the contrary, it varies wildly across some periods and is stable across others.

As a result of this phenomenon and our limited data, we did not experiment with models that explicitly account for lagged features or autocorrelation in their construction (as methods like ARIMA or recurrent neural networks do), but because prior IPC classifications are still correlated with future ones, we did utilize the last IPC value as an input feature alongside data reflecting the current state of the region to predict IPC in the future. Thus, our prediction problem entailed predicting future IPC given present data and implicitly lagged labels.

To explore this problem, we considered multiclass versions of random forests, gradient-boosted trees, support vector machines, gaussian processes, and shallow feedforward neural networks. During training, we also performed upsampling or sample weighting for some models to adjust for class imbalance issues in our dataset given that IPC levels of 1 and 4 are rare (and we never observe an IPC level of 5 in our data). We tuned hyperparameters and class imbalance mitigation for all models via 5 trials of temporal holdout

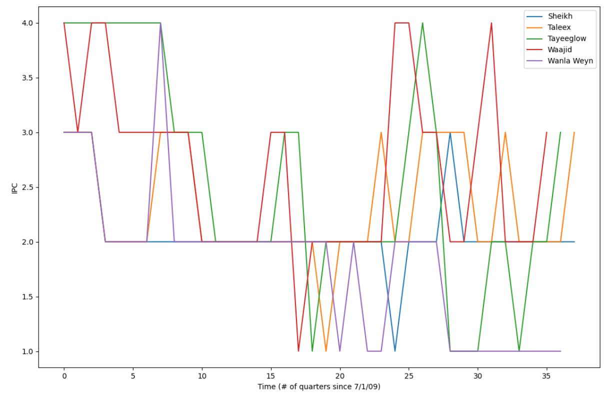

Figure 1: IPC versus Time (in quarters since July 1st, 2009) for several regions of Somalia. IPC lacks seasonal patterns and is sometimes stable from one quarter to the next. We opt against using time series models for these reasons.

validation. As baseline models, we additionally considered dummy models that predict based on the most frequently occurring label in the training set, human predictions from FEWS NET, and random forests that do *not* use the last IPC value as input (inspired by Andree et al. (2020)). We tuned the hyperparameters and imbalance mitigation scheme of the latter random forests baseline model also via temporal holdout validation. We also performed a gridsearch over various hyperparameter configurations for random forests.

We used Keras (Chollet et al. 2018) with TensorFlow (Abadi et al. 2016) for neural networks, XGBoost (Chen and Guestrin 2016) for gradient-boosted trees, scikit-learn (Pedregosa et al. 2011) for all other models, and imbalanced-learn (Lemaître, Nogueira, and Aridas 2017) for upsampling. Hyperparameter values and class imbalance mitigation strategies for each model explored during their respective grid searches are shown in Table 3 located in our Appendix.

Results

For each trial of temporal holdout validation, we evaluated the accuracy of the best models on holdout test data and averaged the results. Table 1 displays these accuracy values alongside the majority class baseline and human performance from FEWS NET.

While the random forests baseline inspired by Andree et al. (2020) performs well, it is outperformed by gradient-boosted trees with sample weighting on the first holdout test set and a shallow neural network on the second. Moreover, these models produce better predictions than those of FEWS NET in the medium-term and long-term for the same time periods.

However, our models, baselines, and human performance alike suffer significantly on the October '20 test set relative to the August '20 test set. This drop in accuracy across all approaches suggests a temporal domain shift that is difficult to handle.

Model	Imb Mit	Aug ’20	Oct ’20
RF Baseline	Rand. Oversampling	0.97	0.56
NN	N/A	0.81	**0.58**
GBT	Sample Weighting	**0.99**	0.52
SVM	N/A	0.73	0.55
RF w/IPC	N/A	0.58	0.52
Maj. Class	N/A	0.69	0.58
GP	N/A	0.97	0.55
FEWS NET	N/A	0.92*	0.54**

Table 1: Models with class imbalance mitigation schemes and performance on holdout test sets of the best models. The drop in performance from Aug ’20 to Oct ’20 suggests temporal domain shift.
*: short-term prediction in June 2020
**: medium-term prediction in June 2020

Since this is a multiclass classification problem, we primarily considered accuracy because it is difficult if not impossible to use evaluation metrics more suited for binary classification and because accuracy captures prediction correctness over all classes. That being said, we also calculated average confusion matrices pertaining to our two best performing models and display the results in Table 2.

Model	Conf. Mat. Aug ’20				Conf. Mat. Oct ’20			
NN	**6.8**	0.8	0	0.4	**12.6**	2.4	0	0
	0	**46.6**	0.4	1	2.2	**23.8**	4.4	0.6
	0	8.2	**0.8**	0	1	19.8	**0.2**	0
	0	0	0	**0**	0	0	0	**0**
GBT	**10**	0	0	0	**10**	4	1	0
	0	**47**	1	0	0	**24**	7	0
	0	0	**9**	0	0	20	**1**	0
	0	0	0	**0**	0	0	0	**0**

Table 2: Confusion matrices with results from the two best models from Table 1 with bolded terms denoting numbers of correctly predicted examples. They suggest that it is hardest to distinguish between IPC classes 2 and 3.

The drop in performance that can be attributed to temporal domain shift can also be seen here in the increase in number of misclassified examples from one test set to the other. In addition, given that the shallow neural network has issues distinguishing between IPC classes 2 and 3 in both test sets and the decrease in accuracy for the gradient-boosted tree largely comes from failing to distinguish between those two classes, these results also imply that IPC classes 2 and 3 are inherently difficult to separate.

Discussion

Modeling Difficulties Forecasting is generally a challenging problem, and our efforts are further limited by the lack of data available. We posit that our models are unable to adapt to the new temporal domain due to this lack of data. Class imbalance also made this problem complex, and imbalance mitigation strategies did not work uniformly across modeling approaches (e.g. random oversampling worked best for the random forests baseline but sample weighting worked best for gradient-boosted trees).

Modeling Uncertainty While limited data and class imbalance negatively impact all of our models, some of them are still able to model the uncertainty in their predictions. Specifically, gaussian processes and neural networks output probability distributions over all classes instead of hard labels instead of fractions of classes at leaf nodes as tree-based approaches do. These probability distributions can be interpreted as uncertainty in predictions and help humans with aid prescription and decision-making. Moreover, in the case of gaussian processes, kernel crafting and prior configuration could be used to factor human expertise into the modeling process as per Wang et al. (2020).

Ethical Considerations We realize these models have ethical and real-world implications. Our models have difficulty distinguishing between IPC levels of 2 and 3, and while this is a pitfall from a modeling perspective, such errors could make the difference between a region receiving or not receiving aid. More generally, modeling predictions can affect the livelihood of millions of people, especially if these models are used for numerous countries. These predictions therefore should be free from as much human bias and error as possible, but this is beyond the scope of our work. If there is a desire to deploy these models, the deployment should be done in collaboration with human experts and relevant stakeholders who have the ability to halt the process if they are uncomfortable with these implications.

Future Work An important next step to continue this research is to improve our modeling approaches, primarily by obtaining more data. Our data only went back as far as July of 2009, which does not translate into many quarters to use for data-hungry modeling algorithms. Using data from other countries or even digitizing hard-copy archives of historic data from NGOs could be viable solutions in this regard. Additionally, more informed application of class imbalance mitigation strategies would also be helpful going forward. Lastly, working alongside NGOs to develop aid allocation strategies given modeling results and ensuring that the modeling outcomes are in-line with human values are also important remaining tasks.

Conclusion In this paper, we have presented a new method to derive predictions for food insecurity. If implemented by international humanitarian aid organizations, it can make humanitarian assistance more rapid, efficient, and more equitably distributed. Building upon existing methods, we showed that automated processes can have higher accuracy than FEWS NET’s manual predictions. Importantly, it allows for daily if not hourly predictions of food insecurity and thereby offers a significant advantage over bi-annual to quarterly predictions that may be published too late for trigger warnings. We conclude that there is significant potential for machine learning to have a positive impact on the field of humanitarian assistance.

Acknowledgments

We thank Prof. Fei Fang for her inputs about our problem and modeling approaches, and we additionally thank Leonardo Milano and Monica Turner from the United Nations Office for the Coordination of Humanitarian Affairs (UN OCHA) for their input about what would be most useful to the NGO community.

References

Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. 2016. {TensorFlow}: A System for {Large-Scale} Machine Learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, 265–283.

ACLED. 2017. Armed Conflict Location Event Data Project (ACLED): Guide for Media Users. *Reliefweb* URL https://reliefweb.int/sites/reliefweb.int/files/resources/ACLED_User-Guide-for-Media_2017_FINAL.pdf.

Andree, B. P. J.; Chamorro, A.; Kraay, A.; Spencer, P.; and Wang, D. 2020. Predicting food crises .

ATARI Initiative. 2021. IPC ATARI Report 1: Key Opportunities for Advanced Technologies and Artificial Intelligence in IPC.

Backer, D.; and Billing, T. 2021. Validating Famine Early Warning Systems Network projections of food security in Africa, 2009–2020. *Global Food Security* 29: 100510. ISSN 2211-9124. doi:https://doi.org/10.1016/j.gfs.2021.100510.

Chen, T.; and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

Chollet, F.; et al. 2018. Keras: The python deep learning library. *Astrophysics source code library* ascl–1806.

FAO. 2022. Food Balances. *FAOSTAT* URL https://www.fao.org/faostat/en/#data/FBS.

Gettliffe, E. 2021. UN OCHA Anticipatory Action. Lessons from the 2020 Somalia Pilot .

Investing. 2022. USD/SOS - US Dollar Somali Shilling. *Investing.com* URL https://www.investing.com/currencies/usd-sos-historical-data.

Lemaître, G.; Nogueira, F.; and Aridas, C. K. 2017. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *The Journal of Machine Learning Research* 18(1): 559–563.

Lentz, E.; and Maxwell, D. G. 2022. Anticipating, Mitigating, and Responding to Crises: How Do Information Problems Constrain Action? *Available at SSRN 4043459* .

Nations, U. 2022. Global issues: Food. *UN* URL https://www.un.org/en/global-issues/food.

NET, F. 2022. Food Security Classification Data. *FEWS NET Data Center* URL https://fews.net/fews-data/333.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* 12: 2825–2830.

Scicchitano, J. S. 2019. International measurement of food security: Enhancing alignment between evidence and assistance programs. *Journal of Public Affairs* 19(3): e1837.

Wang, D.; Andree, B. P. J.; Chamorro, A. F.; and Girouard Spencer, P. 2020. Stochastic modeling of food insecurity .

Westerveld, J. J.; van den Homberg, M. J.; Nobre, G. G.; van den Berg, D. L.; Teklesadik, A. D.; and Stuit, S. M. 2021. Forecasting transitions in the state of food security with machine learning using transferable features. *Science of the Total Environment* 786: 147366.

WFP. 2022a. Food Prices Somalia. *VAM - Food Security Analysis* URL https://dataviz.vam.wfp.org/economic_explorer/prices.

WFP. 2022b. HungerMap LIVE. URL https://hungermap.wfp.org/.

WFP. 2022c. No rain and no resources; millions of families across the horn of Africa pushed closer to catastrophe each day. *WFP News Releases* URL https://www.wfp.org/news/no-rain-and-no-resources-millions-families-across-horn-africa-pushed-closer-catastrophe-each.

World Bank. 2022. Somalia. *World Bank* URL https://data.worldbank.org/country/SO.

Zhou, Y.; Lentz, E.; Michelson, H.; Kim, C.; and Baylis, K. 2021. Machine learning for food security: Principles for transparency and usability. *Applied Economic Perspectives and Policy* .

Appendix

Model	Hyperparams	Imb. Mit.
RF Baseline	Models in Andree et al. (2020)	None Rand. Oversampling
RF w/IPC	Models in Andree et al. (2020)	None Rand. Oversampling
GP		None
SVM	Kernel: Linear, RBF C: $\{$1e-7,1e-6,1e-5, 1e-4,1e-3,1e-2,1e-1,1$\}$	None Rand. Oversampling
GBT	max_depth: i in range [3, 11]	None sample weighting Rand. Oversampling
NN	Structure: $\{$[128, 64, 32, 16, 4]$\}$ Activation: ReLU and softmax	None class weighting Rand. Oversampling

Table 3: Hyperparameter values and class imbalance mitigation strategies used during grid search for optimal models.

Variables	**d**	**Region-varying**	**Description**
IPC level	1	yes	Dependent variable; post facto IPC classification by FEWS NET
Last IPC	1	yes	The last observed IPC value for the subregion
Dummies	89	yes	Two OHE vectors of 18 regions and 73 subregions of Somalia
Time fixed effects	1	no	time variable, one observation per 3-5 months
Spatial fixed effects	2	yes	Latitude and longitude of region
Population	1	yes	Population of region
Weather	5	yes	Precipitation data: absolute and relative to seasonal norm; mean, maximum and minimum
Economic	8	no	Unemployment rates (male and female), child labour (male and female), exports, urban growth, rural water access, public debt; all variables are yearly
Food prices	1	yes	Average price of corn at the two closest recorded markets (in Somali Shilling)
Exchange rate	5	no	Exchange rate of Somali Shilling to US Dollar; absolute and change, quarterly maximum and minimum
Conflict data	8	yes	Fatalities in conflict, battles, explosions, protests, riots, strategic developments, violence against civilians; per capita

Table 4: Dependent and independent variables