Predicting Food-Security Crises in the Horn of Africa Using Machine Learning

Tim Sebastiaan Busker[1], Bart van den Hurk[2], Hans de Moel[1], Marc van den Homberg[3], Chiem van Straaten[1], Rhoda A. Odongo[1], and Jeroen C.J.H. Aerts[4]

[1]Vrije Universiteit Amsterdam
[2]Deltares
[3]510 an initiative of the Netherlands Red Cross
[4]Vrije Universiteit

November 8, 2023

Abstract

The Horn of Africa region has frequently been affected by severe droughts and food crises over the last several decades, and this will increase under projected global-warming and socio-economic pathways. Therefore, exploring novel methods of increasing early warning capabilities is of vital importance to reducing food-insecurity risk. In this study, we present the XGBoost machine-learning model to predict food-security crises up to 12 months in advance. We used >20 datasets and the FEWS IPC current-situation estimates to train the machine-learning model. Food-security dynamics were captured effectively by the model up to three months in advance (R2 > 0.6). Specifically, we predicted 20% of crisis onsets in pastoral regions (n = 84) and 40% of crisis onsets in agro-pastoral regions (n = 23) with a 3-month lead time. We also compared our 8-month model predictions to the 8-month food-security outlooks produced by FEWS NET. Over a relatively short test period (2020–2022), results suggest the performance of our predictions is similar to FEWS NET for agro-pastoral and pastoral regions. However, our model is clearly less skilled in predicting food security for crop-farming regions than FEWS NET. With the well-established FEWS NET outlooks as a basis, this study highlights the potential for integrating machine-learning methods into operational systems like FEWS NET.

Hosted file

`977124_0_art_file_11503580_s2rrpb.docx` available at https://authorea.com/users/677104/articles/674324-predicting-food-security-crises-in-the-horn-of-africa-using-machine-learning

Hosted file

`977124_0_supp_11503400_s2r82p.docx` available at https://authorea.com/users/677104/articles/674324-predicting-food-security-crises-in-the-horn-of-africa-using-machine-learning

Predicting Food-Security Crises in the Horn of Africa Using Machine Learning

Tim Busker,[1] Bart van den Hurk,[2,1] Hans de Moel,[1] Marc van den Homberg,[4] Chiem van Straaten,[1,3] Rhoda A. Odongo[1] and Jeroen C.J.H. Aerts[1,2]

[1] Institute for Environmental Studies (IVM), Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

[2] Deltares, Delft, the Netherlands

[3] Royal Netherlands Meteorological Institute, De Bilt, the Netherlands

[4] 510, an initiative of the Netherlands Red Cross, the Hague, the Netherlands

Corresponding author: Tim Busker (tim.busker@vu.nl)

Key Points

- A machine-learning model is presented to predict food-security crises in the Horn of Africa.
- The model demonstrates high overall performance, and performs similarly to FEWS NET outlooks in the (agro-) pastoral regions.
- This study can be utilized to integrate machine learning in existing early warning systems, to creating hybrid solutions for the future.

CRediT authorship contribution statement

Conceptualization: Tim Busker, Jeroen C.J.H. Aerts, Bart van den Hurk and Hans de Moel
Data curation: Tim Busker and Rhoda A. Odongo
Software: Tim Busker
Formal analysis: Tim Busker
Methodology: Tim Busker, Chiem van Straaten**,** Bart van den Hurk, Hans de Moel, Marc van den Homberg, and Jeroen C.J.H. Aerts
Project Administration: Tim Busker, Jeroen C.J.H. Aerts, Bart van den Hurk and Hans de Moel

Abstract

The Horn of Africa region has frequently been affected by severe droughts and food crises over the last several decades, and this will increase under projected global-warming and socio-economic pathways. Therefore, exploring novel methods of increasing early warning capabilities is of vital importance to reducing food-insecurity risk. In this study, we present the XGBoost machine-learning model to predict food-security crises up to 12 months in advance. We used >20 datasets and the FEWS IPC current-situation estimates to train the machine-learning model. Food-security dynamics were captured effectively by the model up to three months in advance ($R^2 > 0.6$). Specifically, we predicted 20% of crisis onsets in pastoral regions (n = 84) and 40% of crisis onsets in agro-pastoral regions (n = 23) with a 3-month lead time. We also compared our 8-month model predictions to the 8-month food-security outlooks produced by FEWS NET. Over a relatively short test period (2020–2022), results suggest the performance of our predictions is similar to FEWS NET for agro-pastoral and pastoral regions. However, our model is clearly less skilled in predicting food security for crop-farming regions than FEWS NET. With the well-established FEWS NET outlooks as a basis, this study highlights the potential for integrating machine-learning methods into operational systems like FEWS NET.

Plain Language Summary

In the face of increasing droughts and food crises, this study explored the use of machine learning to provide predictions of food crises in the Horn of Africa, up to 12 months in advance. We used an algorithm called "XGBoost", which we fed with over 20 datasets of potential food security drivers. After training the model, we found that food security dynamics were accurately predicted up to three months in advance, especially in pastoral and agro-pastoral regions. Impressively, the model accurately predicted 20% of crisis onsets in pastoral areas and 40% in agro-pastoral regions with a three-month lead time. In agro-pastoral and pastoral regions, our machine learning algorithm showed a similar performance to the established early warning system from FEWS NET. The machine-learning model did not show good performance in crop-farming areas. Nonetheless, this study underscores the potential of integrating machine-learning methods into existing operational systems like FEWS NET. By doing so, it paves the way for improved early warning capabilities, crucial in mitigating the looming threat of food insecurity in the Horn of Africa.

Keywords: early warning, drought, food insecurity, famine, machine learning, Horn of Africa

1 Introduction

The Horn of Africa is one of the world's most vulnerable regions for food security, with around 57 million people experiencing extreme poverty (UNHCR, 2023). The 2020–2023 drought caused by five consecutive failed rainy seasons was the worst in 40 years (World Meteorological Organization [WMO], 2022). It plunged >20 million people into conditions of high food insecurity and caused acute malnutrition among 7 million children (UNHCR, 2023). Recent literature suggests that these extreme droughts may increase in frequency under anthropogenic warming (Baxter et al., 2023; Funk et al., 2023; Kimutai et al., 2023). Together with expected population growth, this change will further increase the number of food-insecure people over the coming decades (Funk & Shukla, 2020). These trends emphasize the importance of strengthening food-security early warning systems and increasing the understanding of drivers of food-security crises in different contexts.

Most of the farmers in the region are dependent on long rains in the March-April-May (MAM) and short rains in the October-November-December (OND) seasons. Meteorological droughts in East Africa have been increasingly observed over the last decades, especially during the MAM season (Funk, Shukla, et al., 2019). These droughts often lead to food insecurity, increasing the demand for seasonal food-security early warning systems. Currently, several drought early warning systems with a focus on food security are operational, such as the Hunger Hotspot early warnings from the World Food Programme (WFP) and the Food and Agriculture Organization of the United Nations (FAO) (WFP and FAO, 2022), and the FAO Global Information and Early Warning System on Food and Agriculture (GIEWS) (FAO, 2023b). The most widely used early warning system for Africa is the U.S. Agency for International Development's Famine Early Warning Systems Network (FEWS NET). FEWS NET uses a combination of observed and forecasted drought indicators, vulnerability indicators and expert judgment to make local and regional assessments of food security. This process is conducted using key Integrated Food Security Phase Classification (IPC) protocols (IPC, 2023). These assessments not only pertain to the food-security assessment of the current situation but also include projections of food security for the near term (up to 4 months in future) and medium term (up to 8 months in future). These projections are summarized in Food Security Outlooks (FEWS NET, 2023b).

Together with national partners, humanitarian agencies such as the Red Cross Red Crescent Movement and WFP have introduced anticipatory action for food insecurity and drought over the last decade (WFP, 2023b). Food-security outlooks and reliable early warning signals are crucial to triggering these anticipatory actions. Although hydrological and agricultural drought predictions—soil moisture (Shukla et al. 2014) and crop yield (Boult et al., 2020; Shukla et al., 2020)—are often accurate, predicting food-security crises remains challenging. For example, FEWS NET food-security outlooks continue to face challenges (Backer & Billing, 2021; Krishnamurthy et al., 2020), as food-security dynamics are often unpredictable because of region-specific or unexpected drivers (e.g. conflict or desert-locust outbreaks).

Recent studies have made advancements in adopting machine-learning approaches to tackle these challenges. Promisingly, certain studies demonstrate that machine learning holds the potential to efficiently monitor (Martini et al., 2022) and predict (Foini et al., 2023; Westerveld et al., 2021) food consumption and food insecurity by utilizing data on their drivers. Nonetheless, significant gaps persist. First, there is a scarcity of studies assessing the accuracy of machine-learning-based food-security predictions, particularly for broad regions like East Africa. Second, the underlying dynamics and potential drivers of food security across various lead times are largely unknown due to a lack of use of explainable machine-learning techniques. Third, machine-learning predictions are rarely directly compared with predictions from existing operational early warning systems. Bridging these gaps is essential to understanding the potential of such machine-learning algorithms in different contexts, and to understand how they can be adopted and integrated into the currently used consensus-based approaches.

Consequently, the primary objective of this study is to develop and test machine-learning models for predicting food security in the Horn of Africa. This process will increase the understanding of lead-time-dependent food-security drivers and reveal the performance of these machine-learning models. We will compare our machine-learning predictions with FEWS NET food-security outlooks to identify where our predictions can provide additional value.

We begin with an outline of the methodological framework (Section 2). In the results (Section 3), we outline the forecast accuracy and the use of explainable machine-learning techniques to elucidate the underlying dynamics of food-security drivers (Section 3). Subsequently, we consider these results in the discussion (Section 4) and provide our main conclusions and recommendations (Section 5).

2. Methods

Figure 1 shows the setup of our approach to predicting food security in the Horn of Africa, specifically Somalia, Kenya, and Ethiopia. We use the XGBoost model (Chen & Guestrin, 2016) as our machine-learning model to predict IPC food-security status on monthly and seasonal time scales (Section 2.1.1). XGboost, or Extreme Gradient Boosting, is an ensemble decision-tree algorithm that is like random forest regressions but able to model more complex interactions due to its ability to boost individual trees. We add multiple explanatory hazard and vulnerability variables (Section 2.1.2) as potential drivers for food insecurity, which we refer to as "features". Subsequently, the XGBoost model (Section 2.2.1) is trained (2009–2020) to predict FEWS IPC food-security states on multiple lead times. These predictions are tested on a separate "hold-out" dataset (2020–2022) of FEWS IPC values (Section 2.2.2). In this stage, three benchmark models are used to which our predictions will be compared: 1) the FEWS NET outlooks, 2) a persistence model (prediction same as now), and 3) a seasonality model (Section 2.5). These steps are visualized in Figure 1.

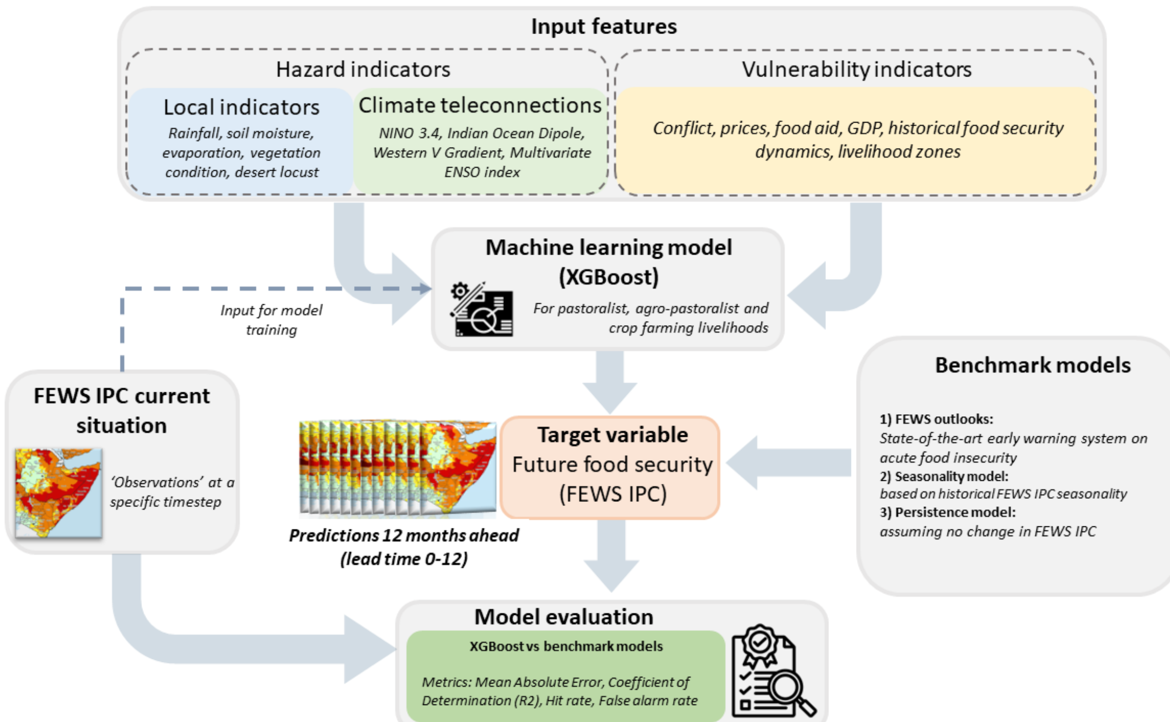

Figure 1 This study design shows the input data employed in the modeling framework, which includes features used (top) and the FEWS IPC current-situation maps (left). These elements feed into the machine-learning model (XGBoost) and are utilized to train the model and make predictions of future food security (center). We compare these predictions to the current situation to evaluate the machine-learning model (bottom). Predictions are also made using three benchmark models (right), one of which is the state-of-the-art outlook from the FEWS NET early warning system.

2.1 Data overview

The modeling framework aims to predict FEWS IPC acute food-insecurity values (target variable; Section 2.1.1). To make these predictions, 20 input variables (hereafter referred to as "features"; Section 2.1.2) are included. The mean of each feature and the median of the FEWS NET data were calculated for each administrative unit (admin-1 or -2 levels). For Somalia and Ethiopia, the admin-2 level was used, but for Kenya, the admin-1 level was used, as the admin-2 regions (290 in total) resulted in an impractical level of spatial differentiation for the modeling. Overall, 17 administrative units with no considerable variation in FEWS IPC food-security status (standard deviation < 0.01) were excluded from the analysis. This resulted in a total of 196 administrative units included in the study.

2.1.1 FEWS IPC food-security outcomes

As described above, the FEWS IPC food-security maps are chosen as the target variable to be predicted with the model. We first provide background information on how the IPC food-security status is determined, after which we outline how the data was pre-processed to be used in the model.

Acute food-insecurity monitoring using IPC

Global food-security assessments consist of four main pillars: food access, food availability, food utilization, and stability (FAO, 2009). The Integrated Phase Classification (IPC, 2023) was developed to represent and evaluate these pillars. The IPC, however, uses three different scales to measure food security and nutrition: acute food insecurity, chronic food insecurity, and acute malnutrition.

In this study, we focus on acute food insecurity. The IPC estimates the magnitude of acute food insecurity and identifies its key drivers (IPC, 2021; Figure 27). Acute food insecurity is measured using internationally recognized scientific standards and cut-offs on a five-phase scale: Phase 1, minimal/none; Phase 2, stressed; Phase 3, crisis; Phase 4, emergency; and Phase 5, catastrophe/famine (IPC, 2021). This system defines first-level food-security outcomes—food-consumption gaps and negative livelihood change—which, if the situation worsens, result in second-level acute malnutrition and mortality outcomes. The combination of these outcomes determines the IPC classification, as malnutrition and mortality can also be caused by factors other than food-consumption gaps. Negative livelihood change is an important indicator because unsustainable livelihood practices, such as the reduction of health expenditures or risky migration, temporarily decrease food-consumption gaps but strongly increase long-term vulnerability. Therefore, IPC can, in this case, still assign a high acute food-insecurity class (IPC, 2021).

The "target variable": FEWS IPC food-security maps

The IPC acute food-security estimates of the current situation, provided by FEWS NET, are used as the main target variable for our machine-learning model (Figure 1, left). The maps are downloaded as shapefiles from the FEWS NET data portal (FEWS NET, 2023a). We use area-level classifications, which assign the highest food-security class faced by at least 20% of the population. From these maps, we calculated the spatial mean per administrative unit.

2.1.2 The "features": Potential drivers of food insecurity

We use a total of 20 features in the models (Table 1), which are classified into hazard (local hazard indicators and climate teleconnections) and vulnerability features.

Hazard data

Rainfall indicators: We used daily CHIRPS rainfall data (Funk et al., 2015) over the period 1981–2022. We calculated total rainfall, total number of wet days (>1mm/day), and maximum dry-spell length (> 5 consecutive dry days) per month. The maximum dry-spell length can be >31 consecutive days if the dry spell extends over several months.

Drought indices: We included three different drought indices that are pivotal for objective drought monitoring at different spatial and temporal scales: the standardized precipitation index (SPI) (McKee et al., 1993), the standardized precipitation evapotranspiration index (SPEI) (Vicente-Serrano et al., 2010), and the standardized soil moisture index (SSMI) (Blauhut et al., 2016; Hao et al., 2014). The SPI measures meteorological drought conditions and is based on CHIRPS. To include wider atmospheric conditions, we calculated the SPEI, which is the standardized difference between precipitation and potential evapotranspiration (PET). The PET was retrieved from the global land evaporation Amsterdam model (GLEAM) (version 3.5a; Martens et al., 2017) and reflects atmospheric conditions such as wind speed, temperature, and humidity. The GLEAM model uses satellite and reanalysis data to estimate land-surface evaporation and soil moisture on a 0.25-degree grid. Additionally, SSMI is derived from GLEAM using the root-zone soil-moisture dataset.

The above drought indices are calculated using the methodology described in Odongo et al. (2023). Specifically, the monthly indices are derived by accumulating the input variables (rainfall for SPI, rainfall and PET for SPEI, and root-zone soil moisture for SSMI) over 1, 3, 6, 12, and 24 months. Subsequently, a distribution was fitted through the accumulated variables, and the data was standardized by comparing these variables with the amount of the variable that would have been expected based on the long-term climatology (1981–2022). For the standardization, we used a statistical distribution that best fitted the data between –3 and +3. Multiple distributions were tested for each of the indices per period and administrative unit. The distribution with the best fit based on the Kolmogorov best-fit test was selected (see Odongo et al., 2023, for details). The calculation of the SPI was corrected for zero values in the distribution, as recommended by Stagge et al. (2015).

Table 1

The features in the model, including local hazard indicators (blue), climate teleconnections (green), and vulnerability indicators (yellow).

Hazard indicators				**Vulnerability indicators**	
Local hazard indicators		***Climate teleconnections***			
Name	*Source*	*Name*	*Source*	*Name*	*Source*
Total rainfall	(CHIRPS; Funk et al., 2015)	Indian Ocean Dipole (IOD)	(NOAA, 2023b)	ACLED: number of conflicts and fatalities	(Raleigh et al., 2010)
Number of dry spells	(CHIRPS; Funk et al., 2015)	Multi-variate ENSO index (MEI)	(NOAA, 2023a)	Food and fuel prices	(WFP, 2023a)
Number of wet days	(CHIRPS; Funk et al., 2015)	NINO3.4	(NOAA, 2023a)	Historical and current food-security situation	(FEWS NET, 2023a)
Standardized precipitation index (SPI)	(CHIRPS; Funk et al., 2015)	Western V gradient (WVG)	(Funk et al., 2023)	Humanitarian food assistance	(FEWS NET, 2023a)
Standardized soil moisture index (SSMI)	(Martens et al., 2017)			Headline and food consumer price index	Somalia (NBS, 2023). Kenya and Ethiopia (Ha et al., 2021)
Standardized precipitation and evaporation index (SPEI)	(Funk et al., 2015; Martens et al., 2017)			Gross domestic product (GDP) per capita	(IMF, 2023)
Normalized Difference Vegetation Index (NDVI)	(NOAA, 2021)				
NDVI croplands	(NOAA, 2021; Pérez-Hoyos, 2018)				
NDVI rangelands	(NOAA, 2021; Pérez-Hoyos, 2018)				
Desert-locust swarms	(FAO, 2022)				

Agricultural indicators: We included NDVI as the agricultural drought indicator, derived from the NOAA STAR Center for Satellite Applications and Research (NOAA, 2021). This dataset contains data from the Advanced Very-High-Resolution Radiometer (AVHRR) sensor. The archive contains validated seven-day composites of smoothed NDVI data at 4 km^2 resolution. We extracted rangeland NDVI and cropland NDVI values using crop and rangeland masks from the Anomaly Hotspots of Agricultural Production (ASAP) system (Pérez-Hoyos, 2018). Subsequently, the NDVI values were expressed as anomalies per month using 2000–2021 as the reference period.

Desert locusts: We included over 10,000 data points on swarms of desert locusts obtained from the FAO Locust Hub (FAO, 2022). The total area affected per administrative unit was calculated for every month and administrative unit across the three countries.

Climate teleconnections: The climate in the Horn of Africa is strongly influenced by sea surface temperatures (SSTs) in the Indian and Pacific Oceans (Funk et al., 2023). Therefore, we included multiple variables representing these SSTs: the Indian Ocean Dipole (IOD) (NOAA, 2023b), the multivariate ENSO index (MEI), and NINO 3.4 ((NOAA, 2023a). Recent research has also discovered a new gradient in the Pacific Ocean called the Western V gradient (WVG), which is linked to the severe drought conditions in East Africa observed over the past several years (Funk et al., 2023). Consequently, we included the WVG as observed during the MAM season. All SST indices used in this research are visualized in Figure S1.

Vulnerability data

Food and fuel prices: We used food and fuel prices from the WFP's price database, the VAM Food Security Portal (WFP, 2023a). We selected maize as the main food crop for each country based on information from the Global Information and Early Warning System from FAO (FAO, 2023a). Other crop prices were not included due to limited data availability in WFP's price database. We also included fuel prices (diesel), as this was an important driver of past food crises (WFP and FAO, 2022). We used the Alert for Price Spikes (ALPS) indicator (WFP, 2014) as a means to obtain standardized prices that are corrected for the long-term (seasonal) trend. Using this indicator, WFP aims to detect price spikes and abnormal price deviations beyond long-term trends. Details of the ALPS-indicator calculation can be found in WFP's ALPS manual (WFP, 2014).

The data is provided for 14 markets in Kenya, 98 markets in Ethiopia, and 29 markets in Somalia. We geolocated these markets and subsequently selected the closest market for each administrative unit. For each time step, data gaps in the closest market for a specific administrative unit are filled by the closest market in the country for which data is available.

Macroeconomic indicators: We include inflation using data from the National Bureau of Statistics (NBS) for Somalia (NBS, 2023) and the World Bank Global Inflation Dataset (Ha et al., 2021) for Kenya and Ethiopia. This includes the consumer price index (CPI) as a measure for overall inflation ("headline CPI") and inflation on food products specifically ("food CPI"). Furthermore, we used the gross domestic product (GDP) per capita as an indicator for national economic growth (IMF, 2023).

Historical and current food-security situation: Upcoming food-security dynamics are dependent on current and past food-security situations. For example, low food-insecurity stages more easily transition to high food insecurity than vice versa (Wang et al., 2020). XGBoost can learn such relationships, so we include the past and current IPC values as features in the model.

Humanitarian food assistance: Data on the impact of humanitarian food assistance has been extracted from the FEWS NET data portal (FEWS NET, 2023a). Such aid includes direct food assistance (i.e., in-kind food transfers) but may also include indirect food assistance (i.e., cash or livestock assistance). The data marks areas that would likely have been one phase more food insecure without significant humanitarian food assistance (FEWS NET, 2023a), which are indicated with an exclamation mark (!) in the food-security maps published by FEWS NET.

Conflicts: Conflict data was extracted from the Armed Conflict Location & Event Data Project (ACLED; Raleigh et al., 2010) dataset. Since 1997, the ACLED dataset has collected events of political violence and protest across 50 states worldwide, such as those resulting from rebels, governments, or militias. Each entry represents a single event of a specific type at a particular location on a given day. We calculated both the total number of conflicts and the total number of fatalities per administrative unit per month.

2.2 The machine-learning model architecture

2.2.1 The XGBoost model

Decision-tree models have shown great potential for impact-based forecasting of climatic shocks (Everingham et al., 2016; Guimarães Nobre et al., 2019; Schoppa et al., 2020; Westerveld et al., 2021). In our study, we selected XGBoost (eXtreme **G**radient **Boost**ing) as the regression model of choice (Chen & Guestrin, 2016; Friedman, 2001). XGBoost is an ensemble tree model that, in contrast to normal decision trees, does not rely on a single tree. Instead, it creates an ensemble of *n* different decision trees and uses a scalable tree-boosting system to optimize predictions. These decision trees are shallow (weak) learners that are iteratively added to minimize the errors of previous predictions while simultaneously being subject to regularization.

We selected this model due to its demonstrated speed and more effective performance compared to other models in many different fields (Chen & Guestrin, 2016), including drought and food-security prediction (Foini et al., 2023; Martini et al., 2022; Westerveld et al., 2021; Zhang et al., 2023). The tree-boosting systems allow XGBoost to better model complex and non-linear relationships, which are often present in food-security systems.

A different XGBoost model was created for each lead time (0, 1, 2, 3, 4, 8, and 12 months), which resulted in seven separate models. The data from the 196 individual administrative units was pooled in three livelihood zones (Figure 3): 1) pastoral, 2) agropastoral, and 3) crop farming. A different model was made for each livelihood zone, which, combined with the different lead times, resulted in 21 unique machine-learning models. The pastoral and crop-farming regions are the largest, with 82 and 89 administrative units, respectively, whereas the agro-pastoral regions consist of 25 different units.

2.2.2 Machine learning setup: Train–test–validation

Figure 2 shows how the machine learning was set up. For each model, the time series are split into a training and test set based on an 80:20 time ratio. This results in a training dataset from 2009–2020 (Figure 2, blue line) and a test dataset from 2020–2022 (Figure 2, red line). The time series presented are for one of the 196 administrative units (Mandera in Kenya). The test dataset is out-of-sample, meaning that we leave it untouched and only use it for model testing. An out-of-sample approach is a beneficial practice in time-series forecasting to ensure temporal independence of the dataset (Cerqueira et al., 2020). We did not shuffle the observations prior to the train-test splitting, because maintaining the original temporal sequence of observations is crucial for time series data (see for example, Snijders, 1988; Cerqueira et al., 2020).

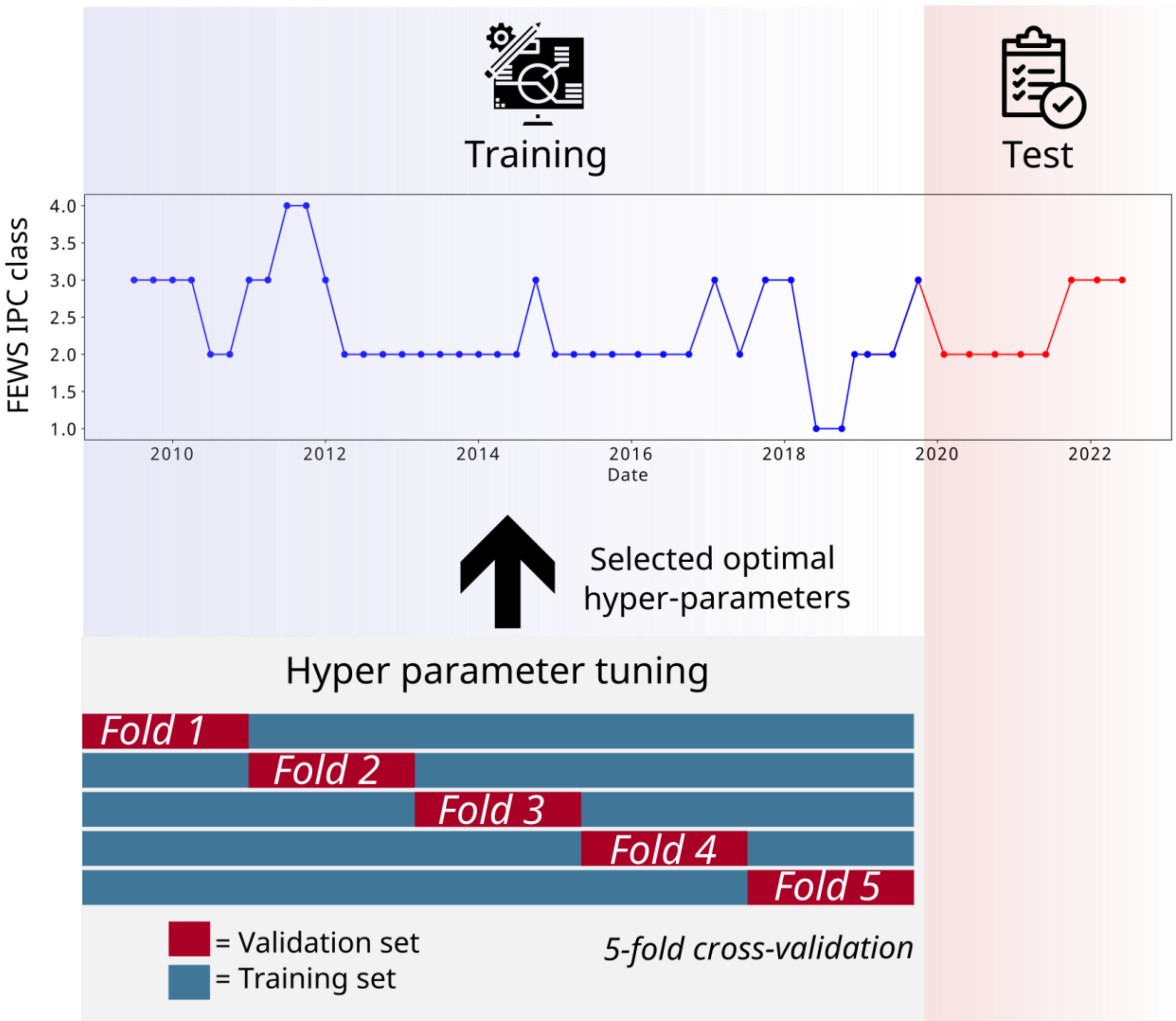

Figure 2 The setup of the machine-learning model splits the features and target variables into training and test parts. The training part is also used for hyperparameter tuning of the model using 5-fold cross validation. As an illustration, the time series shown represents the FEWS IPC values for Mandera in Kenya.

Hyperparameter optimalisation

We executed hyperparameter tuning to optimize model performance, using 5 different validation sets defined in five different folds (Figure 2, bottom). We found the following optimal hyperparameters: maximum tree depth: 4, number of trees/estimators: 400 and learning rate: 0.01 (see Table S1). Extended information on the hyperparameters and the tuning process can be found in the Supporting Information (Text S1). The evaluation scores on the five validation sets (i.e., the five different folds) are shown in Table S2.

2.3 Feature engineering

We further process the features from the data series listed in Section 2.1 using feature engineering (Zheng & Casari, 2018). This brings time and memory effects to the XGBoost model. For each input feature, we computed the rolling average over both the last 4 and 12 months for each individual month. We identified the timing of the main rainy seasons in the Horn of Africa, MAM and OND, which are marked as "1" in the model dataset. Memory effects in the target variable FEWS IPC were accounted for by including values from 1, 4, and 8 months prior, along with the mean FEWS IPC value from the past 12 months. Additionally, we incorporated country names into the model, enabling it to factor in country-specific elements (such as drought-intervention policies) not included in the original data. This resulted in a total of 81 unique features.

2.4 Consideration of lead time

Predictions were made for various lead times (0, 1, 2, 3, 4, 8, and 12 months) using seven distinct XGBoost models. Each model was trained with specific lags corresponding to the respective lead time. Before training and validation, the feature timestamps were adjusted forward based on the lead time, creating a time lag between features and the target variable. This allows the model to learn the relationships between features and FEWS IPC classes separated by the prediction lead time.

Predictions are only generated for months when the FEWS IPC observation is released, which occurs three times annually. When predicting the food-security outcome with a one-month lead time, the features utilize data recorded in the "current" month and the preceding months to make a forecast for the next month. This "historical" data is integrated through feature engineering, as outlined in Section 2.3.

2.5 Benchmark models and performance metrics

The predictions in the test set (2020–2022) were evaluated using the mean absolute error (MAE), the coefficient of determination (R^2), the hit rate, and the false-alarm rate. Three benchmark models were used and served as a performance reference: 1) the state-of-the-art *FEWS NET food-security* outlooks, 2) a *seasonality model* based on historical FEWS IPC observations, and 3) *a persistence model* assuming no change in the FEWS IPC class. The seasonality model makes predictions using the monthly average FEWS IPC value for each administrative unit, as calculated over the training period (2009–2020). Every prediction with the seasonality model is similar for different lead times because they all use the seasonality from the training set. The persistence model relies on the last-observed FEWS IPC value (current situation) for making predictions, which is issued three times a year. However, for lead times beyond 3 months, the model cannot use the last value. Instead, it must utilize the FEWS IPC observation before it. For example, we assume that the FEWS NET current-situation release occurs by the end of the month, and therefore the persistence predictions for October on lead 0 cannot make use of the FEWS NET release in that month. The FEWS NET food-security outlooks require a more detailed explanation, which we provide below.

2.5.1 The FEWS NET food-security outlooks

The FEWS NET food-security outlooks are state-of-the-art projections from FEWS NET. They are the result of a rigorous scenario-development process, which leads to a "most likely" future food-security scenario (FEWS NET, 2018). The outlooks utilize different information sources, such rainfall and temperature observations, but also climate modes, including ENSO. At lead times of 3–6 months, FEWS NET uses long-range seasonal forecasts, such as root-zone soil moisture (Shukla et al., 2020). Local vulnerability is incorporated through knowledge and experience of livelihoods, market dynamics, and nutrition (WMO, 2017).

Two types of food-security outlooks are created through FEWS NET: near term (1–4 months into the future) and medium term (4–8 months into the future). These outlooks, released every month, target the month(s) just prior to the FEWS NET current-situation observations. To validate the outlooks, we compared them to the next available current-situation observation in February, June, or October.

2.6 Interpretation of model results

Machine learning is often criticized as a black box (McGovern et al., 2019), which emphasizes the need to increase model transparency. Therefore, we use the SHAP (Shapley Additive Explanations) (Lundberg & Lee, 2017) framework to interpret model predictions and understand how the model uses the features. Shapley values originate from game theory (Shapley, 1953) and are solutions to the problem of dividing a game's single payout among all players according to their respective contributions. In this case, the payout is the prediction of the statistical model, and features are the contributors. This framework is unique in the sense that it shows the impact of every individual feature on each prediction, which is also called "local feature importance". The SHAP values for every input feature reveal how that feature changed the model prediction at that specific time step compared to the SHAP base values. We used the default SHAP baseline, which reflects the model prediction without using any features (Lundberg & Lee, 2017). Thus, SHAP can reveal the influence of any of the features on any prediction. This differentiates SHAP from the many other explanation methods based on global interpretation that only show the contribution of the features to the model as a whole. Nonetheless, combining all local SHAP values provides a realistic view of global feature importance (Lundberg et al., 2020).

3 Results

3.1 General model evaluation

An illustration of the predictions with the observations demonstrates the ability of the XGBoost model to predict food-security dynamics over different administrative units in the region (Figure 3). This includes (1) the correct timing of the onset of crisis in, for example, Burco, Waajid, and Garissa and (2) the dynamics of phases in low food security (e.g., in Baringo). Both the timing and dynamics are predicted effectively, with R^2 values ranging from 0.41 in Tigray to 0.87 in Afmadow.

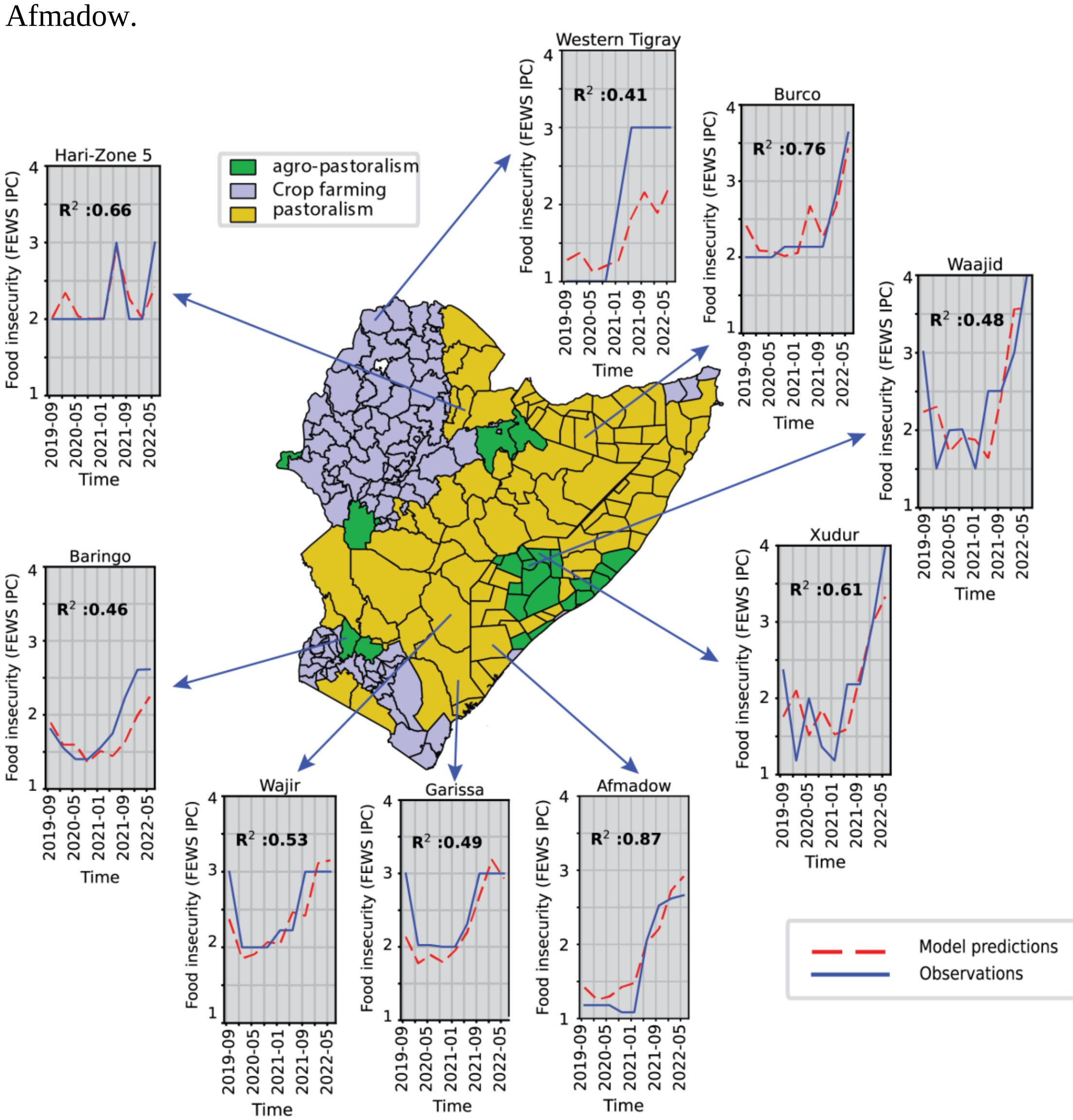

Figure 3 Time series show model predictions of food insecurity based on the FEWS IPC categories, with a lead time of 3 months (red lines) compared to observations (blue lines). Examples of the different administrative units are selected over various regions and livelihood zones. The map displays the three individual livelihood zones for which the machine-learning models are trained.

Figure 4 shows an assessment of the quality of the 3-month forecasts with the XGBoost model across the Horn of Africa region. They have an average MAE (mean absolute error) of 0.36. This is low compared to the range of possible FEWS IPC values (1–5). Outliers with poor skill are the Tigray region in northern Ethiopia and Turkana in north-western Kenya. The Tigray region experienced a sudden increase in food insecurity during our test period (2020–2022) resulting from the outbreak of armed conflict in November 2020. Conflicts are included in the model through the ACLED dataset. However, the relationship between conflict and food insecurity is complex (see Section 4.3) and led to a large underestimation of food insecurity in Tigray (see Figure 4 and Figure 3, Western Tigray).

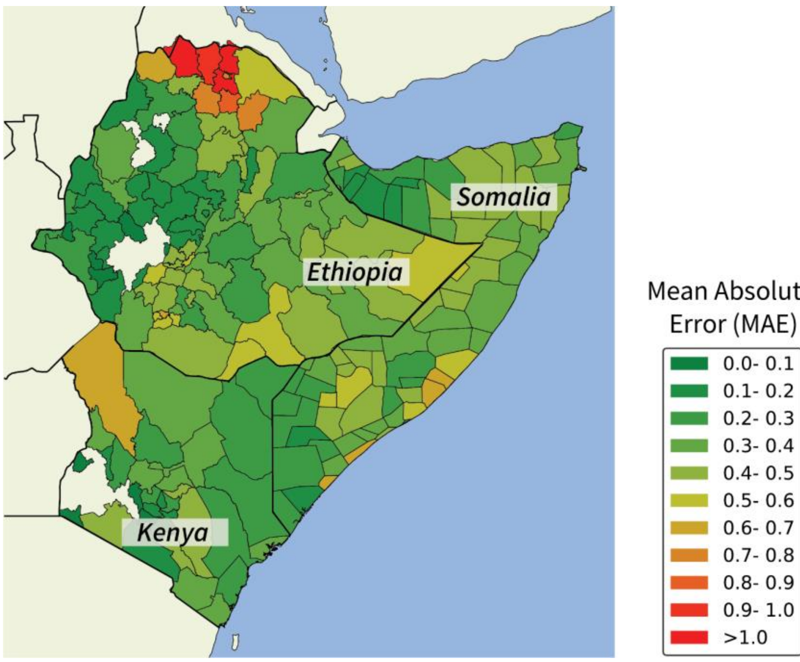

Figure 4 A spatial map of the mean absolute error (MAE) for the XGBoost machine-learning predictions of FEWS IPC food security over the three months ahead of the observation. Metrics are calculated from the test set (2020–2022).

The XGBoost model dynamics were further evaluated using the coefficient of determination (R^2). Over the whole region, high R^2 values were found on short lead times ($R^2 = 0.67$ on 1 month; Figure 5, bottom). We also assessed the model performance in different livelihood zones (Section 2.2.1) as compared to the benchmark models (Figure 5, top). The XGBoost model demonstrates effective performance, with $R^2 > 0.5$ recorded for predictions less than 4 months in advance (Figure 5, green line). For longer leads, the performance decreases. Our model outperforms the persistence and seasonality benchmark models for all three livelihood zones, especially in agro-pastoral and pastoral regions. However, the performance of these baseline models is significantly better in crop-farming regions, which may be related to the low variance in food security in such regions compared to agro-pastoral or pastoral regions. This low variance increases the relative capabilities of the persistence model (Figure 5, black line) for short leads and of the seasonality model (Figure 5, magenta line) for long leads. Note that the R^2 of the seasonality model is negative for the agro-pastoral and pastoral regions, which indicates that FEWS IPC seasonality does not generate any prediction skill. This outcome suggests that the

FEWS IPC dynamics in the train dataset (2009–2020) are notably different than those in the test dataset (2020–2022) and emphasizes the need for more sophisticated predictions.

Subsequently, we compared the XGBoost forecasts to the FEWS NET food-security outlooks. These cutting-edge outlooks are produced by the FEWS NET early warning system and are widely used. The assessment suggests that the XGBoost model has similar performance to the FEWS NET outlooks in the agro-pastoral and pastoral regions. Compared to the FEWS NET outlooks, the skill of the XGBoost predictions seems to reduce less quickly for longer leads. However, for crop-farming regions, the FEWS NET outlooks perform considerably better, particularly for lead times longer than 3 months, where the R^2 of the XGBoost forecasts drops to below 0.4, but the FEWS NET outlooks remain between 0.5 and 0.65.

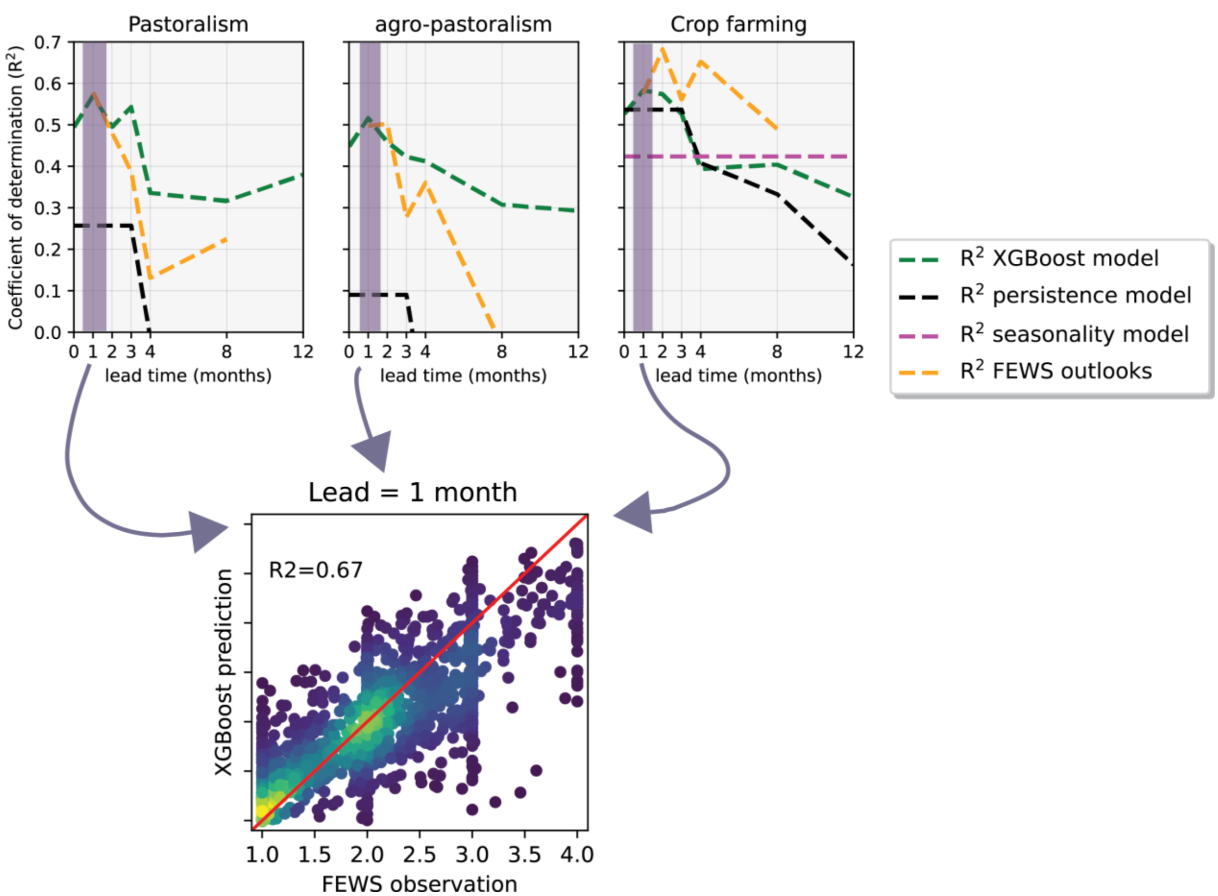

Figure 5 Top: The coefficient of determination (R^2) over lead time for the different livelihood zones are compared to the FEWS outlooks, the persistence model, and the seasonality model. R^2 values <0 are not shown (seasonality model in pastoralism and agro-pastoralism areas). Bottom: Individual predictions and observations are visualized as a scatter plot for a lead time of 1 month. Individual dots represent a prediction or observation for one of the 196 administrative units. Colors represent the density of dots (Gaussian kernel-density estimate).

3.2 Crisis-onset predictions

The above results indicate that general trends in food-security dynamics are captured by the XGBoost model. To use such a model as an early warning system, it is crucial to know to what extent the onset of a food crisis (FEWS IPC > = 3) can be predicted. Therefore, we calculated the hit rate (number of predicted crisis onsets/total number of crisis onsets) and the false-alarm rate (number of false alarms/number of no crisis onsets). Results are shown in Figure 6. We predicted 20% of the food-crisis onsets for pastoral regions (out of 84 onsets in total) and 30–40% for agro-pastoral regions (out of 23 onsets in total) using predictions with a 3-month lead time (Figure 6, top). Predictions for these regions showed a low number of false alarms (approximately 4%). Over all lead times, food-crisis onsets in Somalia were consistently predicted more effectively, followed by those in Ethiopia, while crises in Kenya were nearly never predicted (see Figure 6, bottom left, for a lead time of 3 months). We predicted >25% of all food-crisis onsets in Somalia 3 months in advance with a very low number of false alarms. Predictions with a lead time of 4 months generally showed low performance. The significant drop in skill at 4 months is probably caused by the fact that the FEWS IPC observation of the previous timestep is no longer available to the model 4 months in advance. This is also illustrated in Figure S3 by the lower SHAP values of the "FEWS_CS_lag1" variable for lead 4. Notably, there is an interesting increase at lead times of 8 and 12 months for agro-pastoral regions. This is possibly related to the stronger influence of climate teleconnections in the model predictions at these lead times (Figs. 7 and S3). The model did not predict the start of a food crisis in the crop-farming regions (33 in total). A potential explanation for this is that certain drivers for these regions could not be included. For example, we included proxies for crop yield (e.g., cropland NDVI) but did not have access to crop-yield data with acceptable accuracy.

The results presented in Figure 6 imply that the ability to detect food crises of our XGBoost model is similar to the FEWS NET outlooks for agro-pastoral and pastoral livelihood regions. However, the FEWS NET outlooks clearly outperform our model in crop-farming regions (hit rate > 20% for lead times < 3 months).

This difference in predictive power potentially relates to variations in the input data and dynamic forecast models used in FEWS NET. Specifically, FEWS NET utilizes the G20 Group on Earth Observations Global Agricultural Monitoring (GEOGLAM) crop monitor to generate their outlooks (Funk, Shukla, et al., 2019) as well as crop yield-predictions from the NHyFAS system (Shukla et al., 2020). Additionally, seasonal weather forecasts from NOAA and ICPAC are used in their outlooks. This data is not used in our model. Although the model evaluation is based on many crisis onsets, it reflects a relatively short period (2020–2022). Therefore, these results are indicative and do not have to reflect the true accuracy of the FEWS NET outlooks or the XGBoost model results.

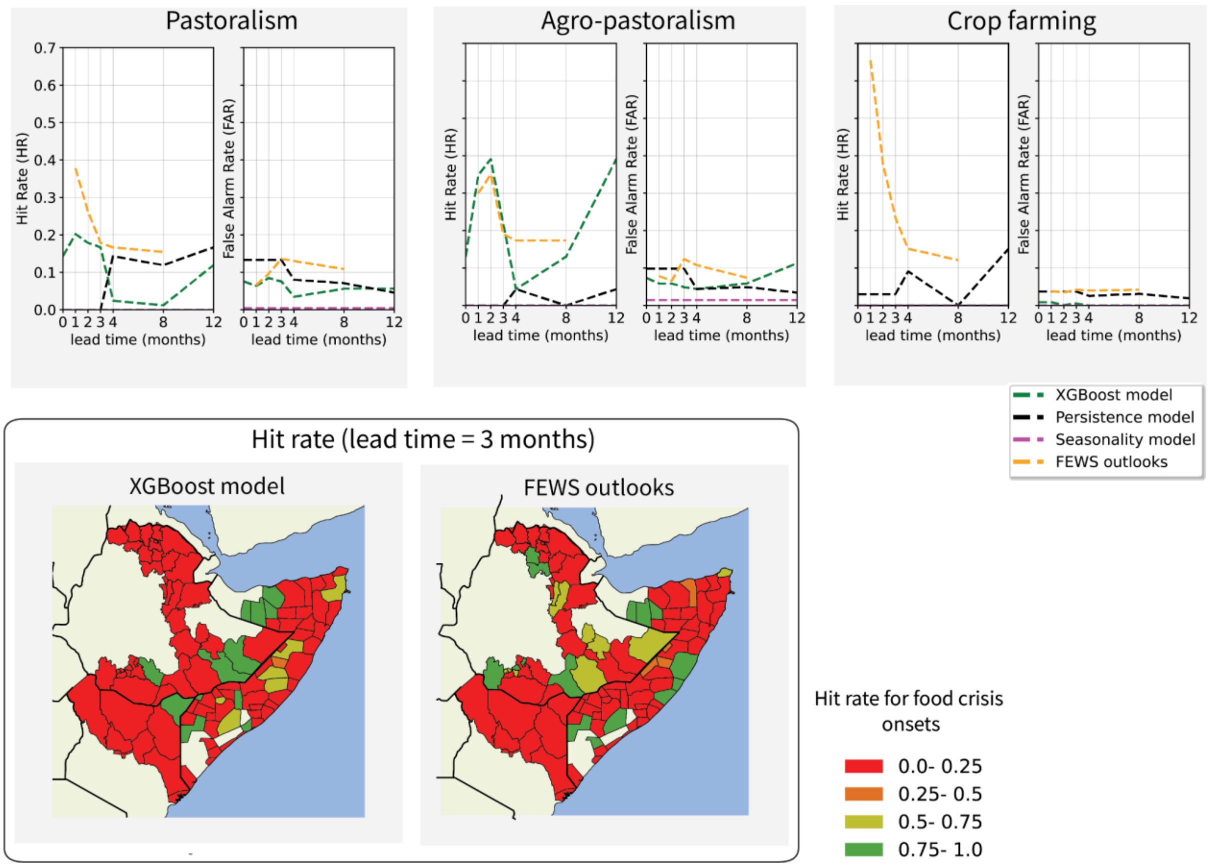

Figure 6. Top: The hit rate and false-alarm rate for the detection of crisis onsets (FEWS IPC > = 3) are represented over multiple lead times for pastoralism (left), agro-pastoralism (middle), and crop-farming (right) livelihoods. Bottom: A map of the hit rates for crisis onsets on lead 3 for the XGBoost model (left) and FEWS NET outlooks (right) is presented. Administrative units without any crisis observations in the test dataset are masked. Metrics are based on the test period (2020–2022).

3.3 Drivers of food insecurity

We assessed the impact of each feature on the predictions using the SHAP framework (Lundberg & Lee, 2017) for lead times of 3 and 8 months. To obtain a realistic insight into potential food-insecurity drivers, we only show the 30% best-performing administrative units (58 in total).

Food security is a multi-hazard impact (Boult et al., 2022), and it is well known that food security can have different potential drivers for different lead times (WMO, 2017). Therefore, we compared the importance of features in our models at different lead times. This resulted in a clear pattern (Figure 7). The previous FEWS NET food-security observation is highly important for predicting the next food-security status for short lead times (Figure 7, left). However, this importance decreases at longer leads (Figure 7, right), although it does not disappear entirely. Climate and weather variables—rainfall, evaporation (SPEI), and soil moisture (SSMI)—are relatively more important for predictability on short lead times (Figure 7, left). In contrast, the predictability of long lead times (Figure 7, right) mostly originates from remote climate teleconnections, such as Nino 3.4 and the WVG, as well as vulnerability indicators, such as the GDP.

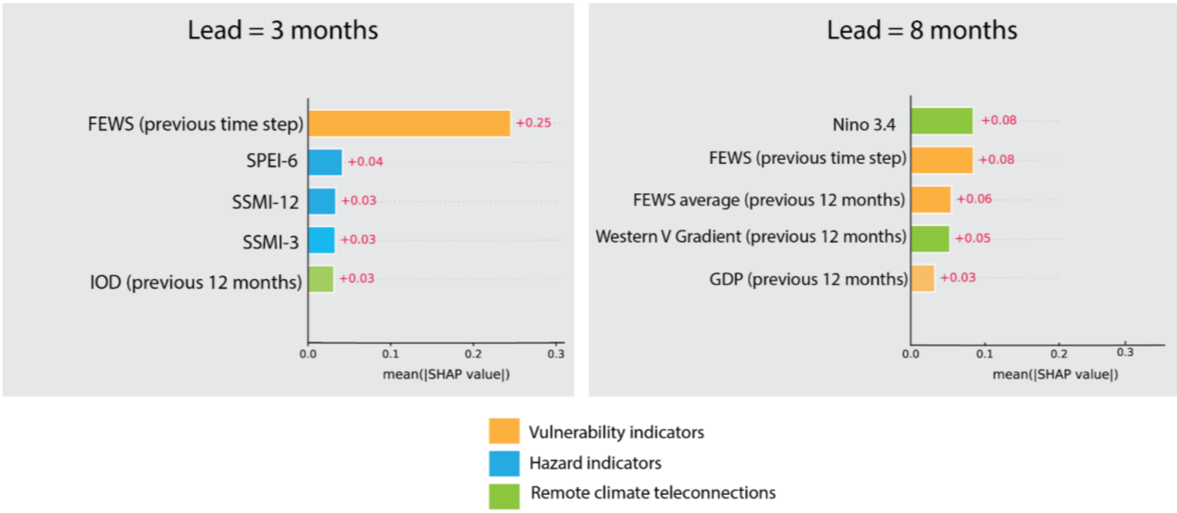

Figure 7. The top five most important features for a lead time of 3 months (left) and 8 months (right) are measured as the mean of the absolute SHAP values. Colors illustrate the different data categories (hazard, vulnerability, and remote climate teleconnections) in accordance with the colors in Table 1 and the framework in Figure 1.

To understand the underlying model interactions, we further examined the relationship between model features and food insecurity. Figure 8 illustrates the impact of the 10 most important features in the model predictions on a 3-month lead time. Nearly all the important features have long accumulation periods: for example, SPI-6 and SSMI-12 use historical data over the last 6 and 12 months, respectively, to generate a forecast. This indicates that food security is mostly influenced by longer and more persistent drought conditions. Interestingly, the only variable that did not show these long accumulation periods was maize prices, which indicates that price spikes over a short period already had a strong effect on food security.

Apart from previous FEWS NET food-security states and maize prices, all selected features have a negative statistical relationship with food insecurity: lower feature values (blue dots, Figure 8) lead to increases in food insecurity (i.e., a positive impact on model output). This is to be expected, as drought (indicated by lower values of the SPI, SPEI, SSMI drought indices) can exacerbate or even trigger food insecurity (Funk, Shukla, et al., 2019). We found that the number of wet days per month is the most important rainfall indicator in the model, outnumbering SPI and the number of dry spells. This suggests that food security has a stronger link with rainfall distribution over the month than with absolute rainfall amounts.

The remote climate teleconnections also show an expected positive relationship with food insecurity: lower indices of Nino 3.4, WVG, IOD, or MEI are all drawing the East African climate towards a dryer state (Funk et al., 2023; Funk, Pedreros, et al., 2019). Negative IOD values imply relatively lower SSTs in the Western Indian Ocean, which result in less evaporating moisture transported into East Africa. Negative MEI and Nino 3.4 imply a La Niña with overall cooler SSTs in the Eastern Pacific and warmer SSTs in the Western Pacific (Figure S1). The

WVG reflects a warm blob in the Pacific Ocean around Indonesia and the Philippines (Figure S1), which have recently been found to be connected to the East African climate on long leads (Funk et al., 2023; Funk, Pedreros, et al., 2019). Our model results are consistent with these findings and suggest that the WVG is also important for such impact-based forecasts, especially on long leads (Figure 7, right). The SHAP values found for the other lead times can be seen in Figure S3. These results demonstrate that the underlying model interactions are physically understandable and reflect intuition and the newest research insights.

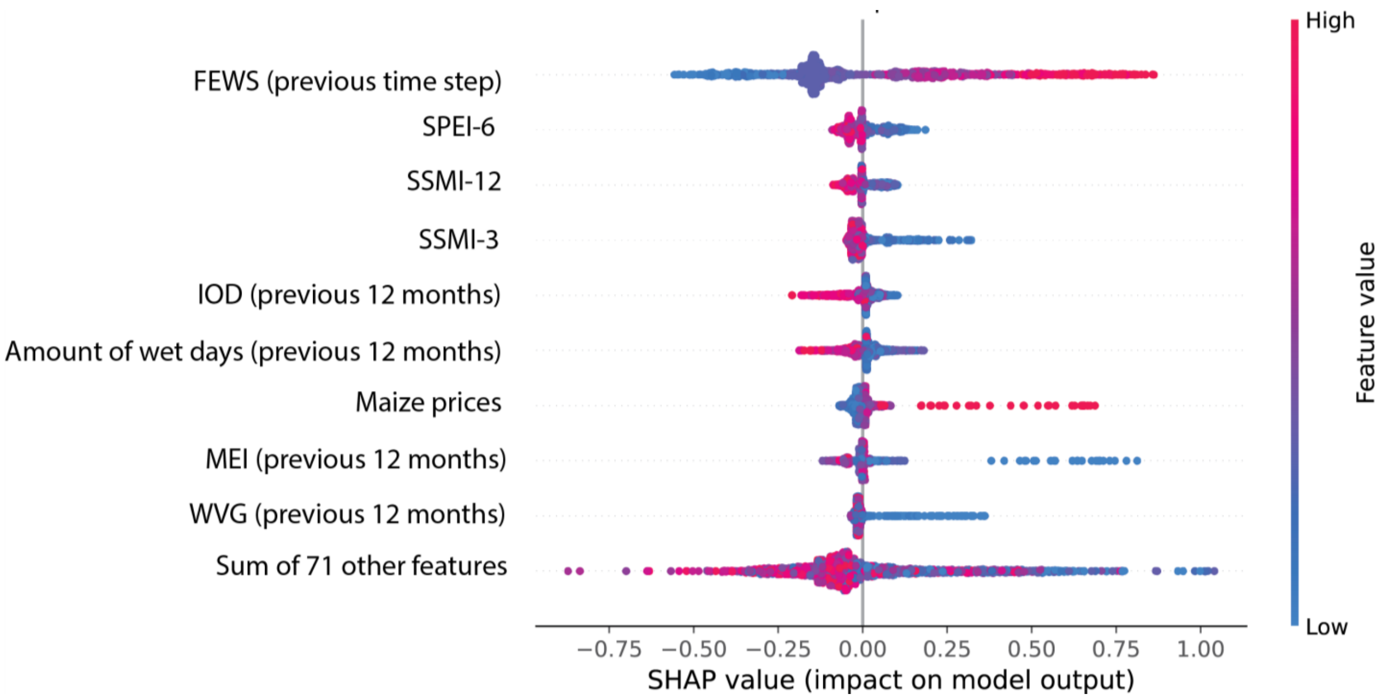

Figure 8. The top 10 most important features for model predictions with a lead time of 3 months, as indicated by the SHAP values for each feature's contribution. Each dot represents a prediction from the model, with the color indicating the value of the feature and the x-axis signifying whether the feature increased (positive SHAP values) or decreased (negative SHAP values) the model prediction of food insecurity.

4. Discussion

4.1 Adding value to FEWS NET

This study indicates that machine-learning models can have similar performance to operational early warning systems such as FEWS NET in some contexts. However, in crop-farming regions, the FEWS NET outlooks clearly outperformed the XGBoost model. This indicates that machine-learning models like the one presented here can complement the existing FEWS NET outlooks. The FEWS NET outlooks are based on a scenario-development process and expert judgments, while our approach is data-driven, enabling a fully transparent generation of early warnings. Another benefit is that potentially a higher initiation frequency can be achieved. Many of our utilized datasets are regularly updated, such as ACLED, which receives weekly updates, and the WPF VAM portal, which is updated biweekly. In practice, this could result in an increased frequency of outlook releases and, therefore, more timely early warnings. In addition to complementing drought early warning systems, explainable AI techniques can reveal new food-security drivers for certain regions or lead times. These insights may then contribute to the scenario-development process for the FEWS NET food-security outlooks. A greater understanding of the drivers can also lead to more informed decisions on the ground and help tailor emergency-response planning. However, by no means can a machine-learning model ever substitute early warning systems like FEWS NET. On the ground expert judgement, local knowledge, and field experience are crucial to the co-production of early warning systems (ICPAC, 2021).

4.2 Machine-learning architecture: key considerations

The model results show that the prediction of acute food insecurity is complex, with many drivers that may contribute to changes in food-security status. Many different machine learning algorithms can be used to capture these drivers. We favored the use of a tree-based model over a neural-network model, as we did not expect strong multi-temporal or multi-spatial dependencies often found in other domains, such as image recognition (Fujiyoshi et al., 2019) or storm-surge modeling (Tiggeloven et al., 2021). Moreover, tree-based models often outperform neural networks on tabular data where features are individually meaningful (Lundberg et al., 2020).

We found that the XGBoost model was more capable of capturing these complex interactions than other tree-based models. We tested our results against random forest models, and the XGBoost model yielded a slightly higher performance, especially for crisis-onset predictions (see Figure S2). This is consistent with other studies. Westerveld et al. (2021) showed that the XGBoost model is superior to other machine-learning models for predicting transitions in food crises.

The feature engineering, which adds features based on existing ones, resulted in a total of 81 different features for the model. Some of these features are related (e.g., total precipitation over the last 4 and last 12 months). Feature selection—the reduction of features based on their (expected) performance—can help address multicollinearity (Chan et al., 2022), which is the correlation between predictor variables. While it can reduce multicollinearity, we opted not to perform feature selection in our study. This decision was influenced by XGBoost's robustness in handling multicollinearity. Additionally, we now retain all features in our 21 models, each tailored to different livelihood zones and lead times. Removing features risked losing vital insights specific to each model's context.

Before we pooled the data of the individual administrative units together (based on the three livelihood types), we tested many other spatial scales for model training. The levels involved were as follows: administrative units (196 individual models for each lead time), livelihood zones within all three countries, the three countries, and last, all data for the 196 administrative units pooled together. We found that increasing the level of spatial pooling significantly improved model performance. This can be explained by the limited period 2009–2019 used for training, which, within one administrative unit, simply does not contain sufficient data to learn the actual drivers of food insecurity. Pooling based on livelihood zones over the whole region performed most effectively, slightly better than pooling all data from every administrative unit together. Although the performance gain was small, making it difficult to draw definite conclusions, this could suggest that different livelihoods are influenced by distinct food-security drivers.

4.3 Limitations of the study

Although this study shows the potential of machine-learning systems for food-security early warning, the data could only be tested for a short period (2020–2022). This brings uncertainty to the verification of the predictions, especially for crisis-onset events. Nonetheless, we evaluated enough crisis onsets—140 in total—to conclude that considerable skill exists.

We created an extensive dataset with food-security drivers from hazard, vulnerability, and remote-climate teleconnections. Although this is a holistic and extensive dataset, certain important drivers could not be included. We did not have access to crop-yield data on a monthly time scale for all the administrative units. Its exclusion may be a reason why the food-security crises could not be predicted effectively in the crop-farming regions (whereas FEWS NET outlooks do predict these crises well). The performance of the model in detecting food crises in Kenya is highly limited (Figure 6), partly because maize prices—an important feature in our model—were not available after 2020 for that country. Thus, although this feature can be used to train the model, it was missing during the model evaluation in the test set (2020–2022). This issue emphasizes the importance of continuous data collection and archiving efforts.

The test period of the study, 2020–2022, also reflects a turbulent period that encompassed the COVID-19 pandemic and the start of the Russian invasion of Ukraine. Both events have had significant economic effects worldwide, with economic fallouts and price raises that in turn have resulted in record levels of malnutrition and food insecurity (WFP, 2022). Because overseas conflicts or health crises are not directly included in the model, it can only indirectly capture these dynamics by incorporating price data.

Although overseas conflicts, such as the Russian invasion of Ukraine, are not directly included in the model, local conflicts are incorporated into the ACLED dataset. However, local conflict was not identified as an important variable in our model, as shown by its absence in the SHAP plots (Figs. 8 and S3). This is surprising, as research and practice show that conflict drives hunger (WFP, 2022). As such, we found that this missing link between conflict and hunger reduced the accuracy of our predictions. We failed to predict the rapid increase in food insecurity in the

Tigray districts after the armed conflict erupted in November 2020 (Weldegiargis et al., 2023), which returned high error scores over this region (MAE, Figure 4). This failure may be explained by the fact that the Tigray conflict was the largest recorded in the ACLED conflict dataset over the 2009–2022 period, and therefore, the model could not train on conflicts of this magnitude. Moreover, some large conflicts in our dataset (e.g., the Mogadishu bombings in October 2017) were followed by a decrease in food insecurity rather than an increase. Conflict has consistently emerged as a complex factor in food-insecurity early warning, as highlighted in prior studies (e.g. Krishnamurthy et al., 2020).

5. Conclusions and recommendations

In this study, we developed an XGBoost machine-learning model to predict food security on monthly timescales over the Horn of Africa. We trained the model on the 2009–2020 period using >20 different datasets and the FEWS IPC current situation as ground truth. Our model predicted 20% of crisis onsets in pastoral livelihood regions (n = 84) and 40% of crisis onsets in agro-pastoral livelihood regions (n = 23) several months in advance with low overall numbers of false alarms. Furthermore, the model predicted general food-security patterns up to 3 months in advance ($R^2 > 0.6$). This underscores the potential of such machine-learning models to complement existing early warning systems, such as FEWS NET.

FEWS NET builds on decades of experience in food-security monitoring and early warning, and the FEWS NET food-security outlooks are widely adopted. To serve as an ultimate benchmark, we compared our predictions with these FEWS NET outlooks over the 2020–2022 period. Results suggest the performance of the XGBoost model is similar to the FEWS NET outlooks for agro-pastoral and pastoral regions. However, the FEWS NET outlooks clearly outperform our model for crop-farming regions. Moreover, machine-learning models need sufficient training data, which limited the data available during the test period (2020–2022). This decreased the robustness of model performance estimates. Thanks to continuing monitoring efforts from FEWS NET, more data will be available in the future to train and test such machine-learning models. This will increase the robustness of the model evaluation and allow for longer training periods, which will likely further improve model performance in the future.

This study shows the potential of food-security predictions made with machine learning to complement existing early warning systems, such as FEWS NET, by allowing more frequent updates and revealing specific drivers in particular regions. Future research could further explore how such machine-learning models can be improved. We expect that the inclusion of dynamical forecasts as features in the machine-learning model will lead to a significant improvement. Soil moisture and yield forecasts (Shukla et al., 2020) can lead to better predictions in crop-farming regions. Moreover, decision-makers can use results from this study to better understand the drivers of food-security crises at different lead times, which may lead to more informed and timely interventions. The organizations operating and developing food security early warning systems can use our results to envision and shape hybrid solutions where a part is automated and based on machine learning, but also a part remains consensus-based.

Acknowledgements

This project was supported by the Horizon 2020 DOWN2EARTH project [grant agreement ID: 869550] and the Horizon 2020 COASTMOVE ERC advanced grant [grant nr. 884442]. Our gratitude goes out to FEWS NET for developing and maintaining an extensive historical archive of food security estimates and projections. We recognize the BAZIS service for the use of the high performance computing cluster. Lastly, we want to acknowledge Brenda Lazarus (FAO) for reviewing the livelihood zone map, using field knowledge and observations.

Open Research

The input data used to run the food-security machine learning model in the study are available at Zenodo via https://doi.org/10.5281/zenodo.10013551 with the Creative Commons Attribution 4.0 International license (Busker, 2023b).

Version 1.0.0 of the machine learning model used to generate the food security predictions is preserved at https://doi.org/10.5281/zenodo.10013666 , available with the Creative Commons Attribution 4.0 International license (Busker, 2023a).

References

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

Baxter, A. J., Verschuren, D., Peterse, F., Miralles, D. G., Martin-Jones, C. M., Maitituerdi, A., et al. (2023). Reversed Holocene temperature–moisture relationship in the Horn of Africa. *Nature*, *620*(7973), 336–343. https://doi.org/10.1038/s41586-023-06272-5

Blauhut, V., Stahl, K., Stagge, J. H., Tallaksen, L. M., De Stefano, L., & Vogt, J. (2016). Estimating drought risk across Europe from reported drought impacts, drought indices, and vulnerability factors. *Hydrology and Earth System Sciences*, *20*(7), 2779–2800. https://doi.org/10.5194/hess-20-2779-2016

Boult, V. L., Asfaw, D. T., Young, M., Maidment, R., Mwangi, E., Ambani, M., et al. (2020). Evaluation and validation of TAMSAT-ALERT soil moisture and WRSI for use in drought anticipatory action. *Meteorological Applications*, *27*(5), 27. https://doi.org/10.1002/met.1959

Boult, V. L., Black, E., Saado Abdillahi, H., Bailey, M., Harris, C., Kilavi, M., et al. (2022). Towards drought impact-based forecasting in a multi-hazard context. *Climate Risk Management*, *35*, 100402. https://doi.org/10.1016/J.CRM.2022.100402

Busker, T. (2023a). Model scripts used for the paper "Predicting Food-Security Crises in the Horn of Africa Using Machine Learning" (Version v1.0.0) [Software]. Zenodo. https://doi.org/10.5281/zenodo.10013666

Busker, T. (2023b). Replication Data for: "Predicting Food-Security Crises in the Horn of Africa Using Machine Learning" (Version v1.0.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.10013551

Cerqueira, V., Torgo, L., & Mozetič, I. (2020). Evaluating time series forecasting models: an empirical study on performance estimation methods. *Machine Learning*, *109*(11), 1997–2028. https://doi.org/10.1007/s10994-020-05910-7

Chan, J. Y.-L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z.-W., & Chen, Y.-L. (2022). Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. *Mathematics*, *10*(8), 1283. https://doi.org/10.3390/math10081283

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). San Francisco California USA: ACM. https://doi.org/10.1145/2939672.2939785

Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. (2016). Accurate prediction of sugarcane yield using a random forest algorithm. *Agronomy for Sustainable Development*, *36*(2), 27. https://doi.org/10.1007/s13593-016-0364-z

FAO. (2009). Declaration of the World Summit on Food Security (WSFS 2009/2). Retrieved June 15, 2023, from https://reliefweb.int/report/world/declaration-world-summit-food-security-wsfs-20092

FAO. (2022). FAO Locust Hub | Swarms. Retrieved June 29, 2023, from https://locust-hub-hqfao.hub.arcgis.com/datasets/hqfao::swarms-1/explore

FAO. (2023a). GIEWS - Global Information and Early Warning System - Country Briefs. Retrieved June 29, 2023, from https://www.fao.org/giews/countrybrief/index.jsp

FAO. (2023b). GIEWS - Global Information and Early Warning System on Food and Agriculture. Retrieved June 13, 2023, from https://www.fao.org/giews/en/

FEWS NET. (2018). Scenario Development for Food Security Early Warning. Retrieved August 8, 2023, from https://fews.net/sites/default/files/documents/reports/Guidance_Document_Scenario_Development_2018.pdf

FEWS NET. (2023a). FEWS NET Data Explorer. Retrieved June 22, 2023, from https://fdw.fews.net/en/

FEWS NET. (2023b). Monitoring & forecasting acute food insecurity. Retrieved June 29, 2023, from https://fews.net/

Foini, P., Tizzoni, M., Martini, G., Paolotti, D., & Omodei, E. (2023). On the forecastability of food insecurity. *Scientific Reports*, *13*(1), 2793. https://doi.org/10.1038/s41598-023-29700-y

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, *29*(5), 1189–1232.

Fujiyoshi, H., Hirakawa, T., & Yamashita, T. (2019). Deep learning-based image recognition for autonomous driving. *IATSS Research*, *43*(4), 244–252. https://doi.org/10.1016/j.iatssr.2019.11.008

Funk, C., & Shukla, S. (Eds.). (2020). Drought Early Warning and Forecasting - Theory and Practice (pp. 215–222). Elsevier. https://doi.org/10.1016/B978-0-12-814011-6.00020-8

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., et al. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. *Scientific Data 2015 2:1*, *2*(1), 1–21. https://doi.org/10.1038/SDATA.2015.66

Funk, C., Pedreros, D., Nicholson, S., Hoell, A., Korecha, D., Galu, G., et al. (2019). Examining the Potential Contributions of Extreme "Western V" Sea Surface Temperatures to the 2017 March–June East African Drought. *Bulletin of the American Meteorological Society*, *100*(1), S55–S60. https://doi.org/10.1175/BAMS-D-18-0108.1

Funk, C., Shukla, S., Thiaw, W. M., Rowland, J., Hoell, A., McNally, A., 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), 1011–1027. https://doi.org/10.1175/BAMS-D-17-0233.1

Funk, C., Harrison, L., Segele, Z., Rosenstock, T., Steward, P., Anderson, C. L., et al. (2023). Tailored Forecasts Can Predict Extreme Climate Informing Proactive Interventions in East Africa. *Earth's Future*, *11*(7), e2023EF003524. https://doi.org/10.1029/2023EF003524

Guimarães Nobre, G., Davenport, F., Bischiniotis, K., Veldkamp, T., Jongman, B., Funk, C. C., et al. (2019). Financing agricultural drought risk through ex-ante cash transfers. *Science of The Total Environment*, *653*, 523–535. https://doi.org/10.1016/j.scitotenv.2018.10.406

Ha, J., Kose, M. A., & Ohnsorge, F. (2021). One-Stop Source: A Global Database of Inflation. *Policy Research Working Paper 9737*.

Hao, Z., AghaKouchak, A., Nakhjiri, N., & Farahmand, A. (2014). Global integrated drought monitoring and prediction system. *Scientific Data*, *1*(1), 140001. https://doi.org/10.1038/sdata.2014.1

ICPAC. (2021). Co-production in Climate Services: What, Why, and How? Retrieved September 20, 2023, from https://www.icpac.net/news/co-production-in-climate-services-what-why-and-how/

IMF. (2023). World Economic Outlook (April 2023) [Data set]. Retrieved from https://www.imf.org/external/datamapper/datasets/WEO

IPC. (2021). *IPC Technical Manual Version 3.1*. Rome. Retrieved from http://www.ipcinfo.org/ipc-manual-interactive/en/

IPC. (2023). IPC Overview and Classification System | IPC - Integrated Food Security Phase Classification. Retrieved from https://www.ipcinfo.org/ipcinfo-website/ipc-overview-and-classification-system/en/

Kimutai, J., Barnes, C., Zachariah, M., Philip, S., Kew, S., Pinto, I., et al. (2023). *Human-induced climate change increased drought severity in Horn of Africa*. Imperial College London. https://doi.org/10.25561/103482

Krishnamurthy, P. K., Choularton, R. J., & Kareiva, P. (2020). 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

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions (Version 2). https://doi.org/10.48550/ARXIV.1705.07874

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence 2020 2:1*, *2*(1), 56–67. https://doi.org/10.1038/s42256-019-0138-9

Martens, B., Miralles, D. G., Lievens, H., van der Schalie, R., de Jeu, R. A. M., Fernández-Prieto, D., et al. (2017). GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017

Martini, G., Bracci, A., Riches, L., Jaiswal, S., Corea, M., Rivers, J., et al. (2022). Machine learning can guide food security efforts when primary data are not available. *Nature Food*, *3*(9), 716–728. https://doi.org/10.1038/s43016-022-00587-8

McGovern, A., Lagerquist, R., Gagne, D. J., Jergensen, G. E., Elmore, K. L., Homeyer, C. R., & Smith, T. (2019). Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning. *Bulletin of the American Meteorological Society*, *100*(11), 2175–2199. https://doi.org/10.1175/BAMS-D-18-0195.1

McKee, T. B., Doesken, N. J., & Kleist, J. (1993). The relationship of drought frequency and duration to time scales. *Proceedings of the 8th Conference on Applied Climatology*, *17*(22).

NBS. (2023). Somali National Bureau of Statistics [Data set]. Retrieved from https://www.nbs.gov.so/home

NOAA. (2021). STAR - Global Vegetation Health Products : Downloading Vegetation Health Products Data. Retrieved August 26, 2022, from https://www.star.nesdis.noaa.gov/smcd/emb/vci/VH/vh_ftp.php

NOAA. (2023a). Climate Indices: Monthly Atmospheric and Ocean Time Series [Data set]. Retrieved from https://psl.noaa.gov/data/climateindices/list/

NOAA. (2023b). Download Climate Timeseries - Dipole Mode Index (DMI) [Data set]. Retrieved from https://psl.noaa.gov/gcos_wgsp/Timeseries/DMI/

Odongo, R. A., De Moel, H., & Van Loon, A. F. (2023). Propagation from meteorological to hydrological drought in the Horn of Africa using both standardized and threshold-based indices. *Natural Hazards and Earth System Sciences*, *23*(6), 2365–2386. https://doi.org/10.5194/nhess-23-2365-2023

Pérez-Hoyos, A. (2018). Global crop and rangeland masks. European Commission, Joint Research Centre (JRC) [Data set]. Retrieved from http://data.europa.eu/89h/jrc-10112-10005

Raleigh, C., Linke, rew, Hegre, H., & Karlsen, J. (2010). Introducing ACLED: An Armed Conflict Location and Event Dataset. *Journal of Peace Research*, *47*(5), 651–660. https://doi.org/10.1177/0022343310378914

Schoppa, L., Disse, M., & Bachmair, S. (2020). Evaluating the performance of random forest for large-scale flood discharge simulation. *Journal of Hydrology*, *590*, 125531. https://doi.org/10.1016/j.jhydrol.2020.125531

Shapley, L. S. (1953). A Value for n-Person Games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the Theory of Games (AM-28), Volume II* (pp. 307–318). Princeton University Press. https://doi.org/10.1515/9781400881970-018

Shukla, S., McNally, A., Husak, G., & Funk, C. (2014). A seasonal agricultural drought forecast system for food-insecure regions of East Africa. *Hydrology and Earth System Sciences*. https://doi.org/10.5194/hess-18-3907-2014

Shukla, S., R. Arsenault, K., Hazra, A., Peters-Lidard, C., D. Koster, R., Davenport, F., et al. (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), 1187–1201. https://doi.org/10.5194/NHESS-20-1187-2020

Snijders, T. A. B. (1988). On Cross-Validation for Predictor Evaluation in Time Series. In T. K. Dijkstra (Ed.), *On Model Uncertainty and its Statistical Implications* (Vol. 307, pp. 56–69). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-61564-1_4

Stagge, J. H., Kohn, I., Tallaksen, L. M., & Stahl, K. (2015). Modeling drought impact occurrence based on meteorological drought indices in Europe. *Journal of Hydrology*, *530*, 37–50. https://doi.org/10.1016/j.jhydrol.2015.09.039

Tiggeloven, T., Couasnon, A., van Straaten, C., Muis, S., & Ward, P. J. (2021). Exploring deep learning capabilities for surge predictions in coastal areas. *Scientific Reports 2021 11:1*, *11*(1), 1–15. https://doi.org/10.1038/s41598-021-96674-0

UNHCR. (2023). Horn of Africa Food Crisis Explained. Retrieved June 13, 2023, from https://www.unrefugees.org/news/horn-of-africa-food-crisis-explained/#1.%20Why%20is%20there%20food%20insecurity%20in%20the%20Horn%20of%20Africa?

Vicente-Serrano, S. M., Beguería, S., & López-Moreno, J. I. (2010). A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index. *Journal of Climate*, *23*(7), 1696–1718. https://doi.org/10.1175/2009JCLI2909.1

Wang, D., Andree, B. P. J., Chamorro, A. F., & Girouard Spencer, P. (2020). *Stochastic Modeling of Food Insecurity*. World Bank, Washington, DC. https://doi.org/10.1596/1813-9450-9413

Weldegiargis, A. W., Abebe, H. T., Abraha, H. E., Abrha, M. M., Tesfay, T. B., Belay, R. E., et al. (2023). Armed conflict and household food insecurity: evidence from war-torn Tigray, Ethiopia. *Conflict and Health*, *17*(1), 22. https://doi.org/10.1186/s13031-023-00520-1

Westerveld, J. J. L., van den Homberg, M. J. C., Nobre, G. 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*, 147366. https://doi.org/10.1016/J.SCITOTENV.2021.147366

WFP. (2014). Technical Guidance Note: Calculation and Use of the Alert for Price Spikes (ALPS) Indicator. Retrieved July 20, 2023, from https://documents.wfp.org/stellent/groups/public/documents/manual_guide_proced/wfp264186.pdf

WFP. (2020). Global Report on Food Crises. Retrieved June 28, 2023, from https://www.wfp.org/publications/2020-global-report-food-crises

WFP. (2022). War in Ukraine drives global food crisis | World Food Programme. Retrieved August 14, 2023, from https://www.wfp.org/publications/war-ukraine-drives-global-food-crisis

WFP. (2023a). Export Data - Prices. Retrieved July 22, 2023, from https://dataviz.vam.wfp.org/version2/economic/export-data/prices

WFP. (2023b). Scaling up anticipatory actions for food security: Anticipatory Action Year in Focus 2022. Retrieved September 20, 2023, from https://www.wfp.org/publications/scaling-anticipatory-actions-food-security-anticipatory-action-year-focus-2022

WFP and FAO. (2022). Hunger Hotspots FAO-WFP early warnings on acute food insecurity: October 2022 to January 2023 Outlook. Retrieved from https://docs.wfp.org/api/documents/WFP-0000142656/download/

WMO. (2017). How climate forecasts strengthen food security. Retrieved September 29, 2023, from https://public.wmo.int/en/resources/bulletin/how-climate-forecasts-strengthen-food-security

WMO. (2022). Meteorological and humanitarian agencies sound alert on East Africa. Retrieved August 2, 2022, from https://public.wmo.int/en/media/news/meteorological-and-humanitarian-agencies-sound-alert-east-africa

Zhang, B., Abu Salem, F. K., Hayes, M. J., Smith, K. H., Tadesse, T., & Wardlow, B. D. (2023). Explainable machine learning for the prediction and assessment of complex drought impacts. *Science of The Total Environment*, *898*, 165509. https://doi.org/10.1016/j.scitotenv.2023.165509

Zheng, A., & Casari, A. (2018). *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*. O'Reilly Media, Inc.