Check for updates

Machine learning can guide food security efforts when primary data are not available

Giulia Martini [1], Alberto Bracci [1,2], Lorenzo Riches[1], Sejal Jaiswal[1], Matteo Corea[1], Jonathan Rivers[1], Arif Husain[1] and Elisa Omodei [1,3] ✉

Estimating how many people are food insecure and where they are is of fundamental importance for governments and humanitarian organizations to make informed and timely decisions on relevant policies and programmes. In this study, we propose a machine learning approach to predict the prevalence of people with insufficient food consumption and of people using crisis or above-crisis food-based coping when primary data are not available. Making use of a unique global dataset, the proposed models can explain up to 81% of the variation in insufficient food consumption and up to 73% of the variation in crisis or above food-based coping levels. We also show that the proposed models can nowcast the food security situation in near real time and propose a method to identify which variables are driving the changes observed in predicted trends—which is key to make predictions serviceable to decision-makers.

C haracterizing the socio-economic status of populations and providing reliable and up-to-date estimates of who the most vulnerable are, how many they are, where they live and why they are vulnerable is essential for governments and humanitarian organizations to make informed and timely decisions on implementation of humanitarian assistance policies and programmes[1]. These data are traditionally collected through face-to-face surveys. However, these are expensive, time-consuming and, in certain areas, not possible to perform due to conflict, disease, insecurity or remoteness. Therefore, during the past few years, researchers have begun to investigate the potential of non-traditional data and new computational methods to estimate vulnerabilities and socio-economic characteristics when primary data are not available. In these studies, mobile phone data[2], satellite imagery[3], a combination of both[4,5], mobile money transaction records[6], geolocated Wikipedia articles[7] or tweets[8] and social media advertising data[9] have been used in combination with state-of-the-art machine learning methods to provide reliable estimates of poverty at different spatial resolutions for several sub-Saharan African countries and southern and southeastern Asian ones.

The methods proposed in these studies provide a unique opportunity to monitor poverty in near real time on a global scale. In this work, we show that similar methods and data can be used to tackle another outstanding form of vulnerability affecting populations worldwide: food insecurity. In 2019, the number of under-nourished people was estimated to be 650 million[10], with 135 million in 55 countries and territories reported to be acutely food insecure[11]. These numbers have substantially increased as a consequence of the COVID-19 pandemic, with at least 280 million people reported to be acutely food insecure in 2020, more than doubling the number from the previous year[12]. To address this global issue, monitoring the situation and its evolution is key. Governments and international organizations such as the World Food Programme (WFP), the Food and Agriculture Organization (FAO) and the World Bank perform food security assessments on a regular basis through face-to-face surveys or, increasingly so, through remote mobile phone surveys (for example, computer-assisted telephone interviews) and further

supporting technologies such as interactive voice response and web surveys[1]. However, as mentioned previously, there are limitations with these approaches given their high costs in both monetary and human resources. In addition, food insecurity is a more dynamic and unstable phenomenon than poverty, with a seasonal component related to agricultural production calendars and subject to swift changes when external shocks hit, therefore requiring more frequent and rapid assessments.

Food insecurity is a multi-dimensional concept, spanning from food availability and access to utilization and stability[13]. Multiple indicators have been developed to characterize household food insecurity levels, each capturing different aspects. In this study, we focus on the Food Consumption Score (FCS) and the reduced Coping Strategy Index (rCSI), the former capturing quantity and diversity of dietary intake and the latter the consequences of constrained access to food, resulting in coping behaviours[14]. Aggregating a representative number of household-level measurements of these indicators makes it possible to characterize the food security situation of a geographical area during a specified time window through the prevalence of people with insufficient food consumption and that of people using crisis or above-crisis food-based coping, respectively. In this study, we show that these two metrics can be estimated from secondary data by means of machine learning algorithms when primary data are not available. This opens the door to food security near-real time nowcasting on a global scale, allowing decision-makers to make more timely and informed decisions on policies and programmes oriented towards the fight against hunger. For example, when the proposed models are predicting increases in the prevalence of food-insecure people, then WFP will trigger rapid assessments through face-to-face or remote surveys and mobilize in-country analysts to gain a better understanding of the situation. The development of these models is motivated by a specific need of WFP to fill a gap that currently exists because of limited resources and inaccessibility, that is, to provide regular information for less reachable places, where food security assessments are carried out only once or twice per year but that nevertheless require a constant flow of information to inform humanitarian operations.

[1]World Food Programme, Research, Assessment and Monitoring Division (RAM), Rome, Italy. [2]Department of Mathematics, University of London, London, UK. [3]Department of Network and Data Science, Central European University, Vienna, Austria. ✉e-mail: omodeie@ceu.edu

Previous work has explored the use of secondary data to investigate specific aspects of food insecurity, such as agricultural production. Statistical crop models and climate modelling have been used to make projections for 2030 of changes in crop productions in 12 food-insecure regions due to climate change[15]. Mobile phone records have been used to analyse monthly mobility in Senegal, leading to the discovery and characterization of seasonal mobility profiles related to economic activities, agricultural calendars and precipitation[16]. Other studies have proposed a characterization of the food security situation based on a variety of secondary indicators, addressing its multi-dimensional aspect and providing annual national-level estimates[17,18]. Famine risk prediction through machine learning and stochastic models has also been the subject of recent investigation. Okori and Obua used household socio-economic and agricultural production characteristics to train several machine learning models to predict households' food security status[19]. The limitation of this approach is that up-to-date household-level data are required not only during model training but also when using the trained models to perform out-of-sample predictions. More recently, the World Bank developed a suite of statistical models to forecast transitions into critical states of food insecurity and famine risk from secondary data[20,21]. In this study, we focus on food security nowcasting, proposing a methodology that allows us to estimate the current prevalence of people with insufficient food consumption and of people using crisis or above-crisis food-based coping at the sub-national level at any given time from secondary data, when primary data are not available. Seminal work by Lentz and collaborators addressed this challenge, obtaining predictions that explain up to 65% of the variation in food consumption, although limited to Malawi only[22]. Similar studies in the context of Ethiopia[23] and Burkina Faso[24] were also recently proposed. Here we make use of a unique dataset of sub-national-level food consumption and food-based coping data collected during the past 15 years across, respectively, 78 and 41 countries and aggregated by first-level administrative country sub-divisions (for example, departments, provinces and so on), allowing the development and validation of nowcasting predictive models of food security indicators on a global scale.

Results

Predictions using previously measured levels and secondary data. The main assumption of this study is that when primary data are not available, levels of insufficient food consumption and of crisis or above food-based coping can be estimated from secondary information, specifically on the key drivers of food insecurity. Experts identify three main causes for food insecurity: conflict, economic shocks and extreme weather events[11]. To build the proposed predictive models, we therefore collected historical data covering all three dimensions: data on conflict-related fatalities, economic information (prices of staple food in local markets, headline and food inflation, currency exchange rates and gross domestic product (GDP) per capita) and data on rainfall and vegetation, including anomalies with respect to historical averages. For each available historical measurement of insufficient food consumption and of crisis or above food-based coping for a given geographical area and time window, we associate as independent variables the corresponding conflict, economic and weather situation for the same area in the previous three-month window. Moreover, we also take into account as independent variables under-nourishment and population density and the target prevalence measured during the most-recent previous food security assessment, whose time frame varies across the different areas.

For each target variable, after having defined and selected the input variables as described in Methods, we fitted $N_b = 100$ bootstrapped models using gradient boosted regression trees[25], employing the first (in temporal terms) approximately 85% of the historical

Table 1 | Model performance metrics

		R^2	MAE
Food consumption	**Prevalence from previous assessment included as independent variable**	**0.81**	**0.07**
	Prevalence from previous assessment *not* included as independent variable	**0.74**	**0.09**
	Baseline model using prevalence from previous assessment only	0.51	0.15
	Baseline model using GDP per capita only	0.40	0.14
	Baseline model using prevalence of under-nourishment only	0.39	0.13
	Baseline model using prevalence of under-nourishment and GDP per capita	0.66	0.11
	Baseline model using weather-related variables only	0.14	0.14
Food-based coping	**Prevalence from previous assessment included as independent variable**	**0.73**	**0.08**
	Prevalence from previous assessment *not* included as independent variable	**0.61**	**0.09**
	Baseline model using prevalence from previous assessment only	0.45	0.12
	Baseline model using GDP per capita only	0.20	0.12
	Baseline model using prevalence of under-nourishment only	0.19	0.12
	Baseline model using prevalence of under-nourishment and GDP per capita	0.39	0.10
	Baseline model using weather-related variables only	0.37	0.13

Coefficients of determination (R^2) and mean absolute errors (MAE) obtained on the test set for each of the four proposed models (in bold) and for the different baseline approaches.

data, as further detailed in Methods. As reported in Table 1, the proposed models are able to explain on the remaining approximately 15% out-of-sample data (that is, corresponding to the past two months of data), 81% of the variation in the prevalence within each country subdivision of people with insufficient food consumption and 73% of the variation in the prevalence of people using crisis or above-crisis food-based coping with a mean absolute error of, respectively, 0.07 and 0.08. Figure 1 (top plots) shows the predicted versus actual prevalence for each observation in the test set. The former is calculated as the median of the predicted values obtained from the N_b bootstrapped models.

As one might expect, in both models, the most predictive variable is the previously measured prevalence (Supplementary Fig. 1). Therefore, the question arises whether the independent variables built from secondary data bring significant additional information
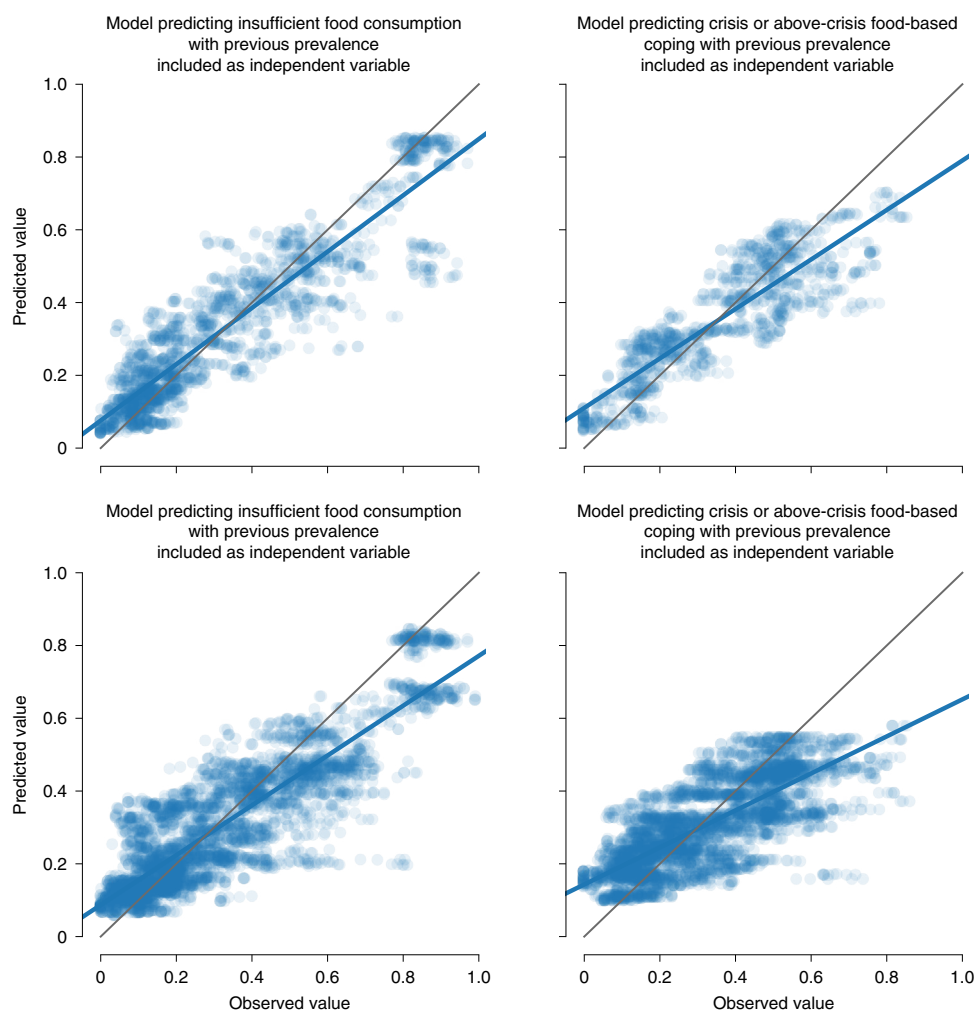
Fig. 1 | Predicted versus observed values in the test set for each of the four models. Each plot shows the predicted value (obtained as the median of $N_b = 100$ predictions, each generated from one of the bootstrapped models) versus the actual value for each observation in the test set. The blue line represents the best fit for the plotted points, whereas the grey line represents where the points would fall if all predicted values perfectly matched the observed ones. The closer the two lines are, the better the model's performance is. The top plots refer to the models that include the prevalence from the previous assessment as an independent variable, and the bottom plots refer to the models that use secondary data only.

into the models or whether most of their explanatory power could be due to the previous assessment variable. To tackle this question, we compare the results of the proposed models with those obtained from a naive approach that uses only the prevalence measured during the previous assessment as an independent variable. We find that this naive model can explain only 51% of the variation in the prevalence of people with insufficient food consumption and 45% of the variation in the prevalence of people using crisis or above-crisis food-based coping, demonstrating the fundamental importance played by secondary data to capture the dynamic nature of food insecurity and to explain the current situation when up-to-date primary data are not available.

Predictions using secondary data only. Having demonstrated the potential of the proposed approach when information on both key drivers and previous values of the target indicator is available and the fundamental role played by the secondary data, we then tested to what extent insufficient food consumption and crisis or above food-based coping levels can be predicted when previously measured prevalence is not available.

We trained two additional models, using the same approach but removing the prevalence from the previous assessment from

the set of independent variables, hence using secondary data only. As reported in Table 1, in this case results show that the proposed models are able to explain, on the test set, 74% of the variance in the prevalence of people with insufficient food consumption and 61% of the variance in the prevalence of people using crisis or above-crisis food-based coping, with a mean absolute error of 0.09. Figure 1 (bottom plots) shows the predicted versus actual prevalence for each observation in the test set. As expected, these latter models have lower explanatory power and slightly higher errors than the former ones; however, the reported metrics are still satisfactory. The advantage of these models is that they also allow prediction of the food security situation in areas where no previous measurement is available, substantially expanding the application horizon of the proposed approach.

To test to what extent the performance of these models is due to variables such as the GDP per capita and the prevalence of under-nourishment, which are strong proxies for a country's socio-economic development, we created a set of baseline models using these individual variables in isolation and in combination. Results show that these variables alone can explain up to 66% of the variation in food consumption and 39% of the variation in food-based coping. This means that these variables are, indeed,

those explaining the largest part of the variation in the two indicators; however, because they are annual national-level figures, they serve as a fundamental baseline but cannot help in predicting the sub-national and rapidly changing dynamics characterizing food insecurity, which is the objective of this study.

Finally, given the recent advances in the use of climate extremes data in famine early-warning systems[26], we also created an additional baseline which includes the weather-related variables only. Results show that these models can explain only 14% of the variation in food consumption and 37% of the variation in food-based coping and cannot therefore be used in isolation. However, the dynamic nature of weather-related features is fundamental to predict the observed rapid changes in the food security situation, as shown in the final section.

Near-real time nowcasts. WFP is currently monitoring the food security situation in near real time in a number of countries, collecting food consumption and food-based coping data through daily remote phone surveys[1,27]. The predictive models proposed in this study aim at serving WFP's need to estimate the situation in additional countries where primary near-real time data are not currently available to provide humanitarian stakeholders with regular and frequent up-to-date global overviews of the food security situation and allow for timely decision-making on resource allocation.

To test the effectiveness of the proposed models in capturing the current situation, we compared insufficient food consumption and crisis or above food-based coping trends measured by WFP's near-real time monitoring systems between 1 August and 30 September 2021 with the corresponding prevalence predicted by the proposed models, which were trained and tested on data collected before 1 August 2021. For areas where the prevalence from a previous assessment—performed before the start of the near-real time monitoring system in the country—is available, we use the models that include this information as an independent variable; for areas where this is not available, we resort to the models that rely on secondary data only.

National-level results for insufficient food consumption are shown in Fig. 2. The red lines represent the target prevalence as measured by WFP's near-real time monitoring systems, the blue lines the predicted prevalence and the green dashed lines the prevalence from the previous assessments, where available. All prevalence levels were first obtained at the spatial resolution of first-level administrative units and then aggregated to obtain national trends. Sub-national trends are reported in Supplementary Figs. 2–16. We can observe that in most cases the prevalence measured by the near-real time monitoring systems falls within the prediction intervals (or within a reasonable distance of less than 5%) for at least part of the trend. In those cases where the actual data line is further from the prediction interval, we can observe that the predicted trend is however visibly closer to the observed one than the prevalence from the previous assessment (for example, Malawi and Zambia). In the remaining case, where no previous assessment is available (that is, Somalia), the predicted and observed trends both fall within the same severity level (>40%) defined by WFP[27]. Similar results can be observed for crisis or above food-based coping in Fig. 3 (Supplementary Figs. 17–31 provide the corresponding sub-national trends), with the exception of Congo and Somalia.

To better understand these findings, a characterization of prediction errors was carried out. As shown in Fig. 4, we classify errors on the basis of how far predicted values are from the observed ones: we classify as correct those predictions that differ from the observed value by maximum ± 5 prevalence points; as high over-estimation (under-estimation) a predicted prevalence >40% (<40%) when the observed prevalence is <40% (>40%); finally, all other regions are classified as low under- and over-estimation. We find that 44.2% (36.8%) of predictions are classified as correct and another 46.0%

(50.5%) are low under-/over-estimations, whereas 2.1% (5.0%) of predicted values highly overestimate and 7.7% (7.7%) highly underestimate the prevalence of people with insufficient food consumption (with crisis or above-crisis food-based coping). Supplementary Figs. 32–33 show the distribution of each independent variable for training data, for correct predicted values and for high over- and under-estimated predicted values. Results indicate that, as one might expect, errors happen when the independent variables take on values that differ the most from those most frequently seen by the models during training. This is particularly evident for the prevalence of under-nourishment, the prevalence from previous assessment and market price alerts. For instance, the high over-estimation of crisis or above food-based coping in Congo and Somalia is probably due to the corresponding extremely high values of the prevalence of under-nourishment. These insights should guide the handling of predictions: when running the models for a new area and date, one should first check the values of the input variables and how they compare to their distribution in the historical data used to train the model. When values falling in the tails of the distributions are observed, care must be taken as predictions are more likely to be farther from the actual unknown prevalence.

Overall, these results show that the proposed models allow WFP to obtain adequate national estimates of the considered food security indicators for the majority of countries of operational relevance for WFP, complementing the information from WFP's near-real time monitoring systems.

Explanations of predicted values and of changes in predicted trends. Machine learning approaches are often seen as black boxes that provide recommendations without the user being able to access the process and rationale that generated them. This is not an acceptable practice when the model outputs are being generated in support of decision-making. Therefore, in this context, proposing methods to explain predicted results is as important as building the models themselves.

Here we make use of SHAP (SHapley Additive exPlanations) values[28,29] to explain how each prediction is obtained. SHAP values make it possible to explain each predicted prevalence as a value obtained starting from the average prevalence observed in the training set (baseline) and then accounting for how much each independent variable contributes to the final prediction by moving the prevalence towards lower or higher values.

We first show how this method is able to demystify the predicted prevalence of people with insufficient food consumption or of people using crisis or above-crisis food-based coping, explaining how the models predict values compatible with what has been measured through WFP's near-real time monitoring systems. In Fig. 5, we show an exemplification of this approach to explain both predicted indicators in the cases of Mali on 1 August 2021, using a waterfall plot approach[29]. Starting from the bottom, each variable's contribution is summed to the baseline $E(f(x))$ to eventually reach the predicted value $f(x)$. Variables are ordered by importance (in terms of the absolute value of their contribution) and coloured by the sign of the contribution: red if increasing and blue if decreasing, with respect to the baseline. In Fig. 5a, we see that the most important variables in determining the high prevalence of people with insufficient food consumption (0.52) in Mali is the prevalence of under-nourishment (>5%), together with the low GDP per capita (US$778.5). Conversely, the lower value of the prevalence measured during a previous assessment (0.20) drives the predicted value down, similarly to what happens in Fig. 5b when considering the prevalence of people using crisis or above-crisis food-based coping. Supplementary Figs. 52–79 report the same plots for the remaining 14 countries shown in Figs. 2–3. It should be noted that each country (and sub-national area) is characterized by a different variable importance order. For instance, in high-intensity conflict areas
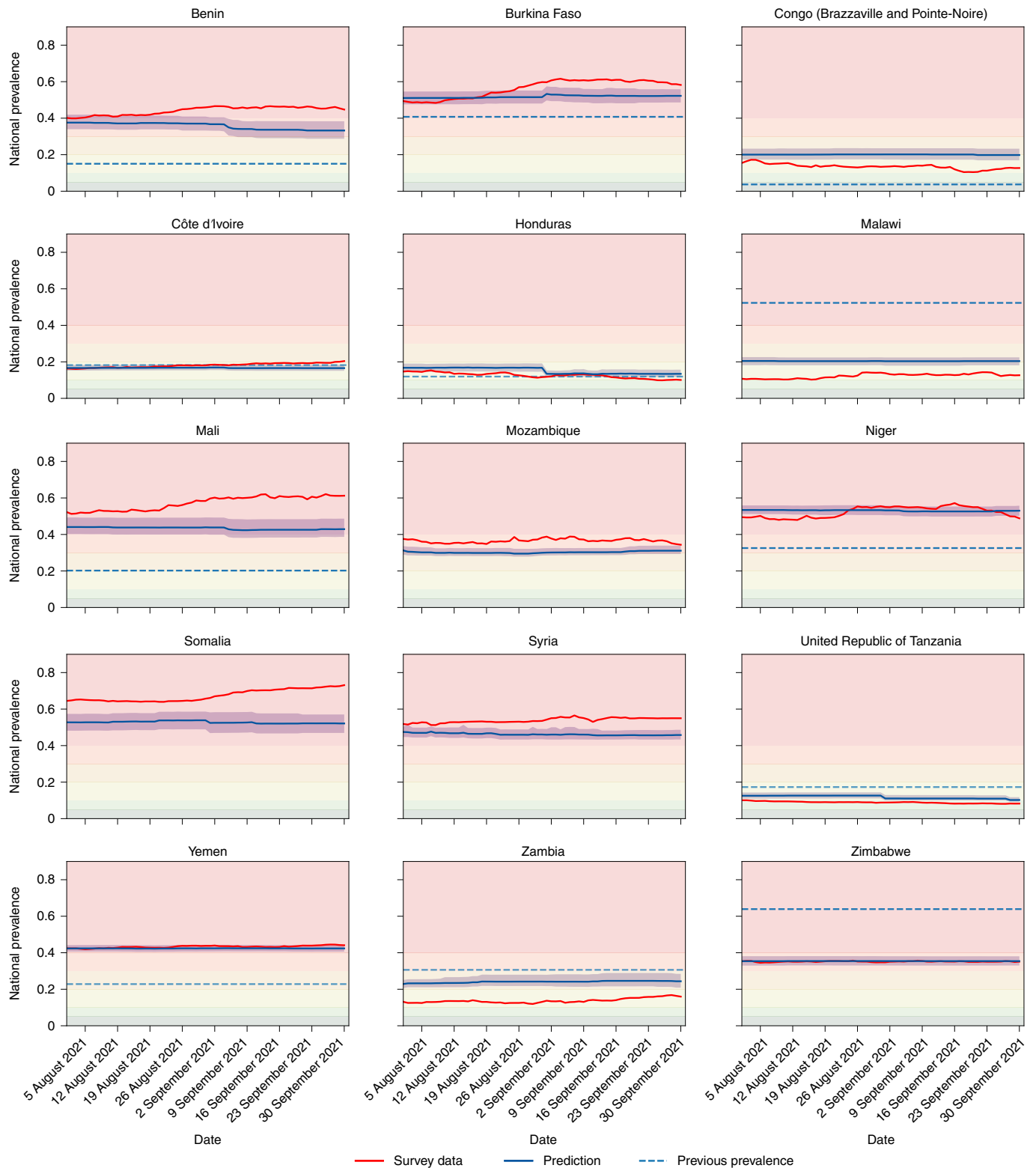
Fig. 2 | Comparison between near-real time monitoring of insufficient food consumption and predicted trends. Each plot shows the prevalence of people with insufficient food consumption between 1 August and 30 September 2021, as measured through WFP's near-real time monitoring systems (red lines) and as predicted by the proposed models (the blue lines represent the median of the $N_b = 100$ bootstrapped models predictions, and the light blue area around them the corresponding 99% confidence interval). The dashed green lines represent the value measured during the previous assessments (performed before the start of the near-real time monitoring system in the country), where available. The background colours represent severity levels in terms of national prevalence as defined by WFP (<5%: very low, 5–10%: low, 10–20%: moderately low, 20–30%: moderately high, 30–40%: high, >40%: very high)[27].
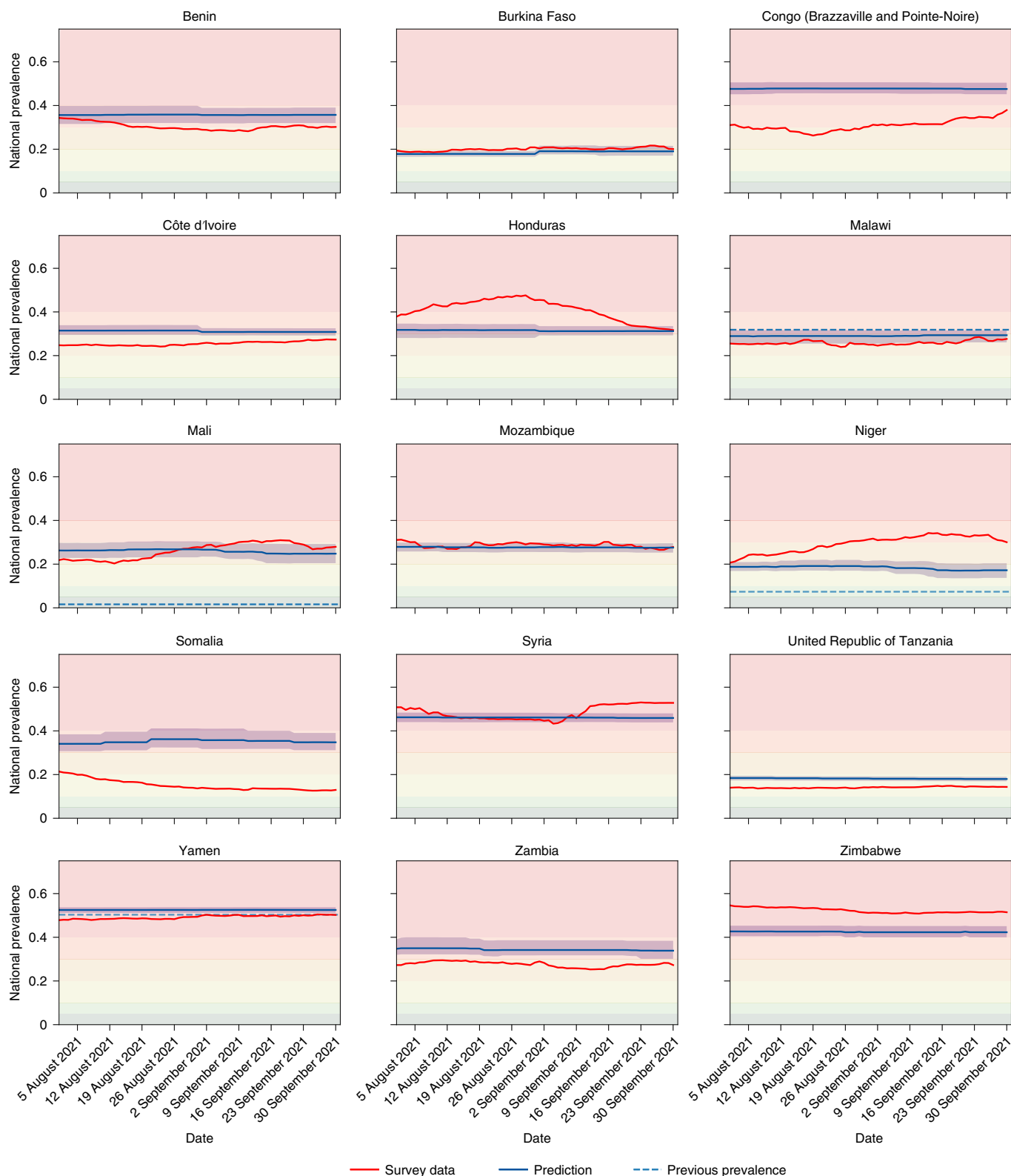
Fig. 3 | Comparison between near-real time monitoring of crisis or above food-based coping and predicted trends. Each plot shows the prevalence of people using crisis or above-crisis food-based coping between 1 August and 30 September 2021, as measured through WFP's near-real time monitoring systems (red lines) and as predicted by the proposed models (the blue lines represent the median of the $N_b = 100$ bootstrapped models predictions and the light blue area around them the corresponding 99% confidence interval). The dashed green lines represent the value measured during the previous assessments (performed before the start of the near-real time monitoring system in the country), where available. The background colours represent severity levels in terms of national prevalence as defined by WFP (<5%: very low, 5−10%: low, 10−20%: moderately low, 20−30%: moderately high, 30−40%: high, >40%: very high)[27].
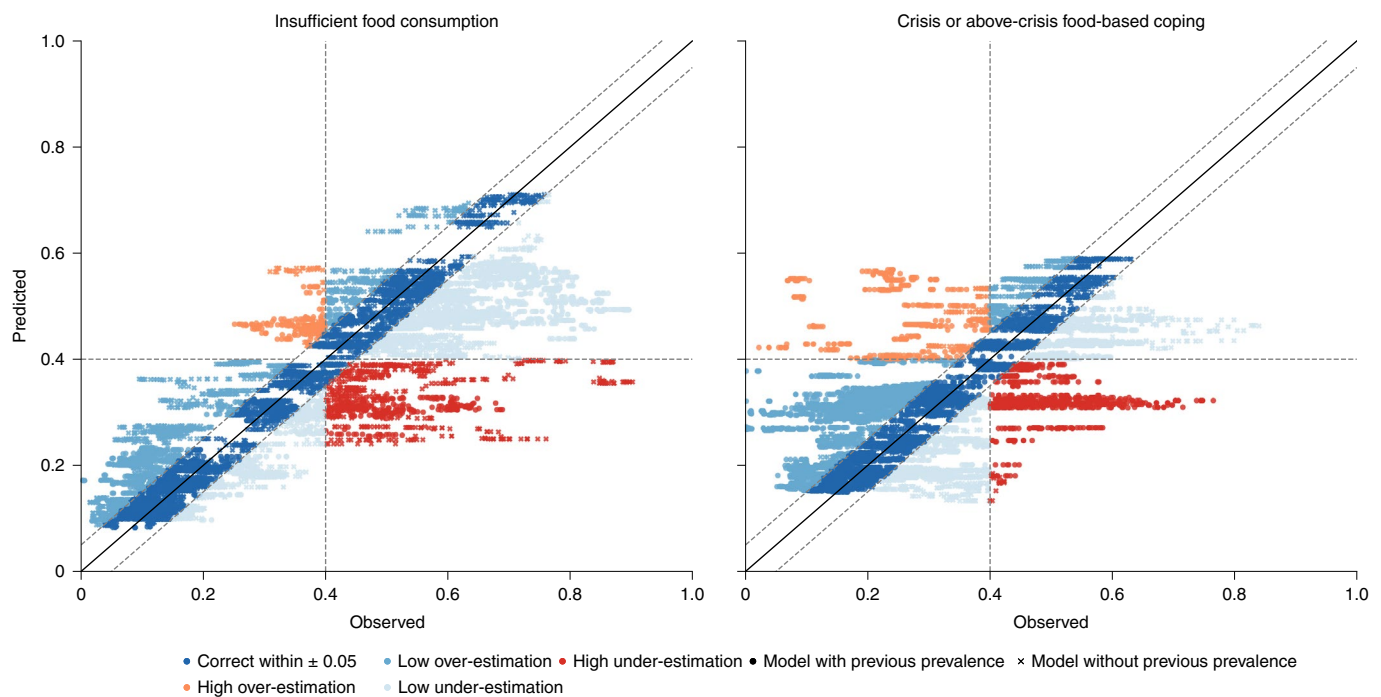
Fig. 4 | Error classification for real time monitoring predictions. Predictions that differ from the observed value by maximum ±5 prevalence points are classified as correct. Predicted prevalence >40% (<40%) when the observed prevalence is <40% (>40%) are classified as high over-estimation (under-estimation). The other regions are classified as low under- and over-estimation. The solid black line indicates where the points would fall if all predicted values perfectly matched the observed ones, and the grey dashed diagonal lines indicate a distance of ±5 prevalence points from it. The grey dashed horizontal and vertical lines indicate the 40% prevalence thresholds.

in countries such as Somalia, Syria and Yemen, the conflict variable is in the top 50% most important variables, as one can see in Supplementary Figs. 80–83.

Beyond using SHAP values to explain individual predictions, in this study, we propose a method based on this framework to measure the relative importance of each independent variable in explaining changes in the predictions of food consumption and food-based coping between two dates. This is important for decision-makers to understand why the models, when deployed to produce near-real time trends, show improvements or deterioration in the food security situation. This is done by exploiting the differences in SHAP values between two dates, and its mathematical formulation is detailed in Methods. In Fig. 6, we show the proposed method applied to a specific example. On the left, we see how the predicted food consumption situation in Indonesia deteriorated between 15 August and 15 September 2021. Our method is able to identify that the most important variable in determining this change has been the change in food inflation, which, as shown in the bottom table, increased within the month under consideration. Other variables also had a smaller impact in the change, for example, an increase in the amount of rainfall and in the greenness of the vegetation, with respect to the historical averages in the same period of the year, played instead a positive role reducing the extent of the predicted deterioration. Let us note that variables that did not change their value during the time period considered can, however, still change their SHAP importance, as this is relative to the values of all variables at each point in time. On the right, we explain an increase in the predicted prevalence of people using crisis or above-crisis food-based coping in the same time period in the same country. In this case, the main causes are an increase in both food and headline inflation, while an increase in the greenness of the vegetation with respect to the historical average contributed to reduce the extent of the deterioration. Sub-national

investigations reveal that the positive contribution of rainfall and vegetation greenness is not homogeneous across provinces, as one can see in Supplementary Figs. 84–151.

To conclude, this kind of analysis will allow WFP to identify the main factors behind a predicted change in insufficient food consumption or in crisis or above food-based coping. More specifically, WFP will be able to divide countries in various tiers of risks by combining levels of prevalence of insufficient food consumption and food-based coping with levels of deterioration over time[30]. This will allow food security experts to evaluate to what extent the situation should be considered concerning and whether more in-depth analyses are required, including, for instance, triggering ad hoc data collection to obtain further information on the different concerned dimensions.

Discussion

In this study, we propose an approach that makes it possible to predict the current sub-national food consumption and food-based coping situation on a global scale from secondary data on the key drivers of food insecurity. We show that when a previous measurement of the target indicator is available and included as an independent variable, the models, as expected, have a higher explanatory power and lower errors than when relying on secondary information only. Importantly, we also demonstrate that these models perform significantly better than a naive approach that uses the prevalence measured during a previous assessment as the predicted prevalence. Moreover, even the models that rely on secondary information show only a satisfactory explanatory power. Having trained and validated the proposed models on historical data, we then further show that they can be used to predict the current prevalence of insufficient food consumption and of crisis or above food-based coping by comparing the data being collected in near real time by WFP during a recent two-month period with the corresponding predicted levels.
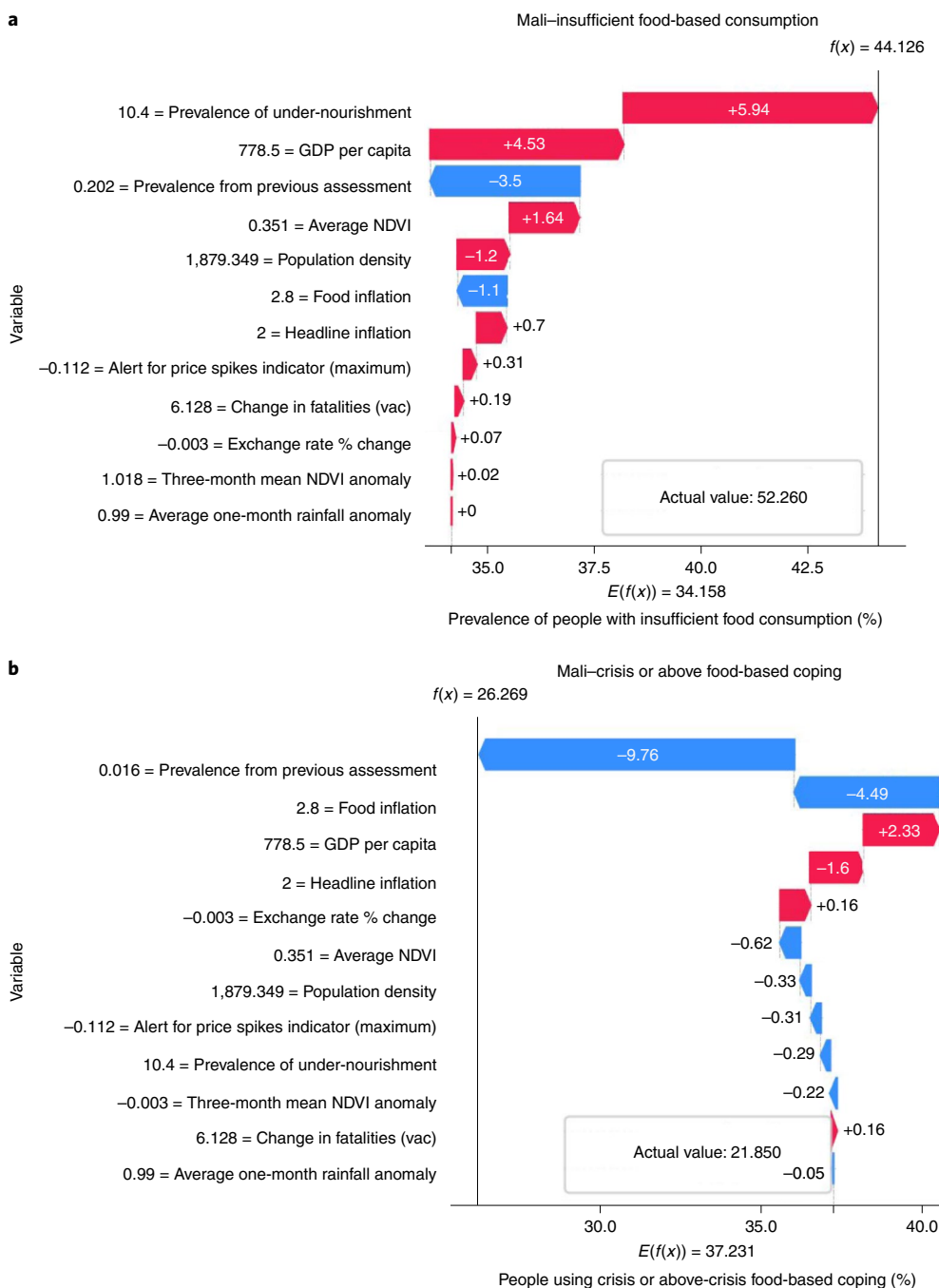
a

Mali–insufficient food-based consumption

f(x) = 44.126

10.4 = Prevalence of under-nourishment ················· +5.94

778.5 = GDP per capita ················· +4.53

0.202 = Prevalence from previous assessment ················· −3.5

0.351 = Average NDVI ················· +1.64

1,879.349 = Population density ················· −1.2

2.8 = Food inflation ················· −1.1

2 = Headline inflation ················· +0.7

−0.112 = Alert for price spikes indicator (maximum) ················· +0.31

6.128 = Change in fatalities (vac) ················· +0.19

−0.003 = Exchange rate % change ················· +0.07

1.018 = Three-month mean NDVI anomaly ················· +0.02

0.99 = Average one-month rainfall anomaly ················· +0

Actual value: 52.260

35.0   37.5   40.0   42.5

E(f(x)) = 34.158

Prevalence of people with insufficient food consumption (%)

b

Mali–crisis or above food-based coping

f(x) = 26.269

0.016 = Prevalence from previous assessment ················· −9.76

2.8 = Food inflation ················· −4.49

778.5 = GDP per capita ················· +2.33

2 = Headline inflation ················· −1.6

−0.003 = Exchange rate % change ················· +0.16

0.351 = Average NDVI ················· −0.62

1,879.349 = Population density ················· −0.33

−0.112 = Alert for price spikes indicator (maximum) ················· −0.31

10.4 = Prevalence of under-nourishment ················· −0.29

−0.003 = Three-month mean NDVI anomaly ················· −0.22

6.128 = Change in fatalities (vac) ················· +0.16

0.99 = Average one-month rainfall anomaly ················· −0.05

Actual value: 21.850

30.0   35.0   40.0

E(f(x)) = 37.231

People using crisis or above-crisis food-based coping (%)

Fig. 5 | Explanation of a single prediction of insufficient food consumption and of crisis or above food-based coping. Waterfall plots[29] show how SHAP values are used to explain the predicted prevalence f(x) on 1 August 2021 in Mali (aggregating first-level administrative unit predictions) as the sum of a baseline E(f(x)) and each variable's contribution, highlighting in red positive contributions and in blue negative ones. Prevalence and variable contributions are expressed as percentages. Next to each variable's name (defined in Methods), its value-averaged weighting by population over all first-level administrative units in the country is shown. The boxes contain the actual values measured through the near-real time monitoring system for the same day. **a**, The prevalence of people with insufficient food consumption. **b**, People using crisis or above-crisis food-based coping. The corresponding sub-national plots are reported in Supplementary Figs. 34–51.

Finally, we show that despite the nonlinear tree-based model structure, it is possible to provide interpretable explanations of predicted figures and of what causes changes over time, even if the models do not have an intrinsic dynamic component.

Despite the encouraging results, several limitations apply. First of all, the proposed models are trained and validated combining together sub-national data from a number of different countries.

Having such rich variety of data made it possible to build a global model that can be used to estimate the food security situation in any area in the world. However, this also means that the models learned patterns in the data that correspond to what is most commonly observed across the different countries, limiting the discovery of less-frequent patterns specific to local contexts. These latter patterns would be more easily discovered by training separate
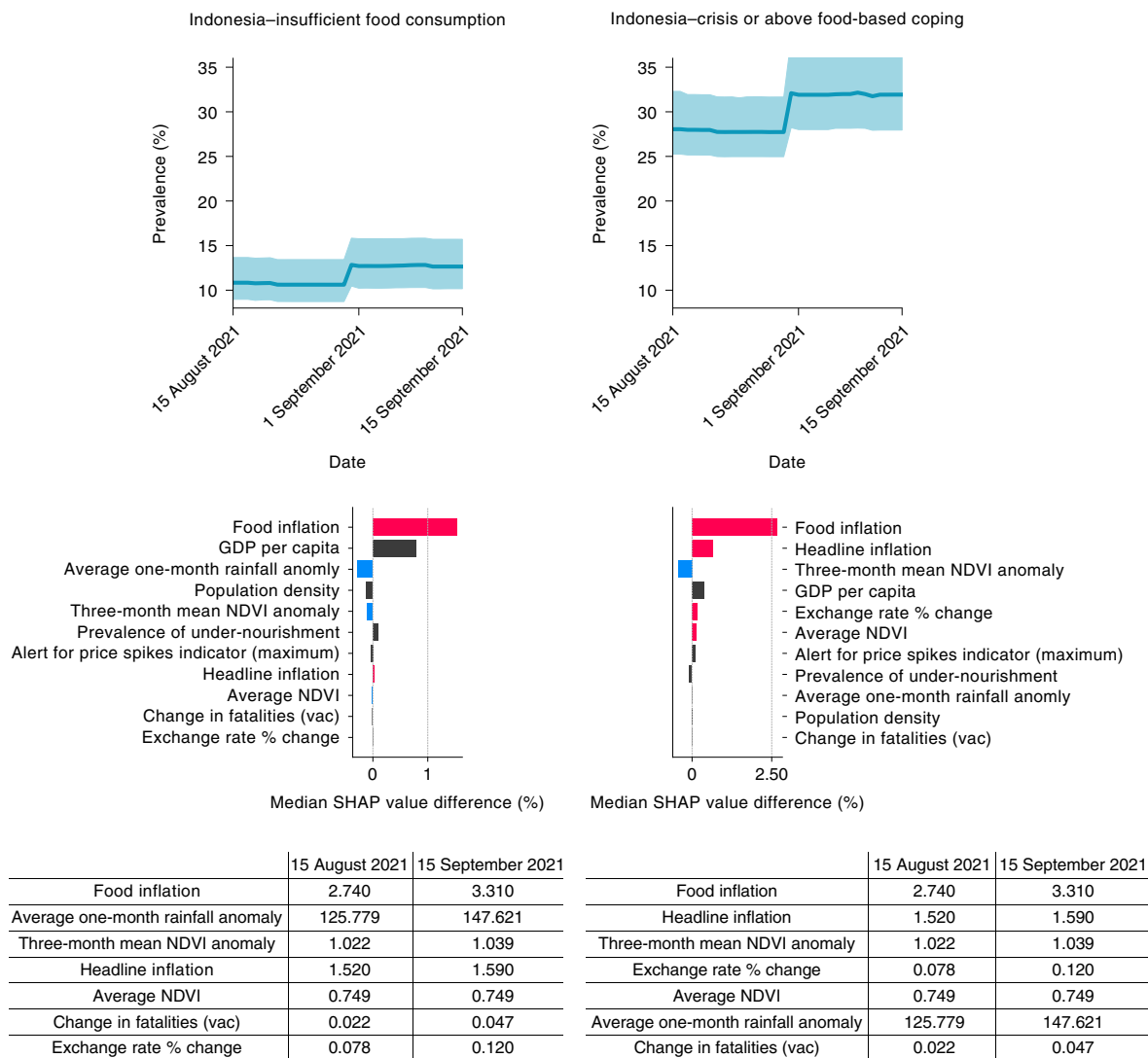
Fig. 6 | Explanation of changes in predicted trends between two dates. A SHAP values-based method was developed to explain why the models are predicting changes in prevalence between two dates. We explain the change in predicted trend for the prevalence of people with insufficient food consumption (left) and of people using crisis or above-crisis food-based coping (right) between 15 August and 15 September 2021 in Indonesia. On the top plots, we can see the predicted trends in blue, with the light blue area around them the corresponding 99% confidence interval. The middle plots show each independent variable's contribution to the change: positive contributions (deterioration) are shown in red, negative contributions (improvements) in blue and variables that did not change value between the two dates are shown in grey. Variables are ordered by importance (in terms of absolute value of their contribution). All prevalence and SHAP value differences are expressed as percentages. The tables in the bottom report the values of the models' independent variables at the two considered dates. The corresponding sub-national plots are reported in Supplementary Figs. 84–151.

models each based on historical data from a specific country only. This would, however, require the availability of large enough samples for each individual country, which is currently not available for a number of countries. Hence the reason for the proposed approach. However, it should be noted that predictions generated by the proposed models should be used with caution and further validated when relative to areas and countries that are not represented in the historical data used to train and test the models, as discussed in previous studies[31]. In this sense, one of the challenges faced in the model development was the unequal availability of food consumption and food-based coping data across different countries. To ensure as much as possible a balanced geographical representation, we resorted to sampling and kept only a subset of the available data for the most data-rich countries, while also ensuring enough data were included to properly train the models. Future works should explore country-specific models for data-rich countries and how

these could be used to complement the current approach through, for example, ensemble modelling.

In regard to the secondary information feeding the proposed models, in this study, we resort to data on the three main drivers of food insecurity: economic shocks, extreme weather events and conflict. Undoubtedly, further information could be included to enrich the models, such as data on displacements, natural hazards, animal and crop diseases and epidemic outbreaks[11]. However, the limited availability of relevant data on a global scale and on a multi-annual time frame restricts the possibilities of expansion to additional independent variables. Given that the time frame covered by the historical data used to train and test the models includes the COVID-19 pandemic, one might expect that we would need to include this information in the models, for example, in the form of case load or death incidence. This is, however, not the approach we adopted, because the objective was to build a

general model for food insecurity, not specific to the current situation. Our assumption is that the effects of this pandemic on the food security situation are already indirectly taken into account by some of the independent variables included in the models, namely those accounting for staple food prices in local markets and macro-economic indicators, which were heavily affected by the pandemic[32,33]. Further investigation should, however, be the subject of future work.

Some challenges and limitations also apply to the secondary data that have been incorporated in the models to build the independent variables. First, the different data sources do not have the same time resolution and update frequency, which range from annual estimates to daily measures, as further detailed in Methods. This means that when generating predictions on a regular basis to nowcast the situation, most variables will not update daily, and therefore, to see notable changes, longer time intervals need to be considered. Secondly, spatial resolution also varies across the different data sources. Whereas population density, rainfall, vegetation and conflict-related fatalities data are available at the first-level administrative unit resolution, macro-economic indicators and prevalence of under-nourishment are national figures, leading to all sub-national areas being assigned the same value. This could be seen as a limitation, but it also allows provision to each first-level administrative unit of some national characterization, which would otherwise be lost in a global model trained with sub-national-level data only. Let us consider, for example, two bordering areas belonging to two different countries, such as Venezuela and Brazil. Given their geographical proximity, they might be highly similar with respect to vegetation and precipitation, but, concurrently, they would be highly different in terms of the economic situation, possibly resulting in profound differences in the food security situation.

From a modelling perspective, a limitation of this study is that it proposes one modelling framework only and does not provide comparison with other models, such as linear regression and decision trees. XGboost was selected due to its well-documented excellent performance on a wide range of problems[25,34,35]. Additional models might, however, be explored in future work.

Finally, while SHAP values are used in this work only to understand how each prediction is obtained by the proposed model, some limitations of this method are worth mentioning to avoid its misuse by decision-makers. For instance, it is important to stress how explaining individual predictions is different from understanding driving factors[36]. Moreover, while SHAP values have been frequently used by the machine learning community in recent years, their adoption is not unanimous, and some limitations on their use as explanatory tools have been pointed out[37].

Having carefully taken into account the challenges and limitations highlighted above, the proposed models have the potential to be used to provide unique information to humanitarian decision-makers when no primary data are available. Predictions should certainly be handled with caution and never considered as ultimate truth. When indicating some level of deterioration, they should serve as triggers for rapid assessments and more in-depth analysis of the situation, rather than being used to prompt immediate decision-making. In this regard, the proposed methods give decision-makers more insights into how the model predicted a certain figure or changes in the predicted trends, allowing for a deeper understanding of the situation. Finally, it should be noted that to ensure continued validity of the proposed models, it is essential to perform regular re-trainings whenever a considerable amount of new primary data is collected and available. This will allow improvement of the model's explanatory power thanks to the increased volume and variety of data the training is performed on, and to eventually learn new emerging patterns, hence remaining representative of the current situation.

Methods

Target indicators. The two target indicators of the proposed predictive models—the prevalence of people with insufficient food consumption and the prevalence of people using crisis or above-crisis food-based coping—are calculated on the basis of two household-level indicators: the FCS[38] and the rCSI[39], respectively.

The FCS is calculated by asking each household how often, during the past seven days, it has consumed items from the different food groups: main staples, pulses, vegetables, fruit, meat and fish, milk, sugar, oil and condiments. Each consumption frequency is then weighted according to its relative nutritional importance to obtain the FCS $= \sum w_i x_i$, where w_i is the weight of food group i, and x_i is the frequency of its consumption by the household, that is, the number of days for which the food group was consumed during the past seven days. Once the food-consumption score is calculated, each household is then assigned a food-consumption group (poor, borderline or acceptable) based on standard thresholds, which can, however, be adapted based on specific consumption behaviours in the country of interest. Food group weights and thresholds are detailed in ref. [38].

To compute the rCSI, households are asked if and how often during the past seven days they had to adopt the following coping behaviours: relying on less-preferred or less-expensive food, borrowing food from relatives or friends, limiting portion sizes, restricting adults' consumption for small children to eat and reducing the number of meals eaten in a day. Coping strategy frequencies are then weighted according to their severity to obtain the rCSI, as detailed in ref. [39].

The available historical data for the two indicators have statistical representativeness at the spatial resolution of first-level administrative units and have been collected through both face-to-face and mobile phone surveys, including those from WFP's near-real time monitoring systems. Because the mode of questionnaire administration can have serious effects on data quality[40], post-stratification weighting schemes are applied by WFP to surveys collected in a remote fashion to mitigate sampling and modality bias, as detailed in ref. [41]. The prevalence of people with insufficient food consumption and the prevalence of people using crisis or above-crisis food-based coping are then obtained as the weighted prevalence of households in the sample with, respectively, poor or borderline food consumption and with rCSI ≥ 19 (ref. [42]).

The insufficient food-consumption data span units across 78 countries from 2006 to July 2021, and the crisis or above food-based coping data units across 41 countries from 2013 to July 2021, with more than 200,000 observations for each indicator. This large volume of data is, however, not equally representative of all covered geographical areas: countries where a WFP's near-real time monitoring system is in place are over-represented because these systems provide data on a daily basis, whereas in the remaining countries, data collection is performed only a few times per year. Therefore, to avoid training the models on an unbalanced dataset, sampling is performed by randomly selecting, for each first-level administrative unit, five observations per month only. The final dataset used to train and test the models where the prevalence from previous assessments is not included as an independent variable is composed of 37,926 observations for food consumption and 35,894 for food-based coping. For the models where the prevalence from a previous assessment is instead included, the size is further reduced because only observations preceded by a previous one can be used, resulting in 24,510 observations for food consumption and 12,570 for food-based coping. The breakdown by country of all of the above-mentioned numbers is reported in Supplementary Table 1.

Feature definition and selection. The initial set of considered features is composed of variables related to food insecurity and its main drivers: economic shocks, extreme weather events and conflict[11]. Because the goal of the models is to provide interpretable insights to decision-makers, the guiding principle for feature definition was expert opinion. To include as much variety as possible in terms of potential drivers, a manual, expert-guided, minimal feature selection process was performed, as detailed below. Let us note that such an approach is not expected to remove all multi-collinearity, but a tree-based method (XGboost) that is robust to multi-collinearity[43] was subsequently used.

Food prices. To capture variations in cereal and tuber prices, we resort to the Alert for Price Spikes (ALPS) indicator[44]. This metric is based on a trend analysis of monthly price data: the idea is to compare the long-term seasonal trend of a commodity's price time series in a market with the last observed price in the same market, providing an indication of the intensity of the difference between the current market price and the historical trend. The higher the difference, the more severe the alert. Price data and the corresponding ALPS calculations are publicly available through WFP's Economic Explorer platform[45]. If more than one market is present within a geographical area, the average ALPS value is considered. If no market is instead monitored in a given area, the national average is considered. From these data, we build a set of three features by taking into account the minimum, maximum and average ALPS value within a three-month window. The length of the window was selected as the shortest time period that minimizes the number of missing values in the training and test set. A one-month lag is applied to ensure data availability when deploying the model in real time. Because the three defined features are different variants of the same indicator, and they display

high levels of positive correlation (Spearman's correlation coefficients $\rho > 0.85$ for all three combinations), only one was selected as independent variable for the models. The selected variable is the maximum ALPS value, being the one with the highest correlation with the target indicators (correlations were computed using the training data only).

Macro-economic indicators. The following four macro-economic features are considered: most-recent available annual GDP per capita in a four-year time window, most-recent available monthly headline and food inflation rates in a six-month time window (applying a one-month lag) and percentage variation between the average value of the currency exchange rate during the past three months and the average value during the previous three, to capture main changes in the situation. The three-month window was selected for consistency with the food price features, and the same applies to the following features too. Being all country-level indicators, the same national value is assigned to all first-level administrative units within a country. Data are obtained from Trading Economics, a website providing publicly available economic and financial indicators, including historical data, for 196 countries[46]. For countries where unofficial currency exchange rate is collected by WFP, these values are used instead of official ones[47], because they provide a more reliable representation of the country's economic situation.

Weather. An initial set of five weather features is built by taking: (1) the average rainfall and normalized difference vegetation index (NDVI) during a 12-month time window, which allows characterization of each first-level administrative unit's climate; (2) rainfall and NDVI anomalies with respect to historical averages. For rainfall, two anomalies have been defined by WFP: the ratio between the amount of rainfall during the past one month or three months and the historical average of the amount of rainfall in the same period of the year. For NDVI, a single anomaly is defined based on the past ten days, because vegetation already integrates the effects of previous rainfall. All three anomalies are provided for each ten-day window of the year, and we take their average during a three-month window as for previous features, applying a ten-day lag. Data are obtained from WFP's Seasonal Explorer platform[48], which provides open rainfall and NDVI time series for a near-global set of administrative units, computed, respectively, from the Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS)[49] and the Moderate Resolution Imaging Spectroradiometer (MODIS)[50] data. In this case, similar to food prices, the defined features are different variants of the same two indicators, and a correlation analysis was performed to reduce their number. Between the two averages during a 12-month time window ($\rho = 0.93$), the average NDVI was selected as the most correlated with the target indicators. Similarly, between the two rainfall anomalies ($\rho = 0.82$), the one-month anomaly was selected. Finally, the NDVI anomaly was also selected given its low correlation with the other four variables ($\rho < 0.4$ in all cases).

Conflict. Conflict data are obtained from the Armed Conflict Location and Event Data Project, a publicly available repository of reported conflict events and related fatalities across most areas of the world[51]. The date, latitude and longitude of each event is reported, allowing it to match to the corresponding first-level administrative unit. To capture deterioration or improvements in the conflict situation, we define the conflict features as the difference between the number of reported fatalities during the past three months and the three months prior, applying a 14-day lag. Only fatalities reported in events involving organized violence (that is 'battles', 'violence against civilians' and 'explosions/remote violence') are considered[52], and a total of seven features is obtained by considering these three categories separately and in combination. Again, a correlation analysis was performed in the same fashion as for the previous cases, and only the 'violence against civilians' feature was selected as an independent variable for the models.

Prevalence of under-nourishment. The most-recent available prevalence of under-nourishment in a four-year time window is considered. This is a national yearly indicator publicly available in the Food and Agriculture Organization Corporate Statistical Database (FAOSTAT)[53]. Being a country-level indicator, the same national value is assigned to all first-level administrative units within a country.

Population density. Yearly population density is also considered and obtained from the Center for International Earth Science Information Network (CIESIN) raster files[54] by averaging all pixel values within each first-level administrative unit. Estimates for years not covered by the dataset are obtained through linear interpolation.

Previous value of the target indicator. Finally, the previously measured prevalence of people with insufficient food consumption and of people using crisis or above-crisis food-based coping are also considered, when available. For first-level administrative units where WFP's near-real time monitoring is in place, only data collected before the start of the near-real time monitoring is used to build this feature. This choice was made because the proposed models are meant to be used in practice in situations where no near-real time monitoring is in place, and hence the last-available value would be from a face-to-face or mobile phone assessment conducted in a specific and limited time window in the past.

Modelling approach. The proposed models predict the probability of having a person with insufficient food consumption or using crisis or above-crisis food-based coping in a given area at a given time. The objective function is therefore a logistic curve. Gradient Boosted Decision Trees[55] was identified as the most suitable algorithm to perform the regressions, given its high performance, flexibility and its capacity to handle complex and nonlinear relationships. The XGboost implementation was used[25].

Four different models were developed: two for the prevalence of people with insufficient food consumption and two for the prevalence of people using crisis or above-crisis food-based coping. In each case, one includes the prevalence from a previous assessment as independent variable, and one does not.

That is, in the first model, the dependent variable is the prevalence of people with insufficient food consumption in a given first-level administrative area a at a given date d, and the corresponding independent variables are: (1) the last-available prevalence of people with insufficient food consumption in area a at a time before d (and before the start of the near-real time monitoring, if this is in place in area a); (2) the most-recent prevalence of under-nourishment (in a four-year time window) available at date d for the country that area a is part of; (3) the most-recent annual GDP per capita (in a four-year time window) available at date d for the country that area a is part of; (4) the most-recent headline inflation rate (in a six-month time window) available at date d for the country that area a is part of; (5) the most-recent food inflation rate (in a six-month time window) available at date d for the country that area a is part of; (6) the percentage variation between the average value of the currency exchange rate during the three months preceding date d and the average value during the previous three for the country that area a is part of; (7) the maximum ALPS value in area a for the three months before date d; (8) the average NDVI in area a during the 12 months before date d; (9) the average one-month rainfall anomaly in area a for the three months before date d; (10) the average NDVI anomaly in area a for the three months before date d; (11) the difference between the number of reported fatalities in violence against civilians events in area a during the three months preceding date d (minus a 14-day lag) and the three months prior; (12) the most-recent population density estimate for area a available at date d. The second model has the same dependent variable and all the same independent variables except the first one (that is, the last-available prevalence of insufficient food consumption). In the third model, the prevalence of people using crisis or above-crisis food-based coping in a given first-level administrative area a at a given date d is the dependent variable, and the independent variables are the same as above, except the first variable, which is instead the last-available prevalence of people using crisis or above-crisis food-based coping in area a at a time before d (and before the start of the near-real time monitoring, if this is in place in area a). Finally, in the fourth model, the dependent variable is the same as in the third, and the independent variables are all the variables (2) to (12).

For each model, the following procedure was performed. The data were split into two parts following their temporal ordering: data until 31 May 2021 (corresponding to ~85% of the data) were used for training and validation, and the remaining two months (~15% of the data) for testing. To tune the model hyper parameters, a walk-forward validation approach was used: four folds were created, each covering one month of data, from February through May 2021, and for each fold, the training set was composed of all the older data up to the end of the previous month. The tuned hyper parameters and the explored values are listed in Supplementary Table 2. The chosen combination of hyper parameters is the one leading to the smallest difference between the average R^2 on the folds used as training set and the average R^2 on the folds used as validation. We opted for this criterion to favour models where the performance on the test is the most similar to the performance on the training set because large differences are often an indication of overfitting. Once the hyper parameters are selected, $N_b = 100$ models are fitted on samples with replacement of the training and validation set (that is, bootstrapping), and the test set is used to evaluate the model's performance. For each observation in the test set, N_b predictions are generated (one per bootstrapped model), and the median value is then used to calculate the model performance metrics, that is, the coefficient of determination R^2 and the mean absolute error (MAE), which are reported in Table 1 and Fig. 1. Supplementary Figs. 152–155 show that convergence for both metrics is reached within 100 bootstraps.

SHAP values. SHAP (SHapley Additive exPlanations)[28] is a framework recently proposed to interpret predictions made by often complex black box machine learning algorithms. SHAP unifies other methods (Lime, DeepLift and so on), and for tree-based models, it allows for writing a prediction as the sum of a baseline value and each feature's contribution[29]:

$$y = \mathrm{f}(x) = \phi_0 + \sum_{i=0}^{M} \phi_i(x) \tag{1}$$

SHAP values for tree-based models such as XGboost have been shown to improve on other local tree explanations, such as visualizing the decision tree, which is not

feasible for tree ensembles, and on model-agnostic local explanations, which are computationally expensive if explaining large datasets[29]. Moreover, SHAP local explanations can be used as building blocks for global explanations, as shown in Supplementary Fig. 1, where we take the mean absolute value of SHAP values across all data points to build a global feature-importance ranking. In this study, we use the Python open-source implementation of the TreeSHAP algorithm[56].

Explaining the single prediction. SHAP values represent each feature's contribution towards the model prediction, and their absolute value can therefore be interpreted as each feature's importance. This method improves on widely used global feature-importance methods such as split-based or gain-based measures, as it allows computation of prediction-specific feature importance. As detailed in previous sections, each of our four models actually consists of $N_b = 100$ different models fitted on different samples (with replacement) of the training data, of which we report the median prediction and confidence interval. To determine the importance of a feature, we then take each feature's median SHAP value across the N_b bootstrapped models. Convergence checks are reported in Supplementary Figs. 156–157.

In this study, predictions are originally obtained at the spatial resolution of first-level administrative units, but they can then be aggregated to display results at the national level, too. To determine feature importance at the country level, we average the SHAP values across all first-level administrative units in a country by weighting each value according to the unit's population. This can be easily interpreted: a SHAP value corresponds to the change in prevalence with respect to the baseline due to one feature. By performing a population-based weighted average, we are computing the change in number of people due to that feature. This operation then sums the number of people across all units and divides it by the country population to return the change in national prevalence. The same operation is also performed on the model baseline. Note that this also allows us to combine predictions coming from areas with and without a previous value of the target variable, even if the underlying models use slightly different sets of features.

Explaining trend changes. This method allows us to compute the feature-importance ranking for a given area in a given day, explaining which features were the most important and how they contributed to the final prediction. However, predictions for the same area can produce a trend in time, which, in turn, can show changes in prevalence due to changes in the input variables. We extend the SHAP framework to explain which features are responsible in determining changes in predicted trends.

Let us take two predictions y^{t1} and y^{t2} corresponding to the same area but two different dates. Following equation (1), we can write the trend change $y^{t2} - y^{t1}$ in terms of the change χ_i in SHAP values relative to each of the M features:

$$y^{t2} - y^{t1} = \sum_{i=0}^{M} \left(\phi_i(x^{t2}) - \phi_i(x^{t1}) \right) = \sum_{i=0}^{M} \chi_i(x^{t1}, x^{t2}) \qquad (2)$$

The features with largest associated SHAP value change are the ones that determined the trend change. Moreover, the sign of the change χ_i also tells us whether that feature has caused an increase or decrease in the prevalence, that is, a deterioration or improvement in the food security situation.

Note that equation (2) is exact when considering a single model for a single first-level administrative area but is only an approximation when considering median SHAP values, as previously mentioned. It is also important to note that this method can give apparently misleading indications due to nonlinear interactions between features. For instance, a feature that does not change value between the two dates can be the one with the largest SHAP value change (that is, determining the trend change). This happens because other features change value, thus changing its relative importance in the two predictions. One could overcome this limitation by computing SHAP interaction values[29], but the computation is not feasible when dealing with our sample size, that is, 400 models (100 bootstrapped iterations per model) and daily computations. Moreover, this would imply dealing with an order of 50 (number of features squared, divided by 2) different interactions, which would greatly undermine our effort to produce explainable predictions for decision-makers.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The data that support the findings of this study are available in a Zenodo repository: https://zenodo.org/record/6953981.

Code availability
The code used to generate the results reported in this study is available in a Zenodo repository: https://zenodo.org/record/6953981.

References

1. *Food Security Analysis* (World Food Programme, 2022); https://www.wfp.org/food-security-analysis
2. Blumenstock, J., Cadamuro, G. & On, R. Predicting poverty and wealth from mobile phone metadata. *Science* **350**, 1073–1076 (2015).
3. Jean, N. et al. Combining satellite imagery and machine learning to predict poverty. *Science* **353**, 790–794 (2016).
4. Steele, J. E. et al. Mapping poverty using mobile phone and satellite data. *J. R. Soc. Interface* **14**, 20160690 (2017).
5. Pokhriyal, N. & Jacques, D. C. Combining disparate data sources for improved poverty prediction and mapping. *Proc. Natl Acad. Sci. USA* **114**, E9783–E9792 (2017).
6. Engelmann, G., Smith, G. & Goulding, J. The unbanked and poverty: predicting area-level socio-economic vulnerability from m-money transactions. In *2018 IEEE International Conference on Big Data* 1357–1366 (IEEE, 2018).
7. Sheehan, E. et al. Predicting economic development using geolocated Wikipedia articles. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 2698–2706 (ACM, 2019).
8. Kondmann, L., Haeberle, M. & Zhu, X. X.Combining Twitter and Earth observation data for local poverty mapping. In *NeuRIPS Machine Learning for the Developing World Workshop* 1–5 (NeurIPS, 2020).
9. Fatehkia, M. et al. Mapping socioeconomic indicators using social media advertising data. *EPJ Data Sci.* **9**, 22 (2020).
10. IFAD, UNICEF, WFP & WHO. *The State of Food Security and Nutrition in the World 2021. Transforming Food Systems for Food Security, Improved Nutrition and Affordable Healthy Diets for All* (FAO, 2021).
11. FSIN. *Global Report on Food Crises 2020* (FSIN, 2020); https://www.wfp.org/publications/2020-global-report-food-crises
12. Emergency Operations Division *WFP Global Operational Response Plan: Update #3–November 2021* (WFP, 2021); https://www.wfp.org/publications/wfp-global-operational-response-plan-update-3-november-2021
13. *Declaration of the World Summit on Food Security* (FAO, 2009).
14. Vaitla, B. et al. The measurement of household food security: correlation and latent variable analysis of alternative indicators in a large multi-country dataset. *Food Policy* **68**, 193–205 (2017).
15. Lobell, D. B. et al. Prioritizing climate change adaptation needs for food security in 2030. *Science* **319**, 607–610 (2008).
16. Zufiria, P. J. et al. Identifying seasonal mobility profiles from anonymized and aggregated mobile phone data. Application in food security. *PloS ONE* **13**, e0195714 (2018).
17. Napoli, M. *Towards a Food Insecurity Multidimensional Index (FIMI)*. MSc thesis, Roma Tre Univ. (2011).
18. Caccavale, O. M. & Giuffrida, V. The proteus composite index: towards a better metric for global food security. *World Dev.* **126**, 104709 (2020).
19. Okori, W. & Obua, J. Machine learning classification technique for famine prediction. In *Proceedings of the World Congress on Engineering*, Vol. 2 4–9 (Citeseer, 2011).
20. Andree, B. P. J., Chamorro, A., Kraay, A., Spencer, P. & Wang, D. *Predicting Food Crises* (World Bank, 2020).
21. Wang, D., Andree, B. P. J., Chamorro, A. F. & Girouard Spencer, P. *Stochastic Modeling of Food Insecurity* (World Bank, 2020).
22. Lentz, E., Michelson, H., Baylis, K. & Zhou, Y. A data-driven approach improves food insecurity crisis prediction. *World Dev.* **122**, 399–409 (2019).
23. Westerveld, J. J. et al. Forecasting transitions in the state of food security with machine learning using transferable features. *Sci. Total Environ.* **786**, 147366 (2021).
24. Deléglise, H. et al. Food security prediction from heterogeneous data combining machine and deep learning methods. *Expert Syst. Appl.* **190**, 116189 (2021).
25. Chen, T. & Guestrin, C. Xgboost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* Pages 785–794 (ACM, 2016).
26. Funk, C. et al. Recognizing the famine early warning systems network: over 30 years of drought early warning science advances and partnerships promoting global food security. *Bull. Am. Meteorol. Soc.* **100**, 1011–1027 (2019).
27. *HungerMap LIVE* (WFP, 2022); https://hungermap.wfp.org/
28. Lundberg, S. M. & Lee, S.-I. in *Advances in Neural Information Processing Systems*, Vol. 30 4765-4774 (eds Guyon, I. et al.) (Curran Associates Inc., 2017).
29. Lundberg, S. M. et al. From local explanations to global understanding with explainable AI for trees. *Nat. Mach. Intell.* **2**, 2522–5839 (2020).
30. *HungerMap LIVE: Global Insights and Key Trends* (WFP, 2022); https://static.hungermapdata.org/insight-reports/latest/global-summary.pdf
31. Blumenstock, J. E. Estimating economic characteristics with phone data. *AEA Pap. Proc.* **108**, 72–76 (2018).

32. Singh, S., Nourozi, S., Acharya, L. & Thapa, S. Estimating the potential effects of COVID-19 pandemic on food commodity prices and nutrition security in Nepal. *J. Nutr. Sci.* **9**, E51 (2020).

33. Jelilov, G., Iorember, P. T., Usman, O. & Yua, P. M. Testing the nexus between stock market returns and inflation in Nigeria: does the effect of COVID-19 pandemic matter? *J. Public Aff.* **20**, e2289 (2020).

34. Ogunleye, A. & Wang, Q.-G. Xgboost model for chronic kidney disease diagnosis. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **17**, 2131–2140 (2019).

35. Zhang, Y. & Hamori, S. Forecasting crude oil market crashes using machine learning technologies. *Energies* **13**, 2440 (2020).

36. Smith, G., Mansilla, R. & Goulding, J. Model class reliance for random forests. In Ao, I.S. et al. (eds) *Adv. Neural Inf. Process. Syst.* 33 22305-22315 Newswood Limited (2020). http://www.iaeng.org/publication/WCE2011/#:~:text=The%20WCE%202011%20takes%20place,engineering%20and%20computer%20science%20subjects

37. Kumar, I. E., Venkatasubramanian, S., Scheidegger, C. & Friedler, S. Problems with Shapley-value-based explanations as feature importance measures. In *International Conference on Machine Learning* 5491–5500 (PMLR, 2020).

38. *Food Consumption Analysis Calculation and Use of the Food Consumption Score in Food Security Analysis* (WFP, 2008); https://documents.wfp.org/stellent/groups/public/documents/manual_guide_proced/wfp197216.pdf

39. *The Coping Strategies Index: Field Methods Manual* (WFP, 2008); https://documents.wfp.org/stellent/groups/public/documents/manual_guide_proced/wfp211058.pdf

40. Bowling, A. Mode of questionnaire administration can have serious effects on data quality. *J. Public Health* **27**, 281–291 (2005).

41. *The World Food Programme's Real-Time Monitoring Systems: Approaches and Methodologies* (WFP, 2021); https://docs.wfp.org/api/documents/WFP-0000135070/download/

42. *Technical Manual Version 3.0. Evidence and Standards for Better Food Security and Nutrition Decisions* (IPC, 2019).

43. Chan, J. Y.-L. et al. Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics* **10**, 1283 (2022).

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

45. *WFP Food Price Forecasting and Alert for Price Spikes* (WFP, 2022); https://dataviz.vam.wfp.org/economic_explorer/price-forecasts-alerts

46. *Trading Economics* (Trading Economics, 2022); https://tradingeconomics.com/

47. WFP Economic Explorer (WFP, 2022); https://dataviz.vam.wfp.org/economic_explorer/macro-economics/exchange_rate

48. *WFP Seasonal Explorer* (WFP, 2022); https://dataviz.vam.wfp.org/seasonal_explorer/rainfall_vegetation/visualizations

49. Funk, C. C. et al. A quasi-global precipitation time series for drought monitoring. *US Geol. Surv. Data Series* **832**, 4 (2014).

50. *MODIS Vegetation Index Products (NDVI and EVI)* (NASA, 2022); https://modis.gsfc.nasa.gov/data/dataprod/mod13.php

51. *Data Export Tool* (ACLED, 2022); https://acleddata.com/data-export-tool/

52. Dowd, C. in Cahill, B.H. and Lawton, J. (eds) *A Skein of Thought* 119–132 (Fordham Univ. Press, 2020).

53. *Food and Agriculture Organization Corporate Statistical Database* (FAOSTAT, 2022); http://www.fao.org/faostat

54. *Gridded Population of the World version 4 (GPWv4): Population Density, Revision 11* (NASA, 2018); https://doi.org/10.7927/H49C6VHW

55. Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).

56. Lundberg, S. *SHAP Documentation* (Read the Docs, 2018); https://shap.readthedocs.io/

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43016-022-00587-8.

Correspondence and requests for materials should be addressed to Elisa Omodei.

Peer review information *Nature Food* thanks Erin Lentz, Gavin Smith and James Goulding for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s): Elisa Omodei

Last updated by author(s): Aug 3, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used by the authors.
Data analysis	Custom Python 3.7 code making use of the following libraries: numpy, scipy, pandas, scikit-learn, xgboost, shap, matplotlib and seaborn.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available in a Zenodo repository: https://zenodo.org/record/6953981

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences   ☒ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Predictive modeling based on quantitative data.
Research sample	Representative sample made of 37926 measurements of the first-level administrative unit prevalence of people with insufficient food consumption collected across 78 countries between 2006 and 2021, and of 35894 measurements of the first-level administrative unit prevalence of people using crisis or above crisis food-based coping strategies collected across 41 countries between 2013 and 2021.
Sampling strategy	Sampling is performed by randomly selecting, for each first-level administrative unit, one observation per month only.
Data collection	Data was collected by governments and international organizations like the World Food Programme (WFP), the Food and Agriculture Organization (FAO) and the World Bank through face-to-face surveys and remote mobile phone surveys (e.g. computer-assisted telephone interviews, CATI).
Timing	From 2006 to 2021.
Data exclusions	No data was excluded from the analysis.
Non-participation	Not applicable.
Randomization	Not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Human research participants
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging