Early Warning and Early Action for Increased Resilience of Livelihoods in the IGAD Region

Report 3. Predictive Analytics and Machine Learning Approaches to Support EW-EA

A Feinstein International Center Working Paper

Erin Lentz, Cori Simmons, Gregory Gottlieb, and Daniel Maxwell

June 2021

Tufts University
Friedman School of Nutrition Science and Policy
Feinstein International Center

Food and Agriculture Organization of the United Nations

Sida

Table of Contents

Acronyms

ACAPS	Assessment Capacity Program
ACLED	Armed Conflict Location & Event Data Project
ADASYN	adaptive synthetic sampling
AI	artificial intelligence
API	application programming interfaces
ARIMA	autoregressive integrated moving average
ATARI	Advanced Technologies and Artificial Intelligence
CERF	Central Emergency Response Fund
CGIAR CIAT	Consultative Group on International Agricultural Research International Center for Tropical Agriculture
CHIRPS	Climate Hazards Group InfraRed Participation with Station data
CNN	convolutional neural networks
CRS-MIRA	Catholic Relief Services Measurement Indicators for Resilience Analysis
DEWS	Drought Early Warning System
DHS	Demographic and Health Data
EA	early action
ENSO	El Nino Southern Oscillation
EW	early warning
EW-EA	early warning/early action
EWS	early warning system
FAM	World Bank Famine Early Action Mechanism
FAO	UN Food and Agriculture Organization
FBF	forecast-based financing
FCS	food consumption score
FEWS NET	Famine Early Warning System Network
GAM	global acute malnutrition
GRACE	Gravity Recovery and Climate Experiment satellites
HDDS	Household Dietary Diversity Score
HFIAS	Household Food Insecurity Access Scale
IGAD	Inter Governmental Authority on Development
IFPRI	International Food Policy Research Institute
IO	international organization
IPC	Integrated Phase Classifications
KII	key informant interview
LASSO	least absolute shrinkage and selection operator
LEAP	Livelihoods, Early Assessment and Protection
LIAS	Livelihood Impact Analysis Sheet
LSMS	Living Standard and Measurement Survey
MERIAM	Monitoring Early Risks Indicators to Anticipate Malnutrition
ML	machine learning
NDVI	normalized difference vegetation index
NGO	non-governmental organization

OCHA	Office for the Coordination of Humanitarian Affairs
PA	predictive analytics
PHYGROW	Phytomass Growth Model
PIN	population in need
PLEWS	Predictive Livestock Early Warning System
PSNP	Ethiopia Productive Safety Net Program
QA/QC	Quality Assurance/Quality Control
rCSI	Reduced Coping Strategies Index
RZSM	root zone soil moisture
SADC	Southern African Development Community
SMART	standardized methods for assessment of relief and transition
SMOTE	synthetic monitoring oversampling technique
SPI	Standardized Precipitations Index
SSA	Sub-Saharan Africa
UN	United Nations
UNICEF	UN Children's Fund
USAID	US Agency for International Development
WASH	water, sanitation, and hygiene
WB	World Bank
WFP	World Food Programme

A. Overview

Predictive analytics (PA) for EW systems hold great potential *for East Africa EW-EA information systems.* The introduction of much higher-speed computers and increasing availability of data has allowed predictive analytics (including machine learning, big data, and statistical modelling) to take off (Varian 2014; Hernandez and Roberts 2020). PA is "technology that learns from experience [historical data] to predict the future behavior of individuals in order to drive better decisions" (Siegel 2016, 15). Predictive analysts often use techniques of **machine learning** (ML) to generate predictions. ML is one type of **artificial intelligence (AI)** used in humanitarian contexts; other forms of AI, including robotics, artificial swarm intelligence and image classification tools, do not prioritize predictions. In a recent review, Hernandez and Roberts (2020) documented 49 projects using PA in the humanitarian space. However, the potential of PA for food security early warning has not yet been reached: only four of the projects in their review focused on food security (Hernandez and Roberts 2020). Nonetheless, numerous projects are in development or are in the "proof of concept" phase.

PA has recently received a great deal of interest in the practitioner, researcher, and donor communities because it offers a way to synthesizing large amounts of data to generate diagnostic evidence. PA can be used at multiple stages in the humanitarian diagnostic and analysis system. It can be used to predict hazards (e.g., future drought) or predict outcomes (e.g., future food security status). These forecasts can feed into scenario building or directly trigger responses (e.g., forecast based financing). PA can also be used to estimate current status ("nowcasting") for locations where data are not available (e.g., WFP's Global Hunger Map). Other AI tools can also facilitate reaching conclusions about an assessment, for example by weighing analysts' responses in an online forum (e.g., artificial swarm intelligence). See Box 1 for more details on machine learning.

Some people propose that PA models are more objective than current EW tools because the findings are generated by the model (O'Neil 2016; Coyle and Weller 2020; 044). Others believe that PA models can bury assumptions. Machine-learning introduces significant challenges regarding transparency and explicability (Coyle and Weller 2020). While all models are driven by data and assumptions, PA models often lack transparency about what assumptions have been made (sometimes implicitly) and how these assumptions influence the outcomes (O'Neil 2016; D'Ignazio and Klein 2020; Coyle and Weller 2020; 052). As a result, it may be harder to build trust in machine-learning and other AI outcomes compared to consensus-based approaches (Coyle and Weller 2020; 019).

Machine learning techniques "optimize a single objective function" and are rigid (Coyle and Weller 2020, p. 1343). That is, they are designed with a specific goal in mind (e.g., trigger funding when the IPC is 3 or above). As a result, it is critical that modelers, decisionmakers and analysts understand and carefully define the goals, risks and assumptions embedded within that objective function. Further, in some cases, these results may not be what policymakers need for decision-making. Policymakers may have (implicit) competing goals or tradeoffs and may prefer to strike a balance among various objectives rather than seek to optimize a single objective function.

The successful incorporation of predictive analytics into a humanitarian diagnostics system will require modelers, analysts and decisionmakers to make choices about what their objectives for the model are, and what they prioritize in their models. Some choices depend on the broader goals of the humanitarian diagnostic system. Others depend on data availability, capacity, and usability. Collaboration and transparency between end users and modelers are essential to ensuring models are solving the right problems.

To make transparent these decisions, tradeoffs, and factors transparent (Coyle and Weller 2020), we first define a few key terms before discussing usability and the role of PA within the broader EW system. All models have an objective, or goal, with an outcome, and require predictors to predict that outcome. The following are definitions used throughout the report.

- **Objectives** are the goals of the model. Models may help end users to triangulate findings, to fill in gaps where data are missing, or to identify and target populations in need. Such models are part of scenario planning. Others PA models provide a "signal" and can trigger a response, such as drought insurance or funding due to a change in IPC status. Models can vary by spatial scale, by predictive window, and by generalizability, among others. These model objectives are discussed in detail in Section III.
- **Outcomes of models** can include malnutrition status (e.g., MERIAM project), food security status (e.g., rCSI or FCS), estimated IPC phases, or estimates of the likelihood of a hazard (e.g., meteorological forecasts of flooding). Models could also predict estimates of PIN, although we are unaware of any models currently doing so.
- **Predictors (independent variables)** can be high frequency observations (e.g., meteorological data, price data), can be low frequency (e.g., changes in roofing material), or even single observations for a particular area (e.g., geographic attributes, such as proximity to a river). These data can also be at high spatial disaggregation (e.g., village level) or more aggregated (e.g., at nation level). Many models use a mixture of high and low frequency data; each predictor is generally scaled to the same spatial level as the outcome (e.g., if outcome is the IPC scale, the predictors may be aggregated to the IPC livelihood-zone).
- **Models:** Predictive Analytics (PA) prioritizes prediction of outcomes using historical data. Machine learning techniques are commonly used tools to generate such predictions. Because most models rely on historical data, any unmodeled changes (e.g., export bans, pandemics) may render findings obsolete. Further, models ought to be rigorously validated; yet this is often a major challenge due to data limitations. Several modeling decisions influence what kind of information PA can provide to early warning analysts and decisionmakers; we highlight these decisions below as well as provide a brief overview of machine learning.
- **Ethics:** Accountability, equity, transparency, and ethical use of data and modeling should inform decisions and model objectives. See **Section 6** for more details.

The remainder of this document discusses several predictive analytics (PA)-related topics and is organized by topics and question within each topic. We first briefly introduce a Table that reviews several recent efforts to model food security, nutrition, and related hazards. We draw on these findings for the remainder of the document. We frame a discussion on usability and the role of PA as a series of questions in Section 2. We provide a brief overview of machine learning for predictive analytics in Section 3. We then provide detailed information on the data for PA in Section 4 and modeling choices available to modelers and end-users in Section 5. Section 6 of the report describes findings, including issues of ethics, biases, privacy, and equity related to PA. Recommendations drawn from these findings are presented in Maxwell et al. (2021a) "Early Warning and Early Action for Increased Resilience of Livelihoods in IGAD Region."

The focus of this review is on predictive analytics for food security EW-EA and resilience. A detailed summary of models, including outcome measures, data requirements, and modeling choices is presented in Table 1. Table 1 is organized by outcome: food security and nutrition measures, IPC-based measures, and hazards.

The table show that there are tradeoffs across outcomes. Several models use either IPC classifications or transitions in IPC scales (e.g., from below 3 to 3 or above). These estimates tend to be at higher administrative levels (e.g., Admin 2) than models using nutrition outcomes or aggregated household food security measures, which can be estimated at Admin 1 and can support more granular targeting. However, models predicting food security or nutrition outcomes tend to have more limited historical data to draw on for estimation and validation. As a result, some models remain in the "proof of concept" phase, reflecting challenges identifying adequate data for external validation. While most food security and nutrition measures reviewed here focus on nowcasting, both early warning forecasting and nowcasting are pursued with IPC-based models. Prediction windows vary across models; there is not a best practice regarding "how early is early enough," although accuracy declines as the predictive window extends (Andree et al. 2020).

There is more research on the prediction of hazards than on the prediction of nutrition or food security outcomes. This likely reflects the richer high frequency data available to support prediction of meteorological hazards and their impacts. Among the hazard studies reviewed, the researchers commonly use secondary data to forecast the likelihood of hazards for scenario development Some pair secondary data with primary data for validation. Much of the modeling of hazards could be valuable for food security and resilience early warning; some models could inform triggers. Some new work has examined the use of hazard-based predictions (e.g., droughts and floods) for early action trigger-based systems, such as forecast-based financing and impact-based forecasting (IFRC 2020).

We briefly describe Sections B-E, to assist readers in identifying sections most of interest to them.

In Section B, we review the usability and role of predictive analytics in an EW system. This section is laid out in terms of addressing seven key issues, including engagement of modelers and EW stakeholders, how predictive analytics will complement other approaches, where the modeling should take place, how accountability to affected populations can be built into predictive analytics, what constitutes "good enough" analytics, bridging the humanitarian/development nexus and the role of donors.

Section C is an overview of machine learning, providing a brief overview of how the techniques of machine learning can support predictive analytics. Section D is about modeling—specifically about outcomes and predictors and how these are chosen, which outcomes and which predictors should be included in models, and when existing data are adequate and when models require new or novel data. Section 5 is on modeling choices and tradeoffs related to the objectives for models. This section is also broken up by key technical questions including the length of the predictive time window, the desired spatial resolution, the choice of scenarios or triggers, location and livelihood specificity of models, the accuracy of models, the interpretation of models and some challenges for the future.

The final section is on findings related to PA/ML. The same findings are in the main report (Report 1). Table 1 at the end provides and overview on selected predictive analytic models that focus on food security and related hazards.

B. Usability and Role of PA within Broader EW System

Sustained, in-country investment and engagement are requirements for PA models to deliver on their promises (069). For models to deliver on promises to fill EW gaps and needs, modelers, EW technical analysts, decisionmakers and donors should consider the following questions on coordination, capacity, and usability, among others.

1. How should modelers engage and coordinate with stakeholders for input and buy-in?

A number of key informants from the modeling community have little formal coordination with end users, reporting uncertainty about who to engage with, when and how (033; 042; 059). A few modelers want to get a model working well before pitching/promoting it. A concern is because end users are so busy, modelers want something relatively polished to share because there may only be one opportunity to make a good impression (042; 059). COVID-19 and locusts contribute to the perception that end users in East Africa are extraordinarily busy. Others aim to have conversations early in the development of models with end users, although modelers report that it can be hard to find the right EW systems personnel with whom to coordinate (059; 060; 070).

When modelers do find the right individuals to speak with, most report it is extremely valuable (033; 060; 061). One modeler (069) based at headquarters argues that working with local country offices is critical to building the most appropriate model and building ownership of that model locally. One modeling group (071) argues that field-research and conversations with NGOs operating in an area, government officials, and impacted community members can help to understand how to turn data into interpretable indicators, for example through the identification of contextually appropriate cut-offs.

2. How will predictive analytics complement other EW approaches?

Few modelers expect EW models to “stand alone,” fully replacing current food security EW systems. This is consistent with Hernandez and Roberts (2020) who found that most humanitarian predictive analytics projects aimed to complement existing systems rather than replace them. A few modelers perceive EW systems as having two different, but related challenges: (1) predicting high-frequency hazards’ (e.g., drought, price fluctuations) impacts on food security and (2) predicting acute, unexpected, rare, and or hard to predict things (e.g., pandemics, conflict, locusts) (044; 069). While opinions vary on where the balance is between what a model can do well and what an analyst should monitor, several modelers propose to models to be well suited to monitoring and predicting higher-frequency events. They suggest that human analysts will be better suited to identifying hard-to-predict drivers (019; 042; 044). This perspective means modelers (implicitly) expect human analysts to read model results against what else is happening (042; 044; 052; 060) and to be able to triangulate findings from several different information-generating processes.

This begs the question: how can human analysts assimilate findings from models with their other analyses? Identifying how to support analysts to make sense of different streams of information has not received much attention thus far by the modeling community. Analysts will likely need think across multiple information streams in order to create coherent scenarios. One important avenue for future research is understanding whether findings from PA models combined with findings from existing EW systems (e.g., IPC process another approach to humanitarian diagnostics) improve the overall accuracy of predictions.

One modeler KII (key informant interviewee) warns Early Warning systems are already in place, and “dropping” a PA model into an existing EW system is not straightforward (060), even if the goal is to provide complementary data for triangulation. It is likely that established systems cannot be easily overhauled to fit a PA model, although a few KII opined that EW systems could benefit from being flexibly designed so that they can incorporate new approaches (033; 060).

3. Will modeling occur in-house? If so, what are the data platform needs? What are the capacity needs?

Recurrent data gathering, data processing, and management of data platforms, and updating and interpreting the model all require institutional commitment and capacity building of both the analysts and decisionmakers. In sum, one respondent described modeling as "the easy part" (060). Several modelers (039, 059, 061, 069, 072) are having initial conversations with governments about government ownership. These conversations raise a host of difficult questions including: Where in the government should models reside? With what funding? How does a model substitute or complement existing systems? How much analytical capacity is required, and will there be a long-term investment to build data skills and capacity? Will data be available for updating it? etc. NGOs, UN agencies, and other IOs using PA face similar questions.

3a. Data platforms

Any EW system that pursues modeling "in-house" will also need to invest in data platforms and data systems that will be easily updated (e.g., application programming interfaces (APIs) to WFP data sources, etc.). Yet, many modelers acknowledge that data sharing, data management and data platforms gets less attention than they should (019; 027; 044; 052; 056; 060; 061). At least one respondent argued that too much attention is being spent on the predictive modeling and more effort should be spent on getting data systems and data sharing right (029). This would include addressing data sharing and management issues. It would also include production of a highly automated data platform that could assess data quality and process and visualize new data in a streamlined approach. This respondent commented "damn the analytics side" (029). Instead, provide the right data and an effective platform. One interviewee (061) argued that from the beginning, development and building of the data platform should be prioritized. Creating a data platform may require substantial computer science skills that needed to be outsourced while modeling could be done in house (061). Deparday et al. (2019) point out that there are emerging platforms for processed data (e.g., data that has classified images into roofs, trees, roads, etc.); relying on these platforms could help streamline data platforms within an EW technical office.

3b. Capacity

The capacity of food security analysts to interpret, use, and update PA models is a major concern (019; 070; 071). Specific concerns included helping analysts understand and communicate performance metrics to decisionmakers; and ensuring that modelers, technical analysts, and decisionmakers reach consensus on what is a statistically meaningful relationship; and reaching consensus on when the evidence is enough to support interventions (060; 070; 071). One way to support analysts is to address issues of capacity and how models fit within the broader humanitarian diagnostics architecture early in modeling decisions.

4. How can accountability to and participation of affected and at-risk populations be better incorporated into PA efforts?

This study has found few examples of incorporating affected populations in model development, in reporting, or in feedback on model outcomes (although see ACAPS South Sudan (2020) study for initiatives to incorporate affected populations). For model developers, the main end users are thought to be in four categories: governments, established EW providers (e.g., FEWS NET), donors and humanitarian actors. Rarely are at-risk communities mentioned as important end users of information (033; 056). Yet, foregrounding affected and at-risk populations can improve models and support fairness and inclusivity (Paul et al. 2019, Hernandez and Roberts 2020).

There are a variety of ways to better incorporate affected populations into the modeling process. One key informant suggested that that during the modeling development and processes it can be helpful to ask: who is in the room and who is speaking? Are people's experiences with hazards being incorporated, etc. (056)? A few modelers described participatory data collection and model development involving community-level actors (033, 039, 056). Qualitative work with individuals at risk of food insecurity may help identify overlooked factors, leading indicators, and local triggers as well as identify the information at risk populations would like communicated to them and when. The CRS-MIRA project (Knippenberg et al. 2019) is one example, which includes efforts to communicate findings back to community groups. Other examples include: developing a pilot livelihood forecast based financing project that seeks to incorporate local perspectives about shocks into the design of triggers etc.; learning what indigenous systems can tell us and which indicators tracked in indigenous systems are sensitive, given climate change and combinations of hazards; and collecting survey based information on local knowledge of drivers of food security in order to develop more locally adapted models and locally appropriate triggers (033; 056; 072).

5. What are the costs of inaction and when is the model good-enough?

An important area for future research is to assess the valued added of models. One approach is to compare, on average, the costs of early action against the costs of later actions. EA will have more inaccuracies (i.e., intervening when it wasn't needed (that is, "false positives") than waiting for a crisis to unfold. But imperfect models may still be good enough. On average, early action may be cheaper, even accounting for the false positives than intervening after crises commence and interventions are costlier (019). For example, suppose three crises occur. A model wrongly predicts five crises (i.e., it is wrong two of the five times). If early action costs $25 million, and regular action costs $50 million, the savings associated with correct early action is $25 million. In this case, being wrong 40 percent of the time is still a net cost savings relative to failing to predict any emergency early. Such calculations can help determine whether the model is "good enough."

Cabot Venton (2020) finds that an earlier and more proactive intervention, based on good early warning (whether PA or more traditional means) could have saved donors up to 30 percent of the costs of subsequent humanitarian spending. This would have amounted to $1.6 billion in just three countries (Kenya, Somalia, and Ethiopia) over the past 15 years.

6. Can PA help bridge the humanitarian-development divide and improve resilience?

Pairing longer term development interventions with shorter term humanitarian interventions has gained traction since the 2016 World Humanitarian Summit. Forecast based financing and impact-based financing have the potential to bridge the development-humanitarian divide by producing predictions that lead to both early action in response to humanitarian crises and longer-term development-based funding (E.g., both OCHA and WB are examining changes in IPC as a funding trigger, but OCHA funds for humanitarian interventions while WB is looking at recovery and longer-term development outcomes). Collaboration among humanitarians, development practitioners, and modelers will be necessary to most effectively leverage PA as a possible bridge. For example, data platforms can be integrated across the humanitarian-development spectrum, although specific data requirements and different PA approaches may be better suited to meeting the objectives of development and humanitarian information systems.

7. What role can donors play?

Several KII discussed the important role(s) donors could play in supporting the use of PA in EW-EA systems. (033, 056, 059, 060, 070). Donors could support the building of PA models, which may require more money up-front, even if the models might save money and or increase speed of data analysis later. Donors could also support the transformation of EW systems to be flexible enough to incorporate such models, for example by addressing data gaps (e.g., limited price data), encourage data sharing, supporting local- level engagement, and supporting data platforms. However, these KII acknowledged that it was challenging for (academic) modelers to work on donor schedule and that donors need education about what can and cannot be done with these models (as do others). While some KII were pursuing private sector partnerships, they acknowledged are challenges regarding tensions between proprietary work and public goods (033, 056, 070).

C. Overview of Machine Learning

Predictive analytics focuses on predicting an outcome, using historical data (Hernandez and Roberts 2020). Machine learning (ML) is a tool that predictive analysts may use to generate their predictions. Machine learning is a subfield of artificial intelligence. There are predictive ML techniques and non-predictive ML techniques that can support synthesis (e.g., artificial swarm intelligence) and classification (e.g., convolutional neural networks or CNN).

ML algorithms can be split into processes that are supervised by humans and processes that learn from the data without human input. Supervised ML models require training and test data. Models are "trained" on the data that include outcomes. The models generated using the training data are then tested on other data. These test data exclude the actual outcomes. Using the test data, the models generate predicted outcomes. Accuracy measures compute the rate of right guesses in the test data relative to the actual outcomes that were removed from the test data. Currently, most ML used for EW predictive analytics rely on supervised techniques.

Unsupervised ML techniques do not have external structure imposed on them; rather, techniques such as clustering analysis or principal component analysis identify categories with similar characteristics. It is up to the analysts to identify or "label" the results (Deparday et al. 2019, p. 8). Deparday et al. (2019) argue this distinction between supervised and unsupervised is blurring. For example, deep learning methods includes convolutional neural networks (used for image classification) and can be either supervised or unsupervised. Results from these unsupervised methods can be included as predictors in predictive analytic models (e.g., CNN is used to classify roofing using satellite imagery. The roofing findings becomes a predictor in a PA model for food security).

Several choices influence model building.

- **Nonlinear relationships and fat tailed data.** Machine learning techniques appear best suited when the underlying relationships are not necessarily linear and when there are fat tails (i.e., data heavily clustered around extreme values). ML may be particularly helpful for food security outcomes, where many people may be food secure (in one tail). Many others might be severely food insecure (in the other tail). We can use ML techniques to prioritize estimating the tail. Further, ML techniques can more easily account for (unknown) nonlinear relationships, such as "tipping points" (Krishnamurthy et al. 2020b), compared to regression models, which require modelers to identify and account for nonlinearities prior to modeling.

- **Interpretability.** Predictive Analytics models lie on a spectrum of interpretability and complexity (or flexibility). Some PA models more easily support the interpretation of coefficients (e.g., LASSO, Random Forest). Other models (e.g., ensemble trees, convolutional neural networks, stacked models that combine multiple types of PA, and others) can produce predictions and measures of fit related to those predictions, but interpreting coefficients or understanding why the model reaches its outcome is not straightforward.
- **Measure of fit.** A common measure of fit for regressions is the R-squared, which is the percent of total variance explained by the model. There are multiple measures of fit for benchmarking PA models. PA models with continuous outcomes commonly report R-squared or Root Mean Squared Error. PA models with binomial outcomes commonly report the Receiver Operating Characteristic Curve. Another common measure is achieving accuracy, which is the percent of outcomes correctly predicted. Other models may seek to achieve a high degree of precision (correctness), which measures the degree of false positives. Recall (completeness) measures false negatives (GFDRR 2018).

ML techniques have been applied to other, related problems such as crop modeling (Burke and Lobell 2017; You et al. 2017; Peng et al. 2020) and poverty mapping (Jean et al. 2016; Head et al. 2017; Yeh et al. 2020). While not reviewed in detail here, findings from efforts such as these could inform identifying and monitoring populations at risk of food insecurity. To date, we are unaware of research that has integrated these types of models into models predicting food security, although this could be a useful direction. Table 2 summarizes PA models for other outcomes, such as Ebola and crop yields.

D. Model Outcomes and Predictors

The choice of outcomes and predictors, ideally, should reflect the model objectives. Often, however, outcome choices are driven by data availability. Outcomes can be the IPC, food insecurity, or malnutrition, or can be hazards (e.g., drought, floods, forage availability, and infestations). Modeling the outcome will vary by objective. For example, models can predict the likelihood of an outcome (often used for scenarios), a change in the outcome (useful for triggers), or the percent of population in need or at-risk (useful for targeting and scenarios). Modelling objectives are described in more detail in the Modelling Choices section (below).

1. What outcome should be modeled?

1a. IPC

Many researchers use the IPC as their outcome variable because it is available and standardized (e.g., Coughlan de Perez et al. 2019; Baez et al. 2019; Krishnamurthy et al. 2020b; Andree et al. 2020; Westerveld et al. 2021). The modeled IPC outcome can be used for both triggers and scenarios. Some researchers predict the IPC as a binary variable (Krishnamurthy et al. 2020b; Andree et al. 2020), which is useful when the model objective is to trigger early action or financing. Krishnamurthy et al. (2020b) argue that a “tipping point” (i.e., a nonlinear change) occurs when the IPC moves from minimal or mild food insecurity (IPC of 1 or 2) to a food crisis (IPC of 3 or above). Machine learning techniques are well suited to modeling nonlinearities. Krishnamurthy et al. (2020b) find that drought can be predictive of changes in IPC status. However, they warn that because each drought – and the factors correlated with it (e.g., prices) is unique, more research is necessary to determine whether remote sensing data can be used to forecast a tipping point in the IPC (Krishnamurthy et al. 2020b). Andree et al. (2020) predict the IPC as a binary variable, where a value of 0 is given when the IPC is 1 or 2, and a value of 1 is given when the IPC is three or above. Westerveld et al. (2021) estimate IPC transitions in Ethiopia, using open-source data.

Perspectives on the value of estimating the IPC are mixed. Some researchers believe that the IPC is the best metric (018, 070); others consider it good enough or the best available (027, 046, 069). One modeler explained that "it is the only thing out there" that allows for cross-national comparisons and national coverage at a regular frequency (046). Another modeler raised a concern that the IPC has not been validated systematically against food security and nutrition measures (042). While the IPC has been validated as differences between forecasted and actual IPC status (Choularton and Krishnamurthy 2019), such an approach could be endogenous because the IPC current status assessment teams know (and may be influenced by) what the forecast IPC value was (042). Some key informants are concerned that IPC categories miss pockets of need because of its high degree of spatial aggregated and that some small populations within a zone may face much more severe outcomes than the status held by 20 percent of the population (039, 042), echoing points raised by Maxwell et al. (2020). A related concern about using the IPC is that the spatial (and, less so, temporal) level of the IPC is too coarse for targeting and response analysis (039, 042). And IPC as an outcome is categorical, but the five categories are not equivalent, so many modelers dichotomize IPC (less than Phase 3/Phase 3 and above) because dummy outcomes variables are easier to work with in ML context, and because the Phase 2/Phase 3 divide is the most relevant to response.

1b. Food security and nutrition measures

Models using food security and nutrition outcomes (other than the IPC) are generally intended to support scenarios. One model uses food security measures such as rCSI and HFIAS, although the model now-casts rather than forecasts (Lentz et al. 2019); Knippenberg et al. (2019) nowcast resilience measures and rCSI. A challenge with food security outcomes generated from household surveys is that these data either cover a national or near-national area for few time periods (Lentz et al. 2019) or are at a high frequency for a few sentinel sites (Knippenberg et al. 2019).

There are several models under development aiming to predict nutrition outcomes. Baez et al. (2019) apply machine learning techniques to assess drought's impacts on stunting collected by DHS. They find drought is associated with deteriorating food security outcomes and intend for their drought shocks predictions to be applied to data sparse areas without stunting information (Baez et al. 2019). Modelers with works in progress using outcomes such as stunting from DHS data or GAM with SMART surveys (029; 049; 060). One modeler reports delays in accessing SMART data and challenges with SMART surveys covering relatively small areas (060). Checchi et al. (2017) also warn that while SMART-led initiatives have improved data quality at point of collection, there is still a wide range of data cleaning procedures, variation in these procedures lead to different estimated rates of malnutrition. Further, as with Maxwell et al.'s finding that household food security measures are imperfectly correlated, (2014), Checchi et al. warn that some nutrition measures overlap but are not perfect substitutes of one another (2017).

A limitation with using food security or nutrition outcome data is the lack of well-sampled food security and nutrition outcome data that the model can be trained on and validated against (042, 046, 049, 060) (see Annex section: "What is Machine Learning" for more information). With the advent of new data collection techniques, this is changing. Kimetrica (2020) proposes to identify undernutrition by combining three-dimensional photography and convolutional neural networks, a machine learning technique; while not yet usable, it is indicative of the interest in identifying new ways to collect nutrition data. WFP has rolled out SMS and cellular phone based data collection for its Hunger Map (World Food Programme 2020). Questions remain about how to best address non-response bias and representativeness in these innovative techniques (061).

1c. Hazards

Hazards are used as outcomes for both triggers and scenarios. When the objective of the model is to estimate a hazard, caution on inferring its impacts is necessary. Not all hazards cause food insecurity, even if there are strong associations in some locations (Funk et al. 2019). Most modeling of hazards focuses on predicting meteorological hazards such as drought and flooding (Gros et al. 2019; McNally et al. 2019; Arsenault et al. 2020; Getirana et al. 2020). Such models can be helpful in producing food security scenarios, as is currently done within the IPC and FEWS. Such models can also be used to predict triggers for **forecast based financing,** which releases funding prior to a crisis for mitigation, for pre-positioning, or other support. (Gros et al. 2019; Coughlan de Perez et al. 2019; van den Homberg et al. 2020). Impact based financing combines weather forecasts with risk assessments to identify those who are most at risk (IFRC 2020).

Meteorological data, compared to other data sources, provide uniquely rich coverage, both spatially and temporally. For this reason, there has been significant modeler attention on these data. There are different families of meteorological data that have different "products" and often start at different points in time (056); many users rely on the CHIRPS-family of data (Funk et al. 2019), which is used by FEWs and others. One modeler warns that because there is an abundance of remotely sensed data, modelers may choose specific data sources without understanding the limitations and opportunities (056). There is also some evidence of path dependency, with prior decisions guiding current ones. For example, decisions about which meteorological products to use are not revisited even as better products emerge (Brown and Brickley 2012). Forecasts of meteorological outcomes identify only one driver of food insecurity, and how these hazards interacts with prices, for example, cannot be inferred.

There are non-meteorological hazards as well. Matere et al. (2020), for example, develop a model combining a wide range of data, including meteorological, ground-truthed water balance estimates, and a forage growth model to estimate a monthly forage condition index for Kenya. Deteriorations in forage conditions can coincide with and or provide warning of deteriorations in local food security status. Other sorts of outcomes (e.g., deteriorations in WASH status, displacement, and the movement of people) may precede severe food insecurity and may be productive areas for modelers and decisionmakers interested in early warning for food insecurity to focus on (021).

2. Which predictors to include?

Modelers include predictors to predict the outcome of interest. These are "x" variables, independent variables, or right-hand side variables. Models often include a mix of dynamic and static indicators.

2a. Dynamic predictors

Dynamic data often include meteorological data (and related products, such as crop production estimates), prices, and conflict data. Meteorological data, which are often more dynamic than modelers even need are often processed (e.g., daily rainfall is aggregated into monthly or weekly rainfall) before included. The specific processing of meteorological data often draws on findings from agronomy and crop production models (e.g., those produced by Burke and Lobell 2017; You et al. 2017; Peng et al. 2020). Other modelers seek to include prices. A challenge with price data is that they are available at inconsistent spatial and or temporal resolution. Conflict data is also dynamic, although how to best incorporate conflict is not known (066 and Maxwell and Hailey 2020a). Both prices and conflict are discussed further below.

2b. Static predictors

Some predictors are static, such as location, night lights, and roofing material. Within the development space, many researchers apply machine learning to predict poverty or create poverty maps using information such as night lights and roofing materials (Jean et al. 2016). These poverty maps, while often successful at providing a snapshot of current poverty status, are not dynamic. As Head et al. (2017) find, modeling temporally varying outcomes (such as education and health) with relatively static data (such a night lights) is less successful. Lee and Braithwaite (2020) concur, arguing that such an approach is useful for cross-country comparisons but may be less so for intertemporal analysis. One modeling team reported that the most informative predictor of nutritional status for their nowcasting model was geographic location (049). However, static or less-dynamic data will not capture rapid changes in people's situations.

2c. Conflict predictors

A recent review finds that conflict is associated with stunting (Brown et al. 2020), although the role of conflict in food insecurity and how to best measure it is not yet well known (Maxwell and Hailey 2020a). In PA models incorporating conflict, ACLED data is common, although conflict can be defined in different ways. For example, one paper uses conflict counts and fatalities (Andree et al. 2020) while another uses conflict counts (Krishnamurthy et al. 2020a). A few interviewees working on models which include conflict as a driver report finding that conflict is not a statistically meaningful driver of food security or has less predictive power than meteorological data (049; 071; 072).

Others argue that the research and practitioner communities don't have a deep enough understanding about how conflict relates to food security to correctly operationalize existing conflict data in models (044; 060; 061; 066; 070; see also Maxwell and Hailey 2020a). For example, a few modeler KIIs described that rather than using conflict counts or fatality numbers, more-precise conflict measures, such as civilian deaths by non-state actors, armed conflict on major roadways, armed conflict near major markets or at certain times of year, or locations that are categorized as peaceful, having ongoing conflict, and have emergent conflict may better predict food security outcomes (060; 066; 071). These respondents warn that we don't yet have the rules of thumb on how conflict can impact food security (i.e., causal theories) and therefore, the lack of observed relationship between conflict and food insecurity may reflect mismeasurement rather than a lack of a relationship. While some initial work suggests idiosyncratic spatial and temporal relationships between conflict and food insecurity (046) Tandon and Vishwanath (2020) note a strong temporal relationship (but not a spatial link) between conflict and food insecurity in Yemen.

2d. Rare events

Models are built using historical information; our literature review did not identify PA models linking either COVID-19 or desert locusts to food security. One reason is the rare and novel events do not readily lend themselves to modeling. Modelers cannot easily model unanticipated events or novel events without prior data (McBride et al. under review). However, there are some innovations in this space.

First, it is possible to use intermediate outcomes (e.g., food price increases) from other events (e.g., the 2008 food price crisis) to understand what might happen to food security if food prices were to increase due to a novel event (e.g., COVID-19). To use historical information, one needs to identify whether prior events are similar enough to the current situation that they can be used to inform current models. A model that showed the impact of price increases similar to the 2008 food price crises on food insecurity today would not be "correct," but it might help decisionmakers understand possible impacts of rapid prices increases due to COVID-19 related supply chain disruptions. To ensure such models are useful requires modelers and analysts to have an-depth prior understanding of what the pathway of impact is likely to be and what historical data might be similar enough (019; 042). Thus, models using historical data from another event to understand a current novel event may not be technically "correct," but may be "useful" in that they can still provide insights for decisionmakers (019).

Second, the novel and rare events of COVID-19 and desert locusts have shown that some models may be best suited to identifying certain drivers of food insecurity but may not be exhaustive (042). For example, meteorological data can inform a global or regional early warning system for drought and its impact on food security (Funk et al. 2019; Coughlan de Perez et al. 2019). Combining them with parallel systems, such as human analysts tracking conflict, pests, diseases, etc. may provide a path forward (Coughlan de Perez et al. 2019). Further, given that models may miss novel or rare events, some modelers position their models as part of a broader EW system rather than stand-alone or entirely replacing current EW approaches (39; 40; 42; 49; 52; 59; 60). The challenges of predicting rare events also reinforce the perspective that real time monitoring systems operating in parallel (or feeding data into predictive models) are likely better suited to identifying unexpected and emergent shocks.

3. Should the model use only pre-existing data ("secondary data"), or should new, additional data be collected that is tailored to the needs of the model ("primary data")?

Currently, many models rely entirely on secondary data (see Table 1). Many KII report that additional data collection is unrealistic leading them to focus on building models with already existing data. Some secondary data (e.g., high resolution satellite imagery or cell phone meta data) are proprietary and costly (see Table 2). Primary data, when collected, is often used for model validation or calibration. For example, Matere et al. (2020) collect water levels in pans to inform and calibrate its measures of water balance that are included in their forage model. Burke and Lobell (2017) estimate maize yields in western Kenya using satellite data and compare results to crop cuts. Knippenberg et al. (2019) collects weekly food security measures from sentinel sites in Malawi. These sentinel sites provide a way to validate predicted model outcomes against actual food security status. High frequency food security data collection at sentinel sites can provide early warning system information, (Barrett and Headey 2014); however, securing sustained funding for such projects is a challenge (039). The specific data included is often driven by availability (i.e., additional data are not collected just for the modeling effort). As a result, the data available may not be the full set of data needed to model the problem.

E. Modeling Choices

The objectives for the model will influence modeling decisions as well as inform what the models can tell us about food security predictions. Here, we highlight several significant decisions and discuss the implications of these decisions for achieving modeling objectives.

1. What is the desired predictive window (i.e., nowcasting, forecasting)?

1a. Nowcasting

Nowcasting seeks to predict outcomes contemporaneous with the predictors, while forecasts predict outcomes at a point in the future. Nowcasting may be useful when aiming to predict outcomes for data sparse areas (Baez et al. 2019), when seeking to estimate at a spatial scale that differs from the scale of the original data (070), or when seeking to validate the model against existing benchmarks (Knippenberg et al.

2019; Lentz et al. 2019). The challenges with measuring the role of conflict is encouraging some modelers to focus on nowcasting to "get it right" before moving on to forecasting (070). Nowcasting for data sparse areas is not without difficulties. If intense conflict or other hazards limit data collection, models built using data from more accessible locations will not be valid in those data sparse areas. Models may produce more positive outcomes than is warranted. In contrast, if an area was not covered in a survey due to random sampling, applying conclusions from similar, sampled areas may be reasonable. In both cases, knowledge of how different or similar the areas are and why data is lacking is critical before using nowcasting techniques.

1b. Forecasting

Forecasts can be between one month ahead to twelve months ahead (Andree et al. 2020). Selecting the prediction window can be driven by data availability and model performance. Generally, models provide more accurate predictions for shorter predictions (e.g., one month ahead rather than twelve months ahead). Predictive accuracy will fall when models predict far out into the future (Andree et al. 2020). There is a lack of consensus of "how early is early enough." Comparing the costs and benefits of early action that accounts for inaccurate predictions against the costs and benefits of regular, post crisis action could help resolve the tradeoffs between greater accuracy and longer predictive windows.

1c. Predictive window for early action

Ideally, the timeframes required for different EA responses should inform the prediction window. That is, if a goal for the EW model is to provide enough EW to deploy contextually relevant responses, that goal could inform the choice of predictive windows (027). **Longer windows will likely result in lower accuracy predictions**; however, this tradeoff may still be worthwhile if EA responses are lower cost than other interventions requiring less warning. For example, if the EW system provides a warning two months in advance but a fodder project requires 6 months to get started, by the time the warning occurs, it will be too late to start a fodder project. Yet, if they are the most cost effective / most adept at addressing crises, it may be better to use a lower accuracy model with a six-month EW window than a model with high precision and a 3 month window.

2. Is the model intended for scenarios or for triggers? Are there appropriate thresholds or triggers for early action?

2a. Triggers

Triggers tie early action funding to specific events. Recent work by the World Bank and OCHA have investigated changes to the IPC as a possible trigger (e.g., moving from below a 3 to 3 or above) (Andree et al. 2020). **Forecast based financing** for droughts (Coughlan de Perez et al. 2019) and floods (Gros et al. 2019) releases funding when hazard-based triggers are predicted. The funding is intended to support the ability of at-risk populations to manage the impending hazard. Such triggers are intended to be relevant to the lives and livelihoods of at-risk populations. Contexts will shape the likely hazards and therefore the appropriate triggers (069). Triggers linked to "softer" outcomes (e.g., food insecurity) compared to sudden onset outcomes (e.g., flooding) require more discussion and collaboration among modelers and end-users in order to reach agreement on how to set the trigger and what to do if the crisis is complex or if there are consecutive crises (069; 072).

Identifying the appropriate trigger can be challenging; research is ongoing (046; 069; 072). Further, even if a trigger is identified, one respondent argues for stress testing model results (027). Another agrees that trigger-based models can introduce some uncertainty but acting on a no-regrets basis and computing returns on investment can clarify when the uncertainty regarding action is outweighed by the possible "returns" on action (019). In other words, if the program implemented is valuable to the at-risk population, regardless of whether the predicted crisis occurs, then implementers will not regret having intervened (Lowcock 2018). A lower-risk form of triggers could be sequential triggers (069), where initial funding is released on very early EW information, with further funding released as the window between the forecast and the event closes. Currently, trigger-based models are used mostly for single-hazard/single-response actions; whether PA can generate reliable and valid triggers for multiple hazards remains to be seen (see IFRC 2020). A few modeler KII (069; 072) propose triangulation across different types of models (e.g., when triggering funds in response to a drought, an analyst could examine changes in the IPC plus specific localized drought indicators).

2b. Scenarios

Scenarios depend on concerted human judgment. PA models intended to complement or inform early warning scenarios provide another source of information to triangulate against (Lentz et al. 2019). For example, an EW analyst may develop a scenario and could compare it against predictions from a model. Or a model could feed into the scenario development process. As Maxwell and Hailey (2020a) note, scenarios are based on a set of assumptions about the future, and they warn that the assumptions are frequently not monitored or tested on a regular basis. Models similarly make assumptions about the future. These can be made explicit and tested (e.g., price increases of 10 percent can be compared against prices increases of 20 percent). Even if a model is intended to inform scenarios and aims to capture multi-hazard, multiple actions contexts, no model is entirely comprehensive and, for that reason, scenario-based models are thought to be complements to EW systems rather than replacements.

3. What is the desired spatial resolution?

The scale of available data influences the usefulness of the model for decisionmakers. Some donors and decisionmakers may prefer to start with larger spatial units before drilling down (070). For example, WFP's Hunger Map (World Food Programme 2020) is at a country level, which may be quite useful for international actors and donors. Country-level reporting is less useful for decisionmakers seeking to identify areas of need within country. Modelers are pursuing a variety of spatial units of analysis, including at village-cluster (Lentz et al. 2019), at the province level (Baez et al. 2019), at the IPC zone (Andree et al. 2020) and at country-level (World Food Programme 2020). FRAYM (2020) is estimating the IPC at a lower geographic scale (e.g., at 1 kilometer scale). Matching data spatially (and temporally) is time consuming and can be challenging, particularly when there is little guidance on best practices (Lentz et al. 2019). Within and across countries, there are richer data pockets (spatially and or temporally) in some locations relative to other locations, making generating models with forecasts spatially disaggregated at the Admin-2 level or lower difficult (071). The choice of spatial scale will depend on how the model will be used and data available. Some end users may wish for greater granularity to support targeting. Others may prefer district level or IPC zone level models to trigger funding.

4. Should PA models be highly specific to certain locations or livelihoods or be consistent across a country or region?

While data availability often drives model development, there is also a question of whether the model should be generalizable across countries or specific to the local context. The former offers comparability. The latter may ensure accuracy for an otherwise hard to monitor factor. The WFP Global Hunger Map provides primarily country-level information for a large number of countries, informing donors, UN agencies, and the media. Andree et al. (2020) develop a model that estimates changes in the IPC for 21 countries, which includes predictors at the Admin 2 level (although Admin 2 levels are not necessarily comparable across countries). A few modeling groups seek to develop general models that allows for apples-to-apples comparisons and possible prioritization for assistance, (042; 049; 071), but also recognize that not all indicators make sense in all contexts.

Local models are more helpful to identify appropriate interventions and causal chains (071). Such models are generally tailored to specific contexts or use highly specialized data sources (027; 042; 069). For example, models developed for Somalia may not be "transportable" because the data in Somalia are quite rich and unique (027; 042). An argument for context specificity is that the main shocks often differ across contexts, and therefore, some of the indicators that are regularly tracked and available, also differ (069). The Pastoral Livestock Early Warning System monthly forage condition index has been tailored to a specific livelihood (pastoralism) in Kenya (Matere et al. 2020). The authors note that to extend it to other pastoral areas in East Africa would require a deeper understanding of other agro-ecological zones. In trigger-based models, some triggers may vary to reflect differing levels of severity across different countries (e.g., the resiliency of different populations to recover from a drought varies across and within countries) (069). A middle path is the Water, Peace and Security project, which offers both global models and regional models to "zoom in" on hotspots that may have more context specific drivers.

A combined approach that generates comparable models across countries (either at country-level or more localized to Admin 2 areas) with more locally specific models (either for country coverage or for specific hot-spots) could provide a useful balance.

Our review did not identify peer-reviewed models that explicitly incorporate cross border information for EW or that seek to ensure results across borders "make sense." This is likely because data are unevenly available and matching these data across different spatial levels and requires care (Matere et al. 2020). However, thinking about countries embedded within regions is important area for future PA work (e.g., if prices are rising in Kenya, what might that mean for its neighbors?).

5. Do decisionmakers prioritize maximizing accuracy, minimizing false negatives, or minimizing false positives?

Several models are evaluated on accuracy, which is the percent of the time the prediction is correct (Choularton and Krishnamurthy 2019; Knippenberg et al. 2019; Lentz et al. 2019; Andree et al. 2020; Matere et al. 2020). Modelers predicting dichotomous outcomes can also choose modeling techniques that seek to minimize false negatives or false positives. False negatives (errors of exclusion) wrongly identify food insecure households or regions as food secure. False positives (errors of inclusion) wrongly identify food secure households or regions as food secure. Modeling to minimize errors can include using innovative sampling techniques (e.g., ADASYN and SMOTE) (Zhou et al. forthcoming) or by weighing prediction errors in favor of minimizing either false positives or false negatives (Andree et al. 2020). An emerging area of research, such models may report measures of precision (true positives divided by the sum of true positives and false positives, which is useful for capturing false positive rate) and recall (true positives divided by true positives and false negatives, which is useful for capturing false negatives).

Andree et al. (2020) point out that whether to minimize false positives or negatives depends on the policy appetite: would a policymaker or donor rather react early to a crisis that turns out to be not as bad as predicted (i.e., a false positive)? Or would they rather react retroactively when the model misses a crisis (i.e., a false negative)? Others (Zhou et al. forthcoming, 042, 052, 071) raise similar questions. Some respondents report that donor and decisionmaker appetites for being wrong vary – one perception is that donors and decisionmakers would rather not be caught spending on crises that do not emerge (i.e., responding to a false positive) (070). One interviewee suggested that given that the costs associated with early action are often significantly lower than the costs associated with regular or late responses, it is better to treat all positives seriously even if some are false positives (019). This decision may also turn on what (if any) "no-regrets" interventions are available.

6. Should modeling techniques that allow for easy interpretation of the predictors be chosen or should the model be chosen to maximize fit?

Some modeling techniques allow for easy interpretation of the coefficients on variables on the right-hand side (which helps with causal analysis) (Zhou et al. forthcoming). Other models maximize fit (which helps improve accuracy) but may not support easy interpretation of coefficients (Paul et al. 2019). Models that do not allow for easy interpretability of predictors are sometimes described as "black box" models and often require additional analyses to generate interpretable findings. Ultimately, whether to prioritize interpretability will depend on the objective for the model. The ability to interpret coefficients may matter less for trigger models, when the goal is to know what is happening rather than why something is happening.

Some techniques, such as models that are extensions of regression (e.g., random forests or Lassos), support relatively easy interpretation of predictors (056, 059, 060) (Knippenberg et al. 2019; Lentz et al. 2019; Andree et al. 2020). In several reviewed studies (e.g., Knippenberg et al. 2019; Lentz et al. 2019; Andree et al. 2020; Zhou et al. forthcoming), modelers select variables for inclusion because they have a rationale, a theory, or heuristic as to why those variables make sense (assuming such data are available). Being able to interpret the predictors enables modelers to assess whether these variables influence the outcome as expected (e.g., poorer households are associated with worse food security outcomes), which is a check on a model's internal validity. Interpretability also helps decision-makers understand the nature of problem and how to respond, and it is desired by some programming staff (027; 059). Further, including dynamic variables that can be influenced by policies and programming can be helpful for end-users. One interviewee reported that a model found that spatial fixed effects and weather-related measures are highly predictive of malnutrition (059). However, these variables provided little insight into which interventions could help, a finding that frustrated program managers and designers. Decisionmakers may not want complex black-box models that are highly predictive but hard to understand; rather, models "must be believable" (059; Coyle and Weller 2020).

Other modeling techniques are "unsupervised"; modelers impose very little structure and the models, such as convolutional neural networks and stacked models, utilize large amounts of data. These techniques can "let the data speak" and identify patterns that modelers would not necessarily know to look for (033). A convolutional neural network model that prioritizes prediction may use a huge amount of different data sources and reliably predict the IPC. This may be adequate to release or trigger funding, although it will not allow an analyst to determine *why* the IPC is changing or to identify which factors (e.g., drought, prices, conflict) are the primary drivers of the deteriorating situation. Thus, black-box models may generate more accurate predictions than other modeling techniques, but they do so at the cost of easy interpretability (Andree et al. 2020; Yeh et al. 2020). Ensemble models, which average across predictions, may provide more confidence in black box models (Varian 2014).

7. What are the future directions for data and models?

7a. Data quality

Deparday et al. (2019) argues that ML algorithms are only as good as the data it is trained on. Checchi et al. (2017) also warn about data quality. They find that most crises do not meet minimum public health data requirements, or the data are not made available in a timely fashion and or are scattered and difficult to access. Without data, models won't work. One perspective on the impact of low-quality data is, "garbage in, garbage out" (042).

7b. Data availability and timeliness impact the usability of a model

Numerous modelers described "data latency" as a major source of frustration. Latent data are slow to arrive or aren't readily shared (019; 029, 033, 049, 060, 066, 069). Several modeler KIIs believe that additional sustainably funded data collection is not realistic, and therefore they need to make do with the data available (027; 033; 049; 070). However, other modelers believe access to better price data could be transformative. Currently, price data is not (reliably) collected, is missing, slow to arrive, or not accessible (027, 029, 042, 044, 046). One respondent (029) indicated that while some institutions champion certain data collection efforts (e.g., UNICEF for nutrition; USAID for DHS; WB for LSMS), currently there is not a champion for price data, noting this could be a space for FAO. In response to the lack of reliable price data, some modelers are studying using innovative data collection, such as SMSing traders. One modeler is investigating the usability of mobile money purchasing patterns, and natural language techniques to parse indigenous radio call-in shows for price-related early warning information (033).

7c. Ensemble models

Varian (2014) points out that average predictions of many models (known as "ensemble modeling" often outperform any single model. Ensemble modeling is commonly used for meteorological hazards (e.g., Arsenault et al. 2020), but less for food security outcomes. One KII (049) reports using joint ensemble prediction improves model fit compared to sequential forecasting. However, one group (072) is unsure whether an ensemble approach would be acceptable to stakeholders because interpreting the role of drivers across multiple models would be difficult.

7d. Emerging data and models

Data availability and modeling techniques are changing quickly. More data are collected via social media and cell phones, and nutrition data collected via camera and processed via machine learning may be emergent (Kimetrica 2020). Several modeler KIIs mentioned an interest in including variables related to WASH, wages, remittances, distress migration, and livestock death and disease (007, 012; 021, 029, 042). A challenge is some innovative forms of data are not publicly available (i.e., cell phone meta data) or are only available for purchase (i.e., commercial satellite data at high resolution) (Blumenstock et al. 2015; Burke and Lobell 2017). Other AI-based or AI-assisted technologies may be transformative, such as natural language processing tools for household surveys, microwave data from cell tower to inform meteorological forecasts, and swarm-based artificial intelligence to help decisionmakers and analysts reach consensus (020, 029, 033, 049).

7e. Data availability and data access often drive approaches

Rather than modeling what is needed, researchers are modeling what is possible. A question is: is this pragmatic or is it hindering development of the type of EW systems that could provide better information for better early action? While relying on available data as a departure point for modeling makes sense from a modeling perspective, it may result in models that are less helpful for EW end users, who may be looking to

PA to provide information on specific problems. This speaks to a need for more collaboration between researchers and EW end users on what the relevant problem is, and what data should and can be brought to bear on the problem. This also speaks to a need for frank assessments on what the data gaps and priorities are and whether additional data, if any, can be collected (with what funding, processed by whom, accessible by whom, etc.).

F. Findings

Modelers and End Users Face a Series of Choices. Applications of PA to food security and food-security related problems are developing quickly. As of yet, there are no established "best practices." What can be done with PA models depends on several factors, including model objectives, data availability, capacity, and ethics, among others. For EW system users, it will be critical to engage modelers, decisionmakers, analysts, and donors interested in using PA in collaborative discussions not only about what is possible and feasible but also to identify assumptions within models that may otherwise remain implicit.

"Looking for Keys under the Street Light" Syndrome. The easy availability of certain data can overly focus attention on these data at the risk of missing other important factors, causing a "**streetlight effect"** (i.e., models are built using only the data that can be easily seen and accessed). A model that starts with (causal) factors and builds out a predictive model may require additional targeted data collection to address shortfalls in coverage of specific causal factors. Price data, for example, is limited and primarily collected from large markets, which may not be accurate proxies for prices in rural areas.

Novel and Rare Events Are Currently Challenging for PA Models to Predict. PA initiatives may be most helpful in identifying and monitoring the "usual" drivers of food insecurity. Human analysts will likely remain essential in monitoring and addressing less easily identifiable drivers and their impacts. Further, PA can be used for **nowcasting and forecasting.** Nowcasting models predict results for areas without data, but places without data may systematically differ from places with data. For example, if conflict limits data collection from some areas but not others, models built using information from conflict-free areas will generate predictions that are inaccurate for conflict locations. If intending to nowcast outcomes for data-sparse areas, analysts and decisionmakers must have an understanding of *why* some areas are data sparse and whether the model results are truly transportable from data-rich to data-sparse locations. At the same time, forecasting accuracy decreases as the forecast window increases. Decision makers will face tradeoffs regarding accuracy and duration of early warning. On average, some early action interventions that require longer early warning windows may be cost effective even if sometimes the model predicts false positives.

Scenario-Driven Models Are Often Complementary to Existing Systems. Even if a model is intended to inform scenarios and capture multi-hazard multiple-actions contexts, no model is entirely comprehensive and, for that reason, scenario-based models are thought to be complements to EW systems rather than replacements. How to interpret PA results and incorporate them into established scenario building has not yet received adequate attention.

Trigger-Based Models Are Increasingly Common. Currently, trigger-based models are used mostly for single-hazard/single-response actions, such as forecast-based financing. Research on PA generating reliable and valid triggers for multiple hazards is needed; work on impact-based forecasting seeks to combine risk assessments with weather forecasts (IFRC, 2020).

Ethics, Bias, Privacy, and Equity Concerns. All data collection and information systems run the risk of failing to capture relevant indicators and accurately interpret them. This is, perhaps, an even greater concern when using PA techniques (O'Neil 2016). It can be difficult for non-modelers to understand the modeling and data choices (Coyle and Weller 2020). At the same time, modelers may not have adequate understanding of implicit biases or gaps in data, or of historical and contemporary contexts that could shape the choice of analyses and interpretation (D'Ignazio and Klein 2020). Below are several ethical concerns. Some are applicable generally to EW systems and some are of particular concern when using PA models.

- **Inaccuracies, Biases, and Inequalities in Historical Data.** If historical data used in predictive analytics have inaccuracies or biases, they can replicate and amplify inequalities, including racism and sexism, and reinforce class or other sources of privilege (O'Neil 2016). Further, using unsuitable proxies may result in inaccurate or confusing predictions (Paul et al. 2019). This issue is not unique to machine-learning tools for predictive analytics. Human-based analytical processes can suffer from similar problems.

- **"Baked in" Inequalities.** When big data incorporates inequalities, it may introduce biases. For example, better-off populations are likelier to have access to cell phones, analyses using phone-based data need to account for response variation by cell phone ownership. Similarly, within resource-constrained households, women's cell phones are less likely to be charged than men's, and they are less likely to own a phone (Paul et al. 2019). This matters for early warning because men and women have different perspectives on food security (Coates 2013). Data drawn from social media face similar critiques. Social media users may be more well-off than others in the population. Relying on big data without recognizing whose voices are being captured (and at what rates) may result in misrepresenting perspectives of individuals without regular access to social media, cell phones, etc. (Hernandez and Roberts 2020).

- **Concerns about Data Privacy.** These are increasing. To the best of our knowledge, EW models are not (currently) using highly disaggregated data, such as street-level-view or drone imagery data. As these data sources become more common, privacy concerns also increase. Different countries are taking different approaches to ensuring privacy of their citizens (Paul et al. 2019). The European Union's General Data Protection Regulation, for example, limits data sharing and selling to a much greater degree than current US standards.

- **Lack of Explainability and Excessive Trust.** Paul et al. (2019) warn of "excessive trust" in ML, which they define as "unquestioning acceptance of model results, which can result in misinformed choices when models do get it wrong" (p. 40). Developing models that prioritize understandability and interpretability can help to minimize these risks (D'Ignazio and Klein 2020). Coyle and Weller (2020) argue that "demanding that ML systems be explainable is likely to make the tradeoffs between different objectives far more explicit than has been the norm previously" (p. 1434). This suggestion is in response to growing concern as models are used to automate decisions.

- **Limited External Validation.** Further, the lack of a clear evaluation of models in humanitarian space introduces both technical and ethical risks. A peer review process ought to prioritize the question of "if models fail, who is impacted and how?" (069). Van den Homberg et al. (2020) report, "Only a bare minimum of relevant accountability standards are currently in place" (p. 2). The Centre for Humanitarian Data (2019) offers a peer review process that could support a consistent reporting-and-accountability mechanism for PA models for humanitarian purposes.

- **Lack of Accountability to Affected Populations.** There remains significant opportunity to better incorporate accountability to affected populations during development, validation, and reporting back to affected populations in PA. This is especially true for marginalized populations.

- **Limited Local Engagement.** Without local-level engagement, there is a risk of techno-colonialism, where data are extracted without benefit to local partners (072). A further benefit of prioritizing local-level engagement is that modelers may have a better understanding of community needs and context-specific data (see ACAPS 2020, 033).

Lack of Gender Disaggregated Data and Blindness to Social Inequalities. Relying on big data can amplify inequalities, and currently few EW diagnostics or PA models explicitly disaggregate data by gender or recognize that food insecurity and famine may be experienced differently—and coped with differently—by gender or other social groupings (de Waal 2018).

Table 1: Selected predictive analytics models that focus on food security and related hazards

Authors	Location	Spatial scale	Modeled outcome	Scenario or Trigger	Inputs	Data sources	Frequency of data collection	Primary method(s)	Predictive window	Limitations or extensions
FOOD SECURITY and NUTRITION MEASURES										
Knippenberg et al. (2019)	Malawi	Household level predictions within several villages	Resiliency; rCSI	Scenario	Assets, amount of land farmed, livestock (Tropical Livestock Unit), living within or outside of flood plain, having fields far from home, demographics.	Primary	Monthly	LASSO and Random Forest (and other regression techniques)	Forecast 1-2 months and use ML to identify future variables	Requires sentinel site data collection.
Lentz et al. (2019)	Malawi	IPC zone, Admin 3 (traditional authority) and Admin 4 (village)	rCSI; HDDS; FCS	Scenario	Previous month of IPC classification, market prices, remotely sensed data, household roof type, cellphone ownership, and household characteristics	Secondary	Mix	LASSO (and other regression techniques)	Nowcast	Training data was collected 3 years prior to the evaluation period; the fitted coefficients may no longer be as accurate.
Baez et al. (2019)	Malawi, Tanzania, Mozambique, Zambia, Zimbabwe	Admin 1 (province)	Children at risk of stunting and wasting	Trigger	Drought measures, NDVI; DHS household measures	Secondary	Mix	Random forest and gradient boosting	Nowcasting drought contingent targeting	Non-monetary measures such as stunting can have significant

										measurement error
Fraym (2020)	Nigeria, Pakistan	1 km squared	Localized Food Insecurity Index	Scenario	Index created from IFPRI's Global Hunger Scale; predictors include IPC, household surveys, satellite imagery, mobility data from network operators, food prices	Secondary	Mix	Human centric QA/QC automation	Nowcasting to lower geographic scale	Doesn't appear to validate out of sample and cannot model areas with < 30 people for 1km-squared
IPC										
Coughlan de Perez et al. (2019)	Ethiopia, Kenya, Somalia	IPC zone	IPC classifications	Scenario and Trigger	12-month Standardized Precipitations Index (SPI) and Standardized Precipitation Evapotranspiration Index	Secondary	High frequency	Chi square tests of SPI index and IPC values (among others)	Forecasts of up to 6 months	Estimation period is 2011 to 2018; climatic events outside of that window would be missed.
Choularton and Krishnamurthy (2019)	Ethiopia	Admin 3 (woreda)	IPC classifications and transitions	Scenario	Actual and forecasted IPC values; results compared against remotely sensed data	Secondary	Quarterly, and then tri-yearly	Compare IPC forecasts against observed IPC current status	Nowcast	IPC forecast accuracy varies within country and by time period.
Krishnamurthy et al. (2020b)	Theoretical	IPC zone	IPC transitions	Scenario and Trigger	Remotely sensed environmental indicators that could be related to droughts or impacts of droughts	Secondary	High frequency	Evaluate autocorrelation, skewness, variance, and identification of a threshold	Between year changes in IPC classifications	Theoretical; may not be able to use remote sensing to forecast a tipping point

										because each drought is different.
Andree et al. (2020)	21 countries	Admin 2	IPC transitions	Scenario and Trigger	Monthly covariates including (imputed) price data, conflict events, population, remotely sensed data	Secondary	Monthly	Random Forest, among others	Forecasts of up to 12 months	Importance of specific covariates varies by country. Attends to false positive and false negatives.
Westerveld et al. (2021)	Ethiopia	Livelihood zone	IPC transitions	Scenario and Trigger	Monthly covariates including climate, land, market, conflict, infrastructure, demographics, and livelihood characteristics	Secondary	Mix	Gradient Boosting, Random Forest, among others	Forecasts of up to 7 months	Relies on open data and could be extended to other locations. Some data are imputed.
HAZARDS										
Dreschler and Soer (2016)	Ethiopia (Theoretical)	Admin 3 (woreda)	Drought	Scenario	Ethiopia has a range of EWS to recognize prospective impact of drought: LIAS, LEAP, hotspot assessments, possibility for bottom-up assessments, prices, weather data, IPC	Mix	Mix	Evaluates data for triggering early warning for drought response through Ethiopia's PSNP	Forecasts of up to 4-5 months	Can apply model results to PSNP to estimate beneficiary numbers.

Funk et al. (2019)	FEWS Net countries	Subnational	Drought	Scenario and Trigger	DEWS data includes agroclimatology, climate outlooks, weather hazards, hydrologic monitoring, and vegetation monitoring	Secondary	Monthly	Describes strategies and components of FEWS' Drought Early Warning System	Forecasts of up to 8 months	More work to be done on integrating new remote sensing resources.
Gros et al. (2019)	Bangladesh	Admin 4 (community)	Poverty and wellbeing	Trigger	Wellbeing measures improved households' access to foods, reduction in debt accrual, and psychological stress. Compared outcomes based on receipt of forecast based financing support or not.	Primary and secondary	Pre flood survey in 2016, Post flood survey in 2017	Impact evaluation of forecast based financing		Variation in beneficiary vulnerability and additional flooding makes establishing impact difficult.
McNally et al. (2019)	Africa	Gridded streamflow per capita	Water scarcity index	Scenario	Water scarcity index aims to include societal demand for water. Meteorological data, streamflow, water stress anomalies, population figures	Secondary	High frequency	Meteorological and hydrological modeling; Mapping	Current status maps, updated monthly	Limitations related to uncertainty in hydrologic or land surface modeling.
Arsenault et al. (2020)	Africa and Middle East	IPC zone	Drought	Scenario	Generate drought and flood prediction maps using meteorological	Primary and secondary	High frequency	Meteorological and hydrological modeling; Mapping	Forecasts up to five months	Plan to include additional models and data sources.

					inputs and satellite inputs					
Getirana et al. (2020)	Niger, Chad: Volta	River basins	Flood prediction	Scenario	Predict streamflow to improve flood prediction: Gravity Recovery and Climate Experiment satellites (GRACE) provide measures of terrestrial water storage variability, and other geographic characteristics.	Primary and secondary	High frequency	Meteorological and hydrological modeling; Mapping	Seasonal forecasts 3-6 months in advance of wet season	Limitation: lacking dynamics between groundwater and standing water.
Kuzma et al. (2020)	Global		Localized conflict	Scenario	Community, economy, and governance, indicators, food price data (WFP), prior conflict (ACLED), and water-related factors potentially correlated with conflict (e.g., drought)	Secondary	Mix	Random forest	Forecasts up to 12 months, updated every 3 months	Signal is noisy with half of predictions being false positives.
Matere et al. (2020)	Kenya	Admin 3 and Admin 4	Forage Condition Index	Scenario	Water balance model, Phytomass Growth model simulation, and spatial analysis	Primary and secondary	Mix	ML to train PHYGROW model to differentiate vegetation	Forecasts of 6 months	Assesses feasibility for integrating PLEWS into drought risk management bulletins.

Shukla et al. (2020)	SADC countries	0.25 degree x 0.25 degree spatial resolution	Drought (Root zone soil moisture)	Scenario	Forecasts root zone soil moisture (RZSM), a crop-yield variable. Predictors include land surface data toolkit and meteorological forecasts	Secondary	High frequency	ARIMA model to generate out of sample crop yield forecasts	Forecasts of 4-5 months ahead of harvest and 12 months ahead of lean season	ENSO is one of the main predictors for SADC's climate.
van den Homberg et al. (2020)	Philippines	Admin 2 (Municipalities)	Typhoons	Trigger	27 historical typhoons, 40 predictors including exposure, vulnerability of housing and individuals, housing damage and meteorological data.	Secondary	Mix	AI classification algorithm to trigger anticipatory action	Forecasts of up to 1 month	Appropriate triggers should be developed with stakeholders.

Note: This table focuses on approaches that are used for EW for food security or could be extended to that problem. Other innovations for humanitarian systems more broadly include WFP's (nowcasted) Hunger Map Live and numerous Forecast based Financing and Impact based Financing projects (Future of Forecasts 2020), which tend to consider the (impacts of) natural hazards. Early-stage research that could also be useful include the IPC's ATARI project, which uses AI for consensus-based decision making; the World Bank FAM project; and CGIAR CIAT's Nutrition Early Warning System, and CERF Anticipatory Action pilots.

Table 2: Selected predictive analytics models predicting food-security relevant outcomes

				Role in system		Timing		Data		
Authors	**Location**	**Spatial scale**	**Modeled outcome**	**Scenario**	**Trigger**	**Fore casting**	**Now casting**	**Secondary**	**Primary**	**Proprietary***
Milivonich et al. (2015)	Sierra Leone, Liberia, Guinea	Country	Ebola	x			x	x		
Blumenstock et al. (2015)	Rwanda	Admin 1 (province)	Poverty and wealth	x			x	x	x	x

Jean et al. (2016)	Nigeria, Tanzania, Uganda, Malawi, Rwanda	Admin 4 (village and ward)	DHS wealth index and others	x			x	x		
Burke and Lobell (2017)	Kenya	1-m imagery	Crop yields	x		x		x	x	x
Head et al. (2017)	Haiti, Nepal, Nigeria, and Rwanda	Sub-regional	DHS wealth index, WASH, and others	x		x	x	x		
Steele et al. (2017)	Bangladesh	Cell tower coverage area	DHS wealth index and others	x			x	x		x
You et al. (2017)	US	Admin 2 (county)	Crop yields	x		x		x		
Lee and Braithewaite (2020)	25 countries in SSA	Admin 4 (village)	DHS wealth index	x			x	x	x	
Peng et al. (2020)	United States	Admin 2 (county)	Crop yields	x		x		x		
Yeh et al. (2020)	23 countries in Africa	Admin 2 (district)	DHS and LSMS wealth index	x			x	x		
Note: *The column "proprietary" indicates data, such as call data records from mobile telephone companies that may be available for purchase. To the best of our knowledge, to date such data have not been used in models predicting nutrition or food security outcomes.										

References

ACAPS. 2020. "South Sudan Analysis Ecosystem." Geneva: ACAPS

Andree, B. P. J., Chamorro, A., Kraay, A., Spencer, P., & Wang, D. (2020). *Predicting Food Crises* (No. 9412; Policy Research Working Paper). The World Bank. https://doi.org/10.1596/1813-9450-9412

Arsenault, K. R., Shukla, S., Hazra, A., Getirana, A., McNally, A., Kumar, S. V., Koster, R. D., Peters-Lidard, C. D., Zaitchik, B. F., Badr, H., Jung, H. C., Narapusetty, B., Navari, M., Wang, S., Mocko, D. M., Funk, C., Harrison, L., Husak, G. J., Adoum, A., … Verdin, J. P. (2020). The NASA Hydrological Forecast System for Food and Water Security Applications. *Bulletin of the American Meteorological Society*, *101*(7), E1007–E1025. https://doi.org/10.1175/BAMS-D-18-0264.1

Baez, Javier E., Varun Kshirsagar, and Emmanuel Skoufias. (2019). *Adaptive Safety Nets for Rural Africa: Drought-Sensitive Targeting with Sparse Data*. World Bank, Washington, DC.

Barrett, Christopher B, and Derek Headey. 2014. "A Proposal for Measuring Resilience in Risky World." *Resilience For Food and Nutrition Security*, no. May: 187–94.

Blumestock, J., Cadamuro, G., & On, R. (2015). Predicting poverty and wealth from mobile phone metadata. *Science*, *350*(6264), 1073–1076. https://doi.org/10.1126/science.aac4420

Brown, M, and Brickley, E. (2012). "Evaluating the use of remote sensing data in the US Agency for International Development of Famine Early Warning Systems Network" *Journal of Applied Remote Sensing.*

Brown, M. E., Backer, D., Billing, T., White, P., Grace, K., Doocy, S., & Huth, P. (2020). Empirical studies of factors associated with child malnutrition: Highlighting the evidence about climate and conflict shocks. *Food Security*. https://doi.org/10.1007/s12571-020-01041-y

Burke, M., & Lobell, D. B. (2017). Satellite-based assessment of yield variation and its determinants in smallholder African systems. *Proceedings of the National Academy of Sciences*, *114*(9), 2189–2194. https://doi.org/10.1073/pnas.1616919114

Cabot Venton, C. (2020). *The Economics of Resilience to Drought*. USAID. https://www.usaid.gov/resilience/economics-resilience-drought

Centre for Humanitarian Data. (2019). *A peer review framework for predictive analytics in humanitarian response. Draft for Consultation as of September 2019*. Centre for Humanitarian Data.

Checchi, F., Warsame, A., Treacy-Wong, V., Polonsky, J., Ommeren, M. van, & Prudhon, C. (2017). Public health information in crisis-affected populations: A review of methods and their use for advocacy and action. *The Lancet*, *390*(10109), 2297–2313. https://doi.org/10.1016/S0140-6736(17)30702-X

Choularton, R. J., & Krishnamurthy, P. K. (2019). How accurate is food security early warning? Evaluation of FEWS NET accuracy in Ethiopia. *Food Security*, *11*(2), 333–344. https://doi.org/10.1007/s12571-019-00909-y

Coates, Jennifer. (2013). "Build It Back Better: Deconstructing Food Security for Improved Measurement and Action." *Global Food Security* 2 (3), 188–94. *DOI.org (Crossref)*, doi:10.1016/j.gfs.2013.05.002.

Coughlan de Perez, E., van Aalst, M., Choularton, R., van den Hurk, B., Mason, S., Nissan, H., & Schwager, S. (2019). From rain to famine: Assessing the utility of rainfall observations and seasonal forecasts to anticipate food insecurity in East Africa. *Food Security*, *11*(1), 57–68. https://doi.org/10.1007/s12571-018-00885-9

Coyle, D., & Weller, A. (2020). "Explaining" machine learning reveals policy challenges. *Science*, *368*(6498), 1433–1434. https://doi.org/10.1126/science.aba9647

de Waal, A. (2018). *Mass Starvation: The History and Future of Famine*. Polity Press.

Deparday, V., Gevaert, C. M., Molinario, G. M., Soden, R. J., & Balog-Way, S. A. B. (2019). *Machine Learning for Disaster Risk Management* (No. 133787). World Bank Group. https://documents.worldbank.org/en/publication/documents-reports/documentdetail

D'Ignazio, C., & Klein, L. (2020). *Data Feminism*. MIT Press. https://mitpress.mit.edu/books/data-feminism

Drechsler, M., & Soer, W. (2016). *Early Warning, Early Action: The Use of Predictive Tools in Drought Response through Ethiopia's Productive Safety Net Programme* (Policy Research Working Paper). The World Bank. https://doi.org/10.1596/1813-9450-7716

Funk, Chris, Shraddhanand Shukla, Wassila Mamadou Thiaw, James Rowland, Andrew Hoell, Amy McNally, Gregory Husak, et al. (2019). "Recognizing the Famine Early Warning Systems Network: Over 30 Years of Drought Early Warning Science Advances and Partnerships Promoting Global Food Security." *Bulletin of the American Meteorological Society* 100 (6): 1011–27.

Fraym. (2020). *Localized Food Insecurity Index: White Paper*. https://fraym.io/lfll-wp/

Getirana, A., Jung, H. C., Arsenault, K., Shukla, S., Kumar, S., Peters-Lidard, C., Maigari, I., & Mamane, B. (2020). Satellite Gravimetry Improves Seasonal Streamflow Forecast Initialization in Africa. *Water Resources Research*, *56*(2), 1–15. https://doi.org/10.1029/2019WR026259

GFDRR. (2018). Machine Learning for Disaster Risk Management. Washington, DC: GFDRR. License: Creative Commons Attribution CC BY 3.0.

Gros, C., Bailey, M., Schwager, S., Hassan, A., Zingg, R., Uddin, M. M., Shahjahan, M., Islam, H., Lux, S., Jaime, C., & Coughlan de Perez, E. (2019). Household-level effects of providing forecast-based cash in anticipation of extreme weather events: Quasi-experimental evidence from humanitarian interventions in the 2017 floods in Bangladesh. *International Journal of Disaster Risk Reduction*, *41*. https://doi.org/10.1016/j.ijdrr.2019.101275

Head, A., Manguin, M. Tran, N. and Blumenstock, J. E. (2017). "Can human development be measured with satellite imagery?" Conference Proceedings: ICTD '17, November 16–19, 2017, Lahore, Pakistan. https://doi.org/10.1145/3136560.3136576

Hernandez, K., & Roberts, T. (2020). *Hernandez and Roberts Predictive analytics in humanitarian action: A preliminary mapping and analysis*. Institute of Development Studies. https://opendocs.ids.ac.uk/opendocs/handle/20.500.12413/15455

IFRC. (2020). "The Future of Forecasts: Impact-Based Forecasting for Early Action." *International Federation of Red Cross and Red Crescent Societies*. January 2020. https://media.ifrc.org/ifrc/document/future-forecasts-impact-based-forecasting-early-action/.

Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., & Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty. *Science*, *353*(6301), 790–794. https://doi.org/10.1126/science.aaf7894

Kimetrica 2020. Webpage on Modeling and Simulation. Accessed September 17, 2020. https://kimetrica.com/services/modeling-and-simulation/

Knippenberg, E., Jensen, N., & Constas, M. (2019). Quantifying household resilience with high frequency data: Temporal dynamics and methodological options. *World Development*, *121*, 1–15. https://doi.org/10.1016/j.worlddev.2019.04.010

Krishnamurthy, K., Fisher, J. B., Schimel, D. S., & Kareiva, P. M. (2020b). Applying Tipping Point Theory to Remote Sensing Science to Improve Early Warning Drought Signals for Food Security. *Earth's Future*, *8*(3), 14. https://doi.org/10.1029/2019EF001456

Krishnamurthy, P. K., Choularton, R. J., & Kareiva, P. (2020a). Dealing with uncertainty in famine predictions: How complex events affect food security early warning skill in the Greater Horn of Africa. *Global Food Security*, *26*, 100374. https://doi.org/10.1016/j.gfs.2020.100374

Kuzma, S., Kerins, P., Saccoccia, L., Whiteside, C., Roos, H., & Iceland, C. (2020). Leveraging Water Data in a Machine Learning–Based Model for Forecasting Violent Conflict. *World Resources Institute*, 40.

Lee, K., & Braithwaite, J. (2020). *High-Resolution Poverty Maps in Sub-Saharan Africa*. https://arxiv.org/abs/2009.00544

Lentz, E. C., Michelson, H., Baylis, K., & Zhou, Y. (2019). A data-driven approach improves food insecurity crisis prediction. *World Development*, *122*, 399–409. https://doi.org/10.1016/j.worlddev.2019.06.008

Lowcock, M. (2018, March 23). *A Casement Lecture: Towards a Better System for Humanitarian Financing - World*. ReliefWeb. https://reliefweb.int/report/world/under-secretary-general-humanitarian-affairs-and-emergency-relief-coordinator-mark-0

Matere, J., Simpkin, P., Angerer, J., Olesambu, E., Ramasamy, S., & Fasina, F. (2020). Predictive Livestock Early Warning System (PLEWS): Monitoring forage condition and implications for animal production in Kenya. *Weather and Climate Extremes*, *27*, 1–8. https://doi.org/10.1016/j.wace.2019.100209

Maxwell, Daniel, & Hailey, P. (2020a). *Towards Anticipatory Information Systems and Action: Notes on Early Warning and Early Action in East Africa*. Centre for Humanitarian Change and Feinstein International Center.

Maxwell, Daniel, Vaitla, B., & Coates, J. (2014). How do indicators of household food insecurity measure up? An empirical comparison from Ethiopia. *Food Policy*, 47: 107-116.

McBride, Linden, et al. (under review) "Predicting poverty and malnutrition for targeting, mapping, monitoring, and early warning."

McNally, A., Verdin, K., Harrison, L., Getirana, A., Jacob, J., Shukla, S., Arsenault, K., Peters-Lidard, C., & Verdin, J. P. (2019). Acute Water-Scarcity Monitoring for Africa. *Water*, *11*(10), 1968. https://doi.org/10.3390/w11101968

Milinovich, G. J., Magalhães, R. J. S., & Hu, W. (2015). Role of Big Data in the Early Detection of Ebola and Other Emerging Infectious Diseases. *The Lancet Global Health* 3 (1), e20–e21. https://doi.org/10.1016/S2214-109X(14)70356-0

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy* (First edition.). Crown.

Paul, A., Jolley, C., & Anthony, A. (2019). Reflecting the Past, Shaping the Future: Making AI Work for International Development. Digital Development, U.S. Agency for International Development. https://www.usaid.gov/digital-development/machine-learning/ai-ml-in-development

Peng, B., Guan, K., Zhou, W., Jiang, C., Frankenberg, C., Sun, Y., He, L., & Köhler, P. (2020). Assessing the benefit of satellite-based Solar-Induced Chlorophyll Fluorescence in crop yield prediction. *International Journal of Applied Earth Observation and Geoinformation*, *90*, 102126. https://doi.org/10.1016/j.jag.2020.102126

Shukla, S., Arsenault, K. R., Hazra, A., Peters-Lidard, C., Koster, R. D., Davenport, F., Magadzire, T., Funk, C., Kumar, S., McNally, A., Getirana, A., Husak, G., Zaitchik, B., Verdin, J., Nsadisa, F. D., & Becker-Reshef, I. (2020). Improving Early Warning of Drought-Driven Food Insecurity in Southern Africa Using Operational Hydrological Monitoring and Forecasting Products. *Natural Hazards and Earth System Sciences* 20 (4), 1187–201. https://doi.org/10.5194/nhess-20-1187-2020

Siegel, Eric. (2016). Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die. Revised and Updated. Hoboken, New Jersey: John Wiley & Sons. https://learning.oreilly.com/library/view/~/9781119145677/?ar

Steele, J. E. et al. 2017 Mapping poverty using mobile phone and satellite data. *J. R. Soc. Interface* 14: 20160690. http://dx.doi.org/10.1098/rsif.2016.0690

Tandon, S., & Vishwanath, T. (2020). The evolution of poor food access over the course of the conflict in Yemen. *World Development*, *130*, 104922. https://doi.org/10.1016/j.worlddev.2020.104922

van den Homberg, Gevaert, C. and Georgiadou, Y. (2020). "The changing face of accountability in humanitarianism: Using artificial intelligence for anticipatory action" *Politics and Governance* 8(4). DOI: 10.17645/pag.v8i4.3158

Varian, H. R. (2014). Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives*, *28*(2), 3–28. https://doi.org/10.1257/jep.28.2.3

Westerveld, J.L., van den Homberg, M.J.C., Guimaraes Nobre, G., van den Berg, D.L.J, Teklesadik, A.D., Stuit, S.M. (2021) "Forecasting transitions in the state of food security with machine learning using transferable features." Science of the Total Environment. 786 (2021):147366. https://doi.org/10.1016/j.scitotenv.2021.147366

World Food Programme (WFP). (2020). *World Food Programme HungerMap LIVE*. https://hungermap.wfp.org/

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., et al. (2020). "Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa." Nature Communications 11 (1): 2583.

You, J., Li, X., Low, M., Lobell, D., & Ermon, S. (2017). Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. *Thirty-First AAAI Conference*. Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17). https://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14435

Zhou, Y., Lentz, E., Michelson, H., Kim, C., and Baylis, K (under review) "Machine Learning for Food Security: Principles for transparency and usability." *Applied Economic Perspectives and Policy.*