Food Policy 84 (2019) 77–91

Contents lists available at ScienceDirect

Food Policy

journal homepage: www.elsevier.com/locate/foodpol

Alternatives to calorie-based indicators of food security: An application of machine learning methods

Marup Hossain[a,*], Conner Mullally[b], M. Niaz Asadullah[c,d,e,f,g]

[a] Research and Impact Assessment Division, International Fund for Agricultural Development, Italy
[b] Food and Resource Economics Department, University of Florida, United States
[c] Faculty of Economics and Administration, University of Malaya, Malaysia
[d] Department of Economics, University of Reading, UK
[e] Centre on Skills, Knowledge and Organisational Performance (SKOPE), University of Oxford, UK
[f] IZA Institute of Labor Economics, Germany
[g] School of Education, Environment and Development (SEED), University of Manchester, UK

ARTICLE INFO

Keywords:
Food security
Poverty
Dietary diversity
Program targeting
Machine learning

JEL Codes:
C52
I32
O12
Q18

ABSTRACT

Identifying food insecure households in an accurate and cost-effective way is important for targeted food policy interventions. Since predictive accuracy depends partly on which indicators are used to identify food insecure households, it is important to assess the performance of indicators that are relatively easy and inexpensive to collect yet can proxy for the "gold standard" food security indicator, calorie intake. We study the effectiveness of different variable combinations and methods in predicting calorie-based food security among poor households and communities in rural Bangladesh. We use basic household information as a benchmark set for predicting calorie-based food security. We then assess the gain in predictive power obtained by adding subjective food security indicators (e.g., self-reported days without sufficient food), the dietary diversity score (DDS), and the combination of both sets to our model of calorie-based food security. We apply machine learning as well as traditional econometric methods in estimation. We find that the overall predictive accuracy rises from 63% to 69% when we add the subjective and DDS sets to the benchmark set. Our study demonstrates that while alternative indicators and methods are not always accurate in predicting calorie intake, DDS related indicators do improve accuracy compared to a simple benchmark set.

1. Introduction

Although access to food has improved globally over the last three decades, a substantial proportion of the population in low-income countries continues to suffer from hunger and malnutrition (FAO, 2014). This is particularly true in developing countries like Bangladesh, where despite tripling of rice production, millions of households remain vulnerable to food insecurity resulting from food price hikes, natural disasters, and other adverse events (Hossain, 2013; Jamora and von Cramon-Taubadel, 2016; Del Ninno et al., 2003; Faisal and Parveen, 2004; Szabo et al., 2016). The effectiveness of social safety net schemes designed to address food insecurity, poverty, or vulnerability crucially depends on a clear framework for identifying targeted households (Barrett, 2010; Hyman et al., 2005; Webb et al., 2006). Even in a situation where there is a consensus around poverty measurement (e.g., a well-defined national poverty line), prevalence of poverty and food insecurity do not always overlap. Consequently, food distribution programs that target the poor can penalize food insecure households (Suryanarayana and Silva, 2007). Accurate classification of households with respect to their food security status can also lead to better targeting of vulnerable groups and improved evaluation of public policy interventions (Maxwell et al., 1999).

In food security analysis, calorie intake is considered the "gold standard" measure (Chung, 1997; Maxwell et al., 1999). But measurement of calorie intake using a food recall module or food diaries is time-consuming, suffers from reporting bias, and has high administrative costs (Hoddinott, 1999). In contexts where time and cost are important considerations (e.g., delivery of disaster relief), identifying food insecure households using calorie intake may not be the right approach. It is therefore important to assess whether there are alternative indicators that are relatively inexpensive to collect yet can adequately proxy for calorie intake in predicting food insecurity. Furthermore, the performance of alternative indicators should be assessed using the best possible prediction methods. In particular, recent developments in machine

* Corresponding author at: Research and Impact Assessment Division, International Fund for Agricultural Development, Via Paolo di Dono 44, 00142 Rome, Italy.
E-mail addresses: md.hossain@ifad.org (M. Hossain), connerm@ufl.edu (C. Mullally), m.niaz@um.edu.my (M.N. Asadullah).

https://doi.org/10.1016/j.foodpol.2019.03.001
Received 14 September 2018; Received in revised form 22 January 2019; Accepted 2 March 2019
Available online 18 March 2019
0306-9192/

learning (ML) may outperform traditional estimation methods, thereby reducing exclusion and inclusion errors in program targeting.

In this study, we test the ability of different variable combinations in predicting calorie-based measures of food security using ML methods. We begin by predicting caloric based indicators using a benchmark predictor set that includes household level variables requiring little effort in data collection, for example, characteristics of the head of household, infrastructure access, and asset holdings (hereafter, the benchmark set). We then estimate the marginal gain in predictive accuracy obtained by adding two different predictor sets to the model, one at a time as well as in combination. The first set is a “subjective” predictor set that includes three self-reported indicators of food shortages. The second set is based on dietary diversity (i.e. the number of food groups consumed) that includes women’s dietary diversity score (WDDS) or the number of food groups consumed by adult women, other members’ above 2 years of age dietary diversity score (ODDS) or the number of food groups consumed by members other than adult women, the household dietary diversity score (HDDS), and the food consumption score (FCS), a version of dietary diversity that weights foods by nutritional value. We refer to this set as the DDS set. Finally, we estimate the marginal gain of adding both the subjective and DDS sets to the benchmark set. In addition to assessing different indicator sets, we test the relative performance of ML and non-ML methods in predicting calorie-based food security. Our ML methods are the random forest (RF) and extreme gradient boosting (XGBoost) algorithms, while our non-ML method combine either ordinary least squares (OLS) or logistic regression with an out-of-sample prediction algorithm (non-ML method). We carry out our analysis at the household and community levels in order to compare the performance of different methods when predicting food security at distinct levels of aggregation.

The study context is Bangladesh, where the government places a high priority on ensuring food security, allocating nearly 2.2% of the gross domestic product (GDP) to safety net and social protection programs (Hossain, 2013). Our data include a nationally representative rural sample of 6,427 households. The households were interviewed as a part of the 2015 Bangladesh Integrated Household Survey (BIHS), administered by the International Food Policy Research Institute (IFPRI). We predict two calorie-based indicators: (1) whether the daily calorie intake in a household is less than the required calorie intake level per adult male equivalent (AME) (calorie poor, hereafter) and (2) the AME-adjusted daily per capita calorie intake (calorie intake, hereafter). We use RF and XGBoost to predict household calorie-poor status and calorie intake level. To compare the marginal gain in prediction obtained when adding alternative predictors to the benchmark set, we estimate the accuracy, sensitivity, and specificity of prediction of being calorie poor.[1] For calorie intake, we assess predictive accuracy by comparing the square root of the sum of squared prediction errors divided by the standard deviation of the target variable, i.e. the normalized root mean square error (NRMSE). We also use the variable-of-importance feature of the RF method to identify the most important predictors of calorie indicators.

When we add the DDS set to the benchmark set, we find that overall accuracy in correctly classifying households as calorie poor or not calorie poor rises from 63% to 69%, while the NRMSE with respect to calorie intake improves slightly from 0.93 to 0.89. The marginal gain in predictive accuracy obtained when adding subjective indicators to the benchmark set is negligible. The benchmark set accounts for 91% and 94% of total prediction accuracy for the RF and XGBoost methods, respectively, when predicting calorie poverty at the household level. Despite the importance of the benchmark set, we find that FCS and HDDS are two of the five most important predictors of calorie indicators according to the importance weights generated by RF. In contrast, none of the subjective indicators are especially important. Our selected non-ML and ML methods perform quite similarly in terms of overall prediction accuracy, perhaps because the total number of predictors is small, limiting the complexity of candidate prediction models. In the case of the non-ML method, both the subjective and DDS sets improve prediction accuracy, but the benchmark set again accounts for as much as 90% of total accuracy in prediction.

While some food security interventions may blanket all areas of a country, in other cases implementers may be forced to limit the geographic scope of an intervention. Therefore an important initial stage of any household-level targeting is an analogous targeting at the aggregate level (e.g., community or regional level). To compare the performance of our predictor sets and methods at a more aggregate level, we predict community-level calorie intake, aggregated from household-level data, using community-level observable characteristics and DDS and subjective food security indicators, also aggregated from the households level data. We find that neither the DDS set nor the subjective set improves prediction accuracy of calorie indicators drastically.

Our study complements as well as adds new evidence to the literature on food security measurement and targeting. We show that a set of predictors comprised of easy-to-observe information does nearly as well in predicting calorie-based indicators of food security as harder-to-collect information like subjective food security and dietary diversity. The accuracy of the benchmark set in correctly classifying households as calorie-poor or not ranges from 63% to 64%. But whether our benchmark set should be viewed as an adequate proxy for “gold standard” calorie-based indicators will be context specific. Policymakers will have to weigh the relatively low collection cost and speed of the benchmark set against the accuracy of calorie-based indicators.

Second, while several studies evaluate whether DDS, self-reported food security, and other alternative indicators can accurate predict food security, few have conducted comparative analyses. Importantly, comparisons are based on correlation indexes, contingency tables, or OLS regressions in many of these studies (e.g., Chung, 1997; Hoddinott, 1999; Maxwell et al., 1999; Hoddinott and Yohannes, 2002; Wiesmann et al., 2009). Other studies compare alternative indicators of food security without considering predictive accuracy with respect to calorie-based indicators (e.g. Headey and Ecker, 2013; Maxwell et al., 2014; Vaitla et al., 2017). In contrast, we compare the performance of different predictor sets when paired with sophisticated ML and non-ML methods, and assess prediction accuracy with respect to calorie indicators.

Lastly, we contribute to the nascent literature on using ML methods to measure dimensions of poverty. Previous examples include predicting local economic outcomes using satellite data (Jean et al., 2016) and crop yield using local weather conditions (Dahikar and Rode, 2014). Some studies use big data to predict various outcomes, for example, individual level wealth using cell phone data (Blumenstock et al., 2015), poverty using mobile phone and satellite data (Steele et al., 2017), and block-level income using images from Google Street View (Glaeser et al., 2018). In the food security literature, Knippenberg et al. (2018) apply ML methods to find predictors of future food insecurity in Malawi. They show that previous food insecurity, living in floodplain areas, and distance to drinking water are important predictors of future food insecurity. In our study, we apply ML methods to examine whether we can accurately predict food security using only basic observable factors and what are the marginal contributions of alternative food security indicators in predicting calorie indicators.

The rest of the paper is organized as follows. Section 2 provides the key definitions and briefly reviews the literature on food security

[1] *Accuracy* refers to overall prediction accuracy, e.g. the proportion of units correctly classified with respect to food security status. *Sensitivity* refers to the proportion of food insecure households correctly identified as such. Poor sensitivity implies a large number of inclusion errors, i.e. identifying households as food insecure when in fact they are not. *Specificity* refers to the proportion of food secure households correctly predicted to be food secure. Poor specificity implies a large number of exclusion errors, i.e. identifying households as food secure when in reality they are not.

indicators. Section 3 presents details on data and measurement issues, followed by the summary statistics in Section 4 and methodology in Section 5. Section 6 presents the findings of our study. In Section 7, we present a comparison of ML and non-ML methods. Section 8 examines a community level targeting, and, finally, Section 9 contains a conclusion.

2. Literature review

2.1. Food security definition and indicators

Food security is "a situation that exists when all people, at all times, have physical, social and economic access to sufficient, safe and nutritious food that meets their dietary needs and food preferences for an active and healthy life" (FAO, 1996). This definition consists of three hierarchical pillars: availability, access, and utilization (Barrett, 2010). Availability is necessary for food security, access reflects individual capacity, and utilization confirms whether households can make effective use of their access.

Food production or purchases and food consumption are the two most widely used indicators of food security, although both are subject to measurement problems. Other common food security indicators are DDS, FCS, Household Food Insecurity Access Scale (HFIAS), coping strategy index (CSI), anthropometric measures, and access to infrastructure and services (Carletto et al., 2013). Among all the food security indicators, calorie intake has been cited as the "gold standard" indicator of food security (Chung, 1997; Maxwell et al., 1999). The detary diversity score (DDS), measured as the number of different foods or food groups consumed by an individual or household over a reference period (Swindale and Bilinsky, 2006; Wiesmann et al., 2009), is commonly used as a proxy of calorie intake. DDS has gained prominence over other indicators because of its close relationship with household access to food and calorie availability (Hoddinott and Yohannes, 2002), household nutrition status (Hatløy et al., 2000; Arimond and Ruel, 2004), and socioeconomic status (Hatløy et al., 2000). However, DDS lacks a quantity dimension, as it varies over any given reference period. FCS is an improved version based on DDS, food frequency, and nutritional importance of food items. Self-reported indicators are becoming common in recent studies because they can capture the psychological and seasonal dimensions of food security. However, self-reported indicators also have limitations such as cross-sectional validity (Deitchler et al., 2010) and sensitivity with the ordering of questions (Headey, 2013).

2.2. Calorie intake and validity of alternative indicators

A typical way of measuring calorie intake is using information of different foods consumed by a household and converting into calories using a calorie conversion table. But it is often difficult to get an accurate measurement of calorie intake because of poor data quality and long time and high-cost requirements in data collection (Hoddinott and Yohannes, 2002; Carletto et al., 2013). Using a comparison of different food security indicators in terms of data collection cost, time requirement, skill requirement, and susceptibility to misreporting, Hoddinott (1999) argues that individual-level calorie intake data requires high costs, time, and skills, but has a low probability of misreporting. Household level calorie intake, on the other hand, requires moderately high costs, time, and skills, but has a high probability of misreporting.

Despite accuracy concerns, basing food security assessment on calorie intake is still viewed as the gold standard, as mentioned above. But given the cost concerns associated with calorie intake data, several studies attempt to identify alternative indicators of calorie intake that minimize the above concerns while having a strong correlation with calorie intake. Hoddinott (1999) uses correlation coefficients, contingency tables, and ordinary least squares (OLS) regression to test the relationship between calorie intake and DDS in Mali. Using a cross-sectional data set from ten countries, Hoddinott and Yohannes (2002) apply OLS regression to examine the relationship between calorie intake and DDS as well. Wiesman et al. (2009) find that the correlation between FCS and calorie intake is positive and statistically significant in Haiti, Sri Lanka, and Burundi. Several other studies include comparative analyses of alternative indicators without explicitly exploring their relationship with calorie intake. For example, Maxwell et al. (2014) compare seven alternative indicators of food security using correlation coefficients and cross-classification. Vaitla et al. (2017) use correlation coefficients, categorical concordance, and factor analysis to compare FCS, HDDS, CSI, and household hunger scale (HHS, an indicator to measure household hunger in food insecure areas that is valid and comparable across cultures and settings (Ballard et al., 2011)).

From the literature review section, it is clear that there is an ongoing effort to find indicators that can proxy calorie intake and other nutritional requirements while reducing data costs and measurement error. While testing the suitability of an alternative indicator against calorie intake or other food security indicators, previous studies are limited to basic estimation strategies like correlation coefficients, contingency tables, and OLS regression.

3. Data and measurement

We use data from the "Bangladesh Integrated Household Survey (BIHS)", a nationally representative survey of rural Bangladesh. The International Food Policy Research Institute (IFPRI) administers the BIHS survey as part of the Feed the Future (FTF) program in Bangladesh. The BIHS survey follows two-stage stratified sampling. In the first stage, using the sample frame developed for the 2001 population census of Bangladesh, IFPRI selects 325 primary sampling units (PSU). In the second stage, based on the sampling weight of the population census of 2011, IFPRI determines the required households from each PSU. We use data from the second round of BIHS conducted from January to June 2015. The second round of BIHS covers 6,427 households. For detailed information on the survey and data collection, see Ahmed et al. (2013).

We estimate calorie intake from household consumption information of different food items acquired through own production, purchases, or any other source over the last seven days. Enumerators collected detailed information on the quantity consumed of 321 pre-listed food items from each household. We then use the "2013 Food Composition Table" for Bangladesh, developed by the Institute of Nutrition and Food Science at the University of Dhaka (Shaheen et al., 2013), to convert food consumption amount into calories. We exclude some items in our estimation that have no caloric value, e.g., salt, water, tea, cigarettes, and betel leaf. We do not consider food consumed away from home due to inadequate data. We expect that the absence of food consumed away from home will not cause a significant underestimation in our study because the share of food consumed away from home in total food consumption is below 2% in rural Bangladesh (Reardon et al., 2014). We estimate per capita calorie intake by dividing household daily calorie intake by household size in adult male equivalents (AMEs). Following Waid et al. (2017), we use adult equivalency weights based on the assumption of a moderate activity level. We define a household as calorie poor if the daily calorie intake level of a household is less than the AME-adjusted required calorie intake as reported by Waid et al. (2017). Table A1 contains the weights and calorie requirements used in this study.

The benchmark set includes education, age, and sex of household head; religion, maximum education, household size, working-age males, and working-age females in a household; number of household members who work in labor, service, non-farm self-employment, and farm self-employment; household, livestock, and agricultural assets; other land and homestead land values; whether the dwelling has bricked walls, floor, and roof; and electricity connection, and access to sealed latrine. The DDS set includes four different indicators estimated at the household level using consumption information for 17 food groups and a seven-day recall period: (1) DDS for women aged 14–49 years (WDDS), (2) DDS for other members above 2 years of age (ODDS), (3) DDS for the entire household (HDDS), and (4) food consumption score (FCS). Using the example of Swindale and Bilinsky (2006), we convert the 17 food groups included in the survey into the following 12 groups: (1) cereals, (2) root and tubers, (3) vegetables, (4) fruits, (5) meat and poultry, (6) eggs, (7) fish and seafood, (8) pulses/legumes/nuts, (9) milk and milk products, (10) oil/fats, (11) sugar/honey, and (12) miscellaneous food items. We estimate WDDS, ODDS, and HDDS as the total number of food items consumed out of the above 12 groups in the last seven days. Finally, we estimate FCS as the weighted sum of the frequency of consumption of the 12 food groups over the last seven days. In FCS estimation, we use weights for different groups following the technical guidance sheet of the World Food Program (2008).

The subjective set consists of three different self-reported indicators: (1) a binary variable taking a value of one if the household reported going a full day with no food at any time in the last four weeks (no-food, hereafter), (2) a binary variable equal to one if any household member went to sleep at night while hungry because of food unavailability in the household in the last four weeks (hungry-night, hereafter), and (3) a binary variable equal to one if any household member went without eating for 24 h because of food unavailability in the last four weeks (hungry-day-night, hereafter).

4. Summary statistics

Table 1 summarizes household characteristics and food security status, as well as differences by household calorie-poor status. Around 19% of households are headed by a woman. A large portion of household heads have no formal education, and average education among household heads is 3.5 years. Average household size is 4.96 members, with working-age female members slightly outnumbering male members of the same age group. Around 60% of the households have access to electricity, while 45% of households use a sealed sanitary latrine.

AME adjusted per capita daily calorie intake level is 2,650 Kcal for a typical household in rural Bangladesh. It is important to note that the average calorie intake level may vary depending on the conversion factors and the number of food items used to convert the reported consumption amounts into calories. For example, FAO (2014) shows that the average daily per capita calorie intake is 2,318 Kcal using the HIES data of Bangladesh. Households consume an average of 7 out of 12 food groups (HDDS), with women consuming an average of 3.75 food groups (WDDS) and other members (aged at least two years) of age consuming 5.20 food groups (ODDS). The average FCS level is 60 and

Table 1
Summary statistics.

	Mean (1)	Mean (Calorie poor = 1) (2)	Mean (Calorie poor = 0) (3)	Diff. (2–3) (4)	P value (5)
Benchmark set					
Female HH head (yes = 1)	0.19	0.16	0.23	−0.07	0.00
Education of HH head (years)	3.50	3.24	3.95	−0.72	0.00
Age of HH head (years)	45.59	45.18	46.31	−1.13	0.00
Maximum education in HH	7.13	7.12	7.16	−0.04	0.68
Religion (non-Muslim = 1)	0.12	0.13	0.11	0.02	0.02
Household size	4.96	5.12	4.68	0.45	0.00
Male working age member	1.22	1.31	1.06	0.25	0.00
Female working age member	1.47	1.50	1.41	0.10	0.00
No. of HH members work in different occupations					
Labor	0.28	0.33	0.18	0.15	0.00
Service	0.11	0.12	0.09	0.03	0.00
Non-farm self-employment	0.39	0.44	0.31	0.13	0.00
Farm self-employment	0.78	0.73	0.87	−0.14	0.00
Infrastructure					
Brick wall (yes = 1)	0.21	0.19	0.25	−0.06	0.00
Brick roof (yes = 1)	0.05	0.04	0.07	−0.03	0.00
Brick floor (yes = 1)	0.16	0.13	0.21	−0.08	0.00
Have electricity (yes = 1)	0.58	0.56	0.63	−0.07	0.00
Use sealed latrine (yes = 1)	0.45	0.43	0.49	−0.06	0.00
Asset value at replacement cost (USD/PPP)					
Household asset	1716.30	1463.16	2161.42	−698.26	0.00
Livestock asset	618.78	520.32	791.92	−271.61	0.00
Agricultural asset	148.70	128.08	184.94	−56.86	0.01
Homestead land	10,755	9628.82	12735.28	−3106.46	0.00
Other land	26,905	21579.61	36269.47	−14689	0.00
Food insecurity indicators					
Calorie (daily/per capita)	1905.43	1579.24	2478.98	−899.74	0.00
Calorie (daily/AME)	2763.81	2219.94	3720.12	−1500.18	0.00
Subjective food security set					
No-food	0.11	0.14	0.06	0.07	0.00
Hungry-night	0.05	0.06	0.03	0.03	0.00
Hungry-day-night	0.02	0.02	0.01	0.01	0.00
Dietiary diversity score (DDS) set					
WDDS	3.75	3.64	3.95	−0.31	0.00
ODDS	5.20	5.03	5.49	−0.46	0.00
HDDS	7.04	6.86	7.35	−0.49	0.00
FCS	64.18	60.21	71.17	−10.96	0.00

Note: N = 6,427. Calorie poor is defined as whether the daily calorie intake in a household is less than the required calorie intake level adjusted for adult male equivalent (AME). In our sample, 64% of households are calorie poor.

71 for calorie-poor and calorie-non-poor households, respectively, with a sample average of 64. 11% of households report having gone 24 hours with no food at sometime in the last four weeks (no food). 5% of households experienced a situation in the past 30 days where a member went to bed hungry due to the unavailability of food (hungry night). Only 2% of households report a member not being able to eat for a whole day and night in the past 30 days due to the unavailability of food (hungry-day-night). Columns 3–5 show differences in household characteristics by calorie-poor status. Calorie-poor households are clearly worse off compared to calorie secure households with respect to most observed characteristics.

5. Methodology

The goal of our study is to examine the ability of different predictor sets, which are relatively easy to quantify and less costly in data collection as compared to a full food consumption survey module, to predict calorie poverty (i.e. being below the minimum in take per AME) and calorie intake. The ease-of-collection criterion rules out household income and expenditure indicators, for example. Our first set of indicators consists of easily observable and quantifiable characteristics of households (the benchmark set), such as characteristics of the household head and other members, assets, housing indicators, access to sanitation, and use of a sealed latrine. Additionally, we have a set of DDS indicators and a set of subjective indicators, as described above. We predict calorie intake and calorie poverty using the benchmark set. We then assess the gain in predictive accuracy obtained when adding either the DDS set or the subjective set, as well as when adding them sequentially.

We use the RF method as the primary prediction method in our study because of its flexibility and superior performance relative to other prediction methods. RF is an algorithm that can be used for classification of discrete variables or prediction of continuous variables. RF generates a set of “decision trees” using a collection of random variables to predict a target variable. Loosely speaking, each tree grown through the RF method splits observations into groups based on their observed characteristics in a way that minimizes prediction error. When predicting a continuous variable, RF assigns each group member the group average of the outcome as its predicted value. When used to classify observations, RF assigns each observation to the most commonly observed class in its group. Each observation is assigned the average prediction over the many decision trees grown by RF for a continuous outcome or is assigned the class predicted by a majority of decision trees for a discrete variable. The RF method has advantages over other prediction methods in reducing overfitting (Zhang and Ma, 2012). We explain the RF method in detail in the Appendix.

To measure the gain in prediction accuracy attributable to alternative predictor sets, we estimate different performance indicators depending on the nature of the target variable. For the binary target variable (calorie poor), we estimate the overall accuracy in classification of calorie poor status (accuracy), the proportion of truly food insecure households predicted to be food insecure (sensitivity), and the proportion of truly food secure households predicted to be food secure (specificity). Additionally, we use the Kappa estimator, which measures agreement for a categorical variable relative to what would be expected by chance. The Kappa estimator is a more robust measure of overall prediction accuracy for a binary target variable (Kuhn, 2008). For the continuous target variable (calorie intake), we estimate the NRMSE of the predicted ($\hat{Y}_i$) and the observed (Y_i) values of calorie intake as follows:

$$\mathrm{NRMSE} = \frac{\sqrt{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2/N}}{sd(Y_i)} \tag{1}$$

We also use the RF method to discover which variables are most important in predicting the calorie indicators. More specifically, each predictor is randomly permuted so as to remove its correlation with the target variable. Then the RF procedure is repeated, with the resulting predictions of the target variable compared to the original pre-permutation predictions. The importance of the permuted variable in predicting the target variable is captured by the degradation in prediction accuracy caused by the permutation. In Appendix A, we describe the technical details of the variable-of-importance technique.

We also use the XGBoost method to predict our calorie indicators. Boosting methods are similar to RF but use an additive strategy rather than generating multiple trees and averaging the predictions. A single tree is fit to the target variable. A second tree is fit to the residuals generated by comparing the first tree to the target variable, and added to the prediction model. The method continues to fit additional trees, with each subsequent tree fit to the residuals of the model created from the preceding trees. XGBoost is primarily distinguished by the speed with which it searches over possible splits of the data in growing the decision trees. This description necessarily glosses over several details of boosting and XGBoost in particular, and more information is given in the Appendix.

In our data, 64% of households are calorie poor, which raises a class imbalance problem. As is the case with traditional econometric classifiers like logistic regression, supervised ML methods like RF and XGBoost place more emphasis on the dominant class when building a predictive model. The more imbalanced the data are towards a particular class, the worse the predictive performance will be of RF and XGBoost with respect to the less common class. One way to overcome the class imbalance problem is to oversample the minority class or undersample the majority class. Oversampling and undersampling have limitations. For example, the former can lead to overfitting while the latter risks discarding valuable information. To overcome these limitations, we use the Synthetic Minority Over-sampling Technique (SMOTE), which draws a subset of the minority class, then generates additional similar synthetic observations, and finally, adds these new observations to the original data (Chawla et al., 2002).

6. Findings

6.1. Prediction of calorie indicators

Panel A of Table 2 reports prediction results when using the RF method. Columns 1–4 shows accuracy, sensitivity, specificity, and Kappa estimators for calorie-poor as the target variable. We find that the overall accuracy of prediction rises from 63% to 69% with the inclusion of DDS set to the benchmark set. The sensitivity indicator increases from 69% to 75% while the specificity indicator rises from 52% to 60%. The inclusion of the subjective set does not affect any of these indicators drastically. The Kappa estimate also increases from 21 to 33 points with the inclusion of the DDS set. In panel B, we show that results obtained using the XGBoost method are similar to those of RF. Overall accuracy rises from 64% to 68%, while sensitivity, specificity, and Kappa rise from 75% to 78%, 46% to 49%, and 0.21 to 0.29, respectively. The marginal gain in prediction accuracy obtained when using the DDS set is not as large with XGBoost as it is with the RF method. However, the DDS set is still the superior to the subjective set regardless of ML method.

Column 5 shows NRMSE for calorie intake as the target variable. Using the RF method, we find that NRMSE is 0.93 when we use only the benchmark predictor set, and that the NRMSE falls to 0.89 when we include add the DDS set to the model. In the case of the XGBoost method, we find that NRMSE falls from 0.94 to 0.89 with the addition of the DDS set to the benchmark set. Once again, we find little to no gain in the prediction of calorie intake with the addition of the subjective set. In Figs. A1 And A2, we show the distributions of original and predicted calorie intake by different predictor sets, starting from the benchmark set to the set that includes all predictors.

Table 2
Prediction of calorie indicators.

	Calorie poor (yes = 1)				Calorie intake
	Accuracy (1)	Sensitivity (2)	Specificity (3)	Kappa (4)	NRMSE (5)
Panel A: RF					
Benchmark	0.63 (0.61–0.65)	0.69	0.52	0.21	0.93
Benchmark + Subjective	0.63 (0.61–0.66)	0.68	0.55	0.23	0.93
Benchmark + DDS	0.69 (0.67–0.71)	0.75	0.59	0.33	0.89
All	0.69 (0.67–0.71)	0.75	0.60	0.35	0.89
Panel B: XGBoost					
Benchmark	0.64 (0.62–0.66)	0.75	0.46	0.21	0.94
Benchmark + Subjective	0.64 (0.61–0.65)	0.75	0.44	0.20	0.93
Benchmark + DDS	0.68 (0.66–0.70)	0.78	0.49	0.29	0.89
All	0.68 (0.66–0.70)	0.78	0.51	0.30	0.89
Panel C: non-ML method					
Basic	0.63	0.62	0.64		0.94
Basic + Subjective	0.65	0.70	0.56		0.94
Basic + DDS	0.69	0.71	0.65		0.89
All	0.69	0.71	0.66		0.89

Note: Panel A shows results using the RF method, panel B shows results using the XGBoost method, and panel C shows results using the non-ML method. Columns 1–4 show statistics when the outcome variable is whether a household is calorie poor. Column 5 shows statistics when the outcome variable is AME adjusted daily per capita calorie intake. Total observation is 6,427, of which 70% are used to train the model and 30% are used as validation data. We use a ten-fold cross-validation technique in both the RF and XGBoost methods. Parentheses in column 1 present the 95% confidence interval. As the two groups are unbalanced for calorie poor indicator, we use SMOTE method to balance the minority class. The benchmark set includes education, age, and sex of household head; religion, maximum education in the household, household size, working-age males, and working-age females; number of household members who work in labor, service, non-farm self-employment, and farm self-employment; household, livestock, and agricultural assets; other land and homestead land values; whether the dwelling has bricked walls, floor, and roof; and electricity connection, and access to sealed latrine. The DDS indicators include WDDS, ODDS, HDDS, and FCS. Finally, the subjective set includes no food, no food at night, and no food at day and night. In the case of a binary dependent variable and non-ML method, we use 0.60 as the cut-off point for prediction. All results in the non-ML method are estimated based on 10,000 bootstrap replications.

6.2. Which individual characteristics best predict calorie intake?

Panel A of Table 3 lists the five most important variables in predicting calorie poor status of households as estimated using RF importance weights. When we use the benchmark set only, we find that household size, agriculture, and livestock assets, age of the household head, and the presence of male members are the most important variables. When we add the subjective set to the benchmark set, we find that none of the subjective indicators are among the five most important predictor variables. We then add the DDS set to the benchmark set and find that FCS and HDDS are two of the five most important predictor variables. Finally, we add both the subjective and the DDS sets to the benchmark set and find FCS and HDDS remain important variables in predicting calorie poor status.

In panel B of Table 3, we list the five most important variables for calorie intake as the target variable. We find that land, livestock, and household assets, education level, and the presence of male members are the most important predictors. Like the calorie-poor case, none of the subjective indicators are important in predicting calorie intake. However, when we add the DDS set to the benchmark set, we find that FCS and HDDS are two of the five most important variables along with household size, household assets, and livestock assets. When adding both the subjective and DDS sets to the benchmark set, we find that FCS remains as the only important predictor variable outside the benchmark set. The findings in this section clearly align with the results in the previous section. The subjective set is not important in predicting calorie indicators, while the DDS set plays an important role in predicting calorie indicators.

Table 3
Variable of importance in predicting calorie indicators.

	Benchmark (1)	Benchmark + Sub. (2)	Benchmark + DDS (3)	All (4)
Panel A: Calorie poor				
1st	HH asset	HH asset	FCS	FCS
2nd	Ag. asset	Ag. asset	Ag. asset	HH asset
3rd	Livestock asset	Livestock asset	Livestock asset	Ag. asset
4th	HH head age	HH head age	HH asset	Livestock asset
5th	Male member	Male member	HDDS	HDDS
Panel B: Calorie intake				
1st	Education	HH asset	FCS	FCS
2nd	Land asset	Male member	HH size	HH size
3rd	Male member	Education	HDDS	Male member
4th	HH asset	HH size	HH asset	Livestock asset
5th	Livestock asset	Livestock asset	Livestock asset	HH asset

Note: Variable of importance is estimated based on the mean decrease in accuracy of prediction. Panel A shows the results when the dependent variable is Calorie poor and panel B shows results when the dependent variable is Calorie intake. Total observation is 6,427, of which 70% are used to train the model and 30% are used as validation data. We use a ten-fold cross-validation technique in both the RF and XGBoost methods. As the two groups are unbalanced for calorie poor indicator, we use SMOTE method to balance the minority class. The benchmark set includes education, age, and sex of household head; religion, maximum education in the household, household size, working-age males, and working-age females; number of household members who work in labor, service, non-farm self-employment, and farm self-employment; household, livestock, and agricultural assets; other land and homestead land values; whether the dwelling has bricked walls, floor, and roof; and electricity connection, and access to sealed latrine. The DDS indicators include WDDS, ODDS, HDDS, and FCS. Finally, the subjective set includes no food, no food at night, and no food at day and night.

7. Machine Learning (ML) versus non-ML methods

In this section, we test whether and to what extent the ML method performs better in predicting calorie indicators relative to the non-ML method. Our chosen non-ML method combines traditional econometric estimation with out-of-sample prediction. We apply a regular logistic or OLS regression model, depending on the nature of the dependent variable, on a training sample (70% of total households) and estimate accuracy, sensitivity, and specificity statistics for the validation sample (remaining 30% of households). We bootstrap the entire process 10,000 times and estimate the average accuracy, sensitivity, and specificity of prediction when the target variable is calorie poor and NRMSE when the target variable is calorie intake. For the calorie poor, we use 0.60 as the cut-off point to classify households into calorie poor and calorie non-poor. Like the ML method, we start with the benchmark set and then include the subjective and the DDS sets sequentially.

Results are presented in panel C of Table 2. We find that the overall accuracy for the non-ML method ranges between 64 and 71%, which is similar to what we observe for the ML method. As with the ML method, we find that the inclusion of the DDS set improves prediction accuracy. In addition, sensitivity improves sharply with the inclusion of the subjective and DDS sets. Contrary to the ML method, we find that the specificity of prediction drops with the inclusion of the subjective and DDS sets. For calorie intake, we find that NRMSE falls from 0.94 to 0.89 with the inclusion of the DDS set, which is again like the ML method. Two important points arise from the ML versus non-ML methods comparison. First, the two methods perform quite similarly in predicting overall accuracy. Second, unlike the ML method case, the subjective set improves prediction like the DDS set.

We use the average proportion of the calorie-poor households in the sample as the cut-off point. Obviously, a smaller cut-off point will increase the sensitivity of prediction (in other words, reduce the probability of exclusion error), but it may reduce the specificity of prediction (increase the probability of inclusion error). In Table A2, we show that when we reduce the cut-off point to 0.50 or 0.40 relative to 0.60, the sensitivity of prediction increases at the cost of specificity. Therefore, the final cut-off point will depend on the coverage capacity of the respective program.

8. Community-level targeting

An important initial stage of any household-level targeting is an analogous targeting at the aggregate level (e.g., community or regional level). Because of budget constraints, government or non-government organizations need to identify most food insecure regions or communities first followed by identifying most food insecure households within each region or community. In this section, we predict community level (i.e., village) calorie indicators using predictor sets also measured at the community level. At the community level benchmark set, we include total households in the community, the proportion of households headed by a female, average education and asset levels, infrastructure, and access to credit and health facility related indicators. We estimate the DDS and the subjective sets at the community level by taking average values of household level observations. The target variables at the community level analysis are the proportion of calorie poor households and the average level of calorie intake in a community. Note that both the target variables in this section are continuous variables; therefore, we will use only NRMSE to estimate the marginal gain in the prediction of calorie indicators by different predictor sets. Like the household level analysis, we use RF, XGBoost, and non-ML methods for prediction. In our data, we have a total of 323 communities. We use 70% of the community sample as the training sample and the remaining 30% as the validation sample.

Panel A in Table 4 shows results where the target variable is the proportion of calorie poor households in a community. In the case of the RF method, we find that NRMSE is 0.96 when we use the benchmark set only, but NRMSE rises to 0.98 with the inclusion of the subjective and the DDS sets. In the case of the XGBoost method, we again find that the NRMSE is lowest when we use the benchmark set only and is highest when we add both the subjective and DDS sets to the benchmark set. However, we find that the NRMSE drops when we include the DDS set to the benchmark set for the non-ML method.

Panel B in Table 4 shows results where the target variable is the average level of calorie intake in a community. We find that NRMSE is 0.92 when combining the benchmark set with the RF method. The NRMSEs are 0.92, 0.90, and 0.91 when we include the subjective, DDS, and both sets together with the benchmark set, respectively. The XGBoost method, on the other hand, shows that NRMSE is lowest when we use only the benchmark set as predictors of calorie intake. The non-

Table 4
Prediction of calorie indicators at the community level.

	Benchmark (1)	Benchmark + Sub. (2)	Benchmark + DDS (3)	All (4)
Panel A: Calorie poor				
RF	0.96	0.96	0.97	0.98
XGBoost	0.94	0.95	0.95	0.98
Non-ML	0.99	0.98	0.95	0.95
Variable of importance				
1st	Ag. asset	Ag. asset	FCS	FCS
2nd	Land asset	Land asset	Ag. asset	Ag. asset
3rd	Female head (%)	HH asset	Land asset	Land asset
4th	HH asset	Contract farming	ODDS	Contract farming
5th	Brick floor (%)	Sleep Hungry	Female head (%)	ODDS
Panel B: Calorie intake				
RF	0.92	0.92	0.90	0.91
XGBoost	0.93	1.07	0.99	1.00
Non-ML	0.97	0.97	0.93	0.94
Variable of importance				
1st	Female head (%)	Female head (%)	FCS	FCS
2nd	Ag. asset	Land asset	Female head (%)	Female head (%)
3rd	Land asset	Ag. asset	Ag. asset	Land asset
4th	Contract farming	No food	Land asset	Ag. asset
5th	HH asset	HH asset	HDDS	No food

Note: In panel A and panel B, the dependent variables are the proportion of calorie poor households in a community and average calorie intake level, respectively. Total observation is 323, of which 70% are used to train the model and 30% are used as validation data. We use a ten-fold cross-validation technique in both the RF and XGBoost methods. For the non-ML method, results are estimated based on 10000 bootstrap replications. The benchmark set includes household proportions of non-Muslims, female heads, dwellings with bricked walls, floors, and roofs; electricity connection and access to sealed latrine; average education level; average household, livestock, agricultural assets; and other land, and homestead land values. At the community level, the benchmark set includes the number of households and road conditions in the community; number of private and government clinics, and pharmacies; number of informal lenders, NGOs, input dealers, and contract farming dealers. The DDS indicators include the average level of WDDS, ODDS, HDDS, and FCS. Finally, the subjective set includes the proportion of households with no food, no food at night, and no food at day and night.

ML method follows the RF method that NRMSE falls with the inclusion of the DDS set to the benchmark set.

From the list of important variables using the RF method, we find that household, agriculture, and livestock assets, the percentage of female-headed households, and household infrastructure are the most important variables. FCS and ODDS remain as two of the important variables when we include both the subjective and DDS sets to the benchmark set. We find a similar set of important variables when the target variable is the average level of calorie intake in a community. We find that FCS and no-food (one of the subjective indicators) remain as two of the important variables when we include both the subjective and DDS sets to the benchmark set.

In Figs. A3–A6, we show the distributions of original and predicted proportion of calorie poor households and the average level of calorie intake in a community by different predictor sets, starting from the benchmark set to the set that includes all predictors. In Figs. A7 And A8, we show the percentage of the incorrect predictions for the community level proportion of calorie poor households by the RF and XGBoost methods. From the community level analysis, we do not find a clear improvement in prediction accuracy with the inclusion of the subjective and DDS sets. Nevertheless, as in our household-level analysis, we find that DDS set still has some positive marginal gain over the benchmark set.

9. Conclusion

In this paper, we revisit the issue of measuring food security. Specifically, we test the performance of alternative predictor sets that are relatively inexpensive to collect as proxies for calorie-based indicators. We use a nationally representative dataset from Bangladesh and compare three alternative predictor sets. The benchmark set includes household level variables that are easy to collect and quantify. We then sequentially add a subjective set and a DDS set to the benchmark set and test the extent of improvement in the prediction of calorie indicators.

We find that overall prediction accuracy with respect to calorie-poor status rises from 63% to 69% when we add the subjective and DDS sets to the benchmark set. We also find that the DDS indicators are relatively more important in predicting calorie status of households as compared to the subjective set. Overall, our results indicate that the marginal gain from the addition of the DDS and subjective sets are small. We also show that with an adequate sample and correct-methodology, a non-ML method performs similar to the ML method in terms of overall accuracy but generates a large exclusion error. From the community level analysis, we show that the NRMSEs remain lowest when we include only the benchmark set.

One of the important findings in our study is that the benchmark set accounts for approximately 90% of the total prediction accuracy with respect to household calorie indicators. Therefore, an important question is why the benchmark set is so powerful and in what contexts we would expect the subjective and DDS sets to add predictive power? One possible explanation is that both the subjective and DDS sets are correlated with the predictors in the benchmark set, so that the marginal contribution of the subjective and DDS set are small. Some earlier studies show a significant correlation between DDS (or subjective) and household level characteristics. For instance, DDS is related to socioeconomic characteristics and asset ownership (Hatløy et al., 2000; Anzid et al., 2009; Rashid, Smith, and Rahman, 2011). Other studies find that CSI, a variant of the subjective food security indicator, is strongly correlated with household assets (Maxwell et al., 2008; Jones et al., 2013). In other words, the same assets and household characteristics that make it possible for a household to afford sufficient calories also make it more likely that a household will be able to purchase a diverse diet and avoid feelings of food insecurity. When viewed in this light, it is unsurprising expected that the marginal gain in predictive accuracy from using the subjective and DDS sets on the top of the benchmark set is small. However, the subjective set has an advantage of capturing psychological and seasonal dimensions of food security (Coates et al., 2007), while the DDS set has a strong relation to household nutrition status (Hatløy et al., 2000; Arimond and Ruel, 2004). Future work could assess the ability of the subjective and DDS sets to predict not just calorie-based indicators of food security, but micronutrient intake and detailed psychological measures of food security as well.

We find that overall prediction accuracy ranges between 60% and 70% for our selected ML and non-ML methods. The sensitivity of prediction (predicted calorie poor to true calorie poor households) ranges 70%–80% while the specificity (predicted calorie non-poor to true calorie non-poor households) of prediction ranges 50%–70% depending on the method used. Whether the level of accuracy obtained in our analysis is acceptable would depend on the study context. The additional gain in accuracy obtained by collecting detailed food consumption data have to be weighed against the associated time and monetary costs. Regardless, given that we find similar predictive performance for the DDS and subjective sets across various estimation methods, we believe our results can serve as a valuable guide in deciding what data to use in measuring food security.

We find that the ML and non-ML methods perform quite similarly in predicting overall accuracy, where our non-ML method consists of either OLS or logistic regression combined with out-of-sample prediction. At first glance, it may seem surprising that our non-ML method would fair comparably to our ML method. The success of the OLS and logistic regression as predictive tools depends on how closely the conditional mean (for a continuous outcome) or the odds ratio (for a binary classification problem) can be approximated by a linear function. Our predictor sets have a fairly small number of covariates from which to choose, and several of the predictors are binary, limiting the scope for model complexity and making it more likely that a simple linear model will yield an adequate approximation. With a larger pool of predictors, we might find that RF and XGBoost begin to outperform our non-ML methods. Which methods are likely to work best will depend on the pool of available predictor sets as well as the complexity of the functional forms linking predictors to outcomes.

Appendix A

A.1. Random forest (RF) method

We briefly explain the RF method here following Zhang and Ma (2012) closely. Let X be the predictor set and Y be the calorie indicator. RF uses a prediction function, $f(X)$, that minimizes the expected loss function, $E_{XY}(L(Y, f(X)))$, where a loss function (L) measures the closeness of $f(X)$ and Y. For a continuous target variable (calorie intake), a typical loss function (L) is a squared error loss, $L(Y, f(X)) = (Y - f(X))^2$. For a binary target variable (calorie poor), a common loss function (L) is as follows:

$$L(Y, f(X)) = I(Y \neq f(X)) = \begin{cases} 0 \text{ if } Y = f(X) \\ 1 \text{ otherwise} \end{cases}$$

The prediction function ($f(X)$) is determined by J number of trees, $h_1(x)$, ……,$h_j(x)$. More specifically, $f(X)$ is the average of all J trees for a

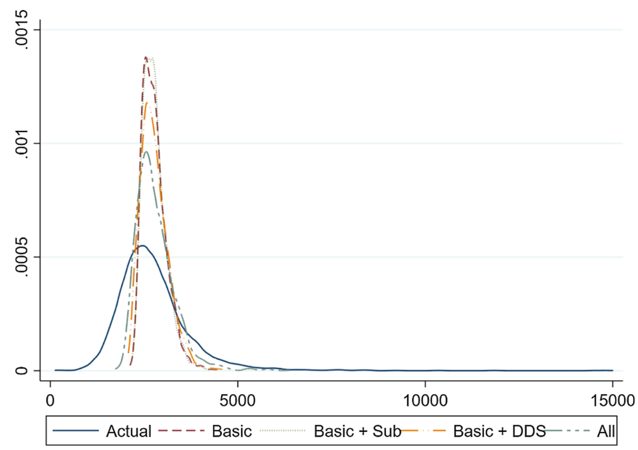

Fig. A1. Distribution of predicted calorie intake by alternative predictor sets (RF method). Note: Total observation is 6,427, of which 70% are used to train the model and 30% are used as validation data. We use a ten-fold cross-validation technique in estimating the RF method. As the two groups are unbalanced for calorie poor indicator, we use SMOTE method to balance the minority class. The benchmark set includes education, age, and sex of household head; religion, maximum education in the household, household size, working-age males, and working-age females; number of household members who work in labor, service, non-farm self-employment, and farm self-employment; household, livestock, and agricultural assets; other land and homestead land values; whether the dwelling has bricked walls, floor, and roof; and electricity connection, and access to sealed latrine. The DDS indicators include WDDS, ODDS, HDDS, and FCS. Finally, the subjective set includes no food, no food at night, and no food at day and night.

Table A1
Energy requirement and AME by age and sex.

	Male		Female	
Age	Energy (kcal)	AME	Energy (kcal)	AME
0	635.1	0.22	582.1	0.20
1	917.5	0.32	835.9	0.29
2	1117.2	0.39	1031.4	0.36
3	1231.5	0.43	1150.5	0.40
4	1340.7	0.47	1260.8	0.44
5	1447.0	0.51	1352.6	0.47
6	1568.3	0.55	1452.5	0.51
7	1694.8	0.59	1562.7	0.55
8	1825.6	0.64	1688.7	0.59
9	1965.3	0.69	1826.0	0.64
10	2097.7	0.73	1942.9	0.68
11	2226.6	0.78	2044.4	0.72
12	2380.3	0.83	2142.6	0.75
13	2555.9	0.90	2223.0	0.78
14	2723.5	0.95	2280.9	0.80
15	2862.3	1.00	2318.4	0.81
16	2967.4	1.04	2341.2	0.82
17	3042.0	1.07	2355.2	0.83
18	3095.6	1.08	2364.8	0.83
19–29	2906.6	1.02	2239.6	0.78
30–59	2854.1	1.00	2298.8	0.81
60+	2352.0	0.82	2039.0	0.71

Note: Source: Waid et al. (2017). AME is shown for a moderate activity level.

continuous target variable (i.e., calorie intake) and most frequently predicted category for a binary target variable (i.e., calorie poor).

A typical RF-based prediction method randomly splits data into training and validation subsets at the beginning and performs the following steps:

1. Draw a bootstrapped sample of size N with replacement from the training data and fit a tree using the binary recursive partitioning as follows:
   (a) Select k predictors out of m available predictors and find the best binary split among all potential splits, that is, the split that decreases the within-node residual sum of squares most often or using Gini criteria.
   (b) Split the node into descendants as above until a terminal node is defined and build a tree.
2. Repeat steps a and b “n” number of times to generate “n” number of trees.
3. Make predictions as $\hat{f}(X) = \arg\max \sum_{j=1}^{J} I\ (Y = h_j(x))$ for tree-based RF and $\hat{f}(X) = 1/j\ \sum h_j(x)$, j = 1. ..j for regression-based RF.

A.2. Variables of importance

RF uses some holdout a sample from the bootstrapped sample in Step 1, called “out-of-bag data”. RF can also list variables in terms of their importance in prediction using two steps as follows:

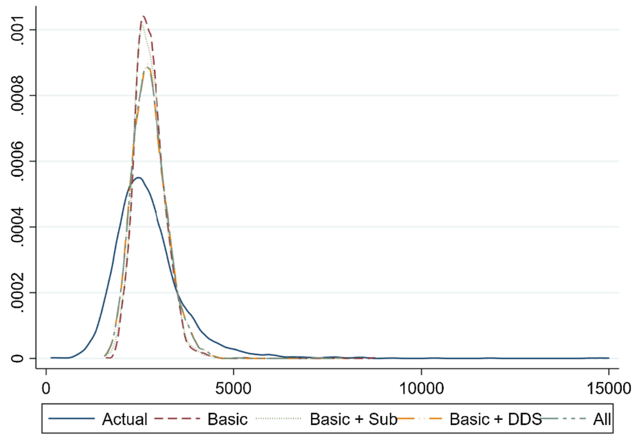

Fig. A2. Distribution of predicted calorie intake by alternative predictor sets (XGBoost method). Note: Total observation is 6,427, of which 70% are used to train the model and 30% are used as validation data. We use a ten-fold cross-validation technique in estimating the XGBoost method. As the two groups are unbalanced for calorie poor indicator, we use SMOTE method to balance the minority class. The benchmark set includes education, age, and sex of household head; religion, maximum education in the household, household size, working-age males, and working-age females; number of household members who work in labor, service, non-farm self-employment, and farm self-employment; household, livestock, and agricultural assets; other land and homestead land values; whether the dwelling has bricked walls, floor, and roof; and electricity connection, and access to sealed latrine. The DDS indicators include WDDS, ODDS, HDDS, and FCS. Finally, the subjective set includes no food, no food at night, and no food at day and night.

Table A2
Prediction of calorie indicators using the non-ML method.

	Accuracy (1)	Sensitivity (2)	Specificity (3)
Cut-off = 0.40			
Basic	0.66	0.73	0.53
Basic + Subjective	0.66	0.97	0.10
Basic + DDS	0.69	0.95	0.24
All	0.69	0.95	0.24
Cut-off = 0.50			
Basic	0.64	0.68	0.58
Basic + Subjective	0.68	0.89	0.29
Basic + DDS	0.71	0.87	0.44
All	0.71	0.87	0.44
Cut-off = 0.60			
Basic	0.63	0.62	0.64
Basic + Subjective	0.65	0.70	0.56
Basic + DDS	0.69	0.71	0.65
All	0.69	0.71	0.66

Note: Total observation is 6,427, of which 70% are used as training data and 30% are used as validation data. All statistics are estimated based on 1000 bootstrap replications. The benchmark set includes education, age, and sex of household head; religion, maximum education in the household, household size, working-age males, and working-age females; number of household members who work in labor, service, non-farm self-employment, and farm self-employment; household, livestock, and agricultural assets; other land and homestead land values; whether the dwelling has bricked walls, floor, and roof; and electricity connection, and access to sealed latrine. The DDS indicators include WDDS, ODDS, HDDS, and FCS. Finally, the subjective set includes no food, no food at night, and no food at day and night.

1. For each tree in step b above, the out-of-bag observations are passed down the tree and predicted values are stored.
2. Values of the predictor k are randomly permuted and predicted values are stored again at the new value holding all other variables fixed. The difference of MSE is computed for each observation and averaged over all observations to determine the importance of the variable k.

A variant of the mean decrease in accuracy is the mean decrease in GINI, where the RF method uses a variable in each time to split a node and the Gini coefficient is calculated and compared to the GINI of the original node.

A.3. Extreme gradient boosting (XGBoost) method

The XGBoost method also follows similar stages like the RF method. However, XGBoost does additional error correction in subsequent trees based on the prediction errors in previous trees. The additional steps are as follows:

1. Fit decision tree on the training sample
2. Estimate error in prediction, $e_1 = y_i - \hat{y}_1$
3. Model e_1 as target variable with the same predictors and predicted error, $\hat{e}_1$

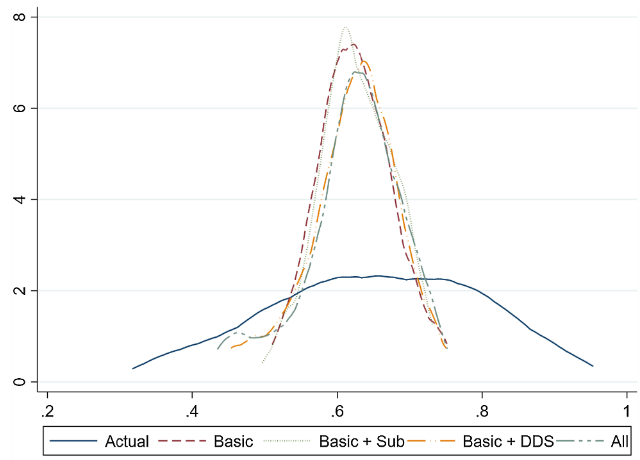

Fig. A3. Distribution of predicted proportion of calorie poor households at the community by alternative predictor sets (RF method). Note: Total observation is 323, of which 70% are used to train the model and 30% are used as validation data. We use a ten-fold cross-validation technique in estimation. The benchmark set includes household proportions of non-Muslims, female heads, dwellings with bricked walls, floors, and roofs; electricity connection and access to sealed latrine; average education level; average household, livestock, and agricultural assets; and other land and homestead land values. At the community level, the benchmark set includes the number of households and road conditions in the community; number of private and government clinics, and pharmacies; number of informal lenders, NGOs, input dealers, and contract farming dealers. The DDS indicators include the average level of WDDS, ODDS, HDDS, and FCS. Finally, the subjective set includes the proportion of households with no food, no food at night, and no food at day and night.

4. Add e_1 to $\hat{y}_1$, $\hat{y}_2 = \hat{y}_1 + \hat{e}_1$
5. Estimate residual error in prediction $e_2 = y_i - \hat{y}_2$ and repeat steps 2 to until it starts overfitting, or the sum of residuals become constant.

A.4. Non-ML method

1. Randomly split data into training (70%) and validation (30%) samples
2. Run a logistic or OLS regression depending on the nature of the target variable using the training data only.
3. Predict calorie poor or calorie intake for the validation sample using the estimated coefficients for the training sample
4. Estimate the accuracy, sensitivity, and specificity statistics when the target variable is calorie poor and NRMSE when the target variable is calorie intake
5. Bootstrap the entire process (step 1–4) N number of times and estimate an average of all statistics.

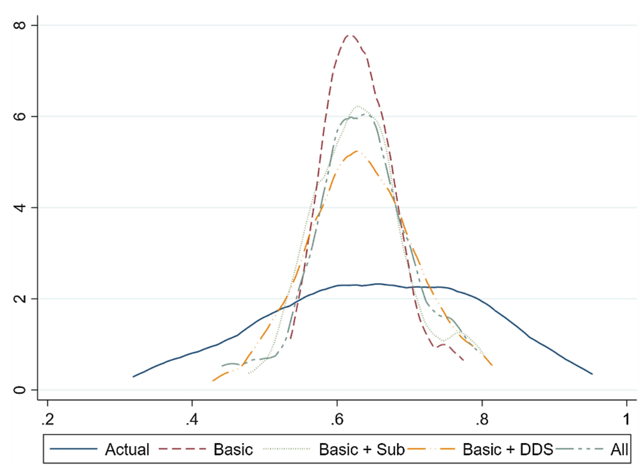

Fig. A4. Distribution of predicted proportion of calorie poor households at the community by predictor sets (XGBoost method). Note: Total observation is 323, of which 70% are used to train the model and 30% are used as validation data. We use a ten-fold cross-validation technique in estimation. The benchmark set includes household proportions of non-Muslims, female heads, dwellings with bricked walls, floors, and roofs; electricity connection and access to sealed latrine; average education level; average household, livestock, and agricultural assets; and other land and homestead land values. At the community level, the benchmark set includes the number of households and road conditions in the community; number of private and government clinics, and pharmacies; number of informal lenders, NGOs, input dealers, and contract farming dealers. The DDS indicators include the average level of WDDS, ODDS, HDDS, and FCS. Finally, the subjective set includes the proportion of households with no food, no food at night, and no food at day and night.

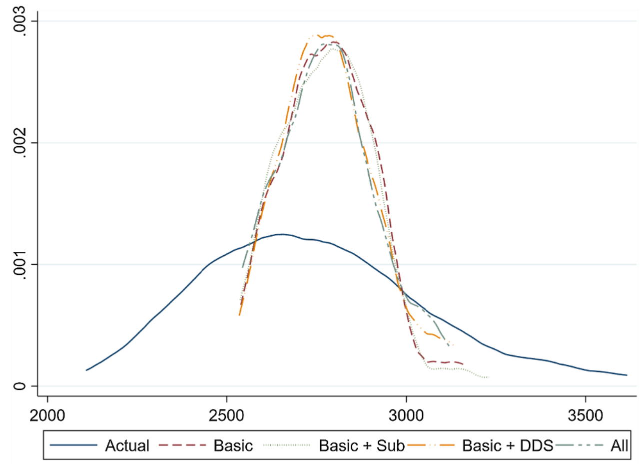

Fig. A5. Distribution of predicted average calorie intake at the community level by alternative predictor sets (RF method). Note: Total observation is 323, of which 70% are used to train the model and 30% are used as validation data. We use a ten-fold cross-validation technique in estimation. The benchmark set includes household proportions of non-Muslims, female heads, dwellings with bricked walls, floors, and roofs; electricity connection and access to sealed latrine; average education level; average household, livestock, and agricultural assets; and other land and homestead land values. At the community level, the benchmark set includes the number of households and road conditions in the community; number of private and government clinics, and pharmacies; number of informal lenders, NGOs, input dealers, and contract farming dealers. The DDS indicators include the average level of WDDS, ODDS, HDDS, and FCS. Finally, the subjective set includes the proportion of households with no food, no food at night, and no food at day and night.

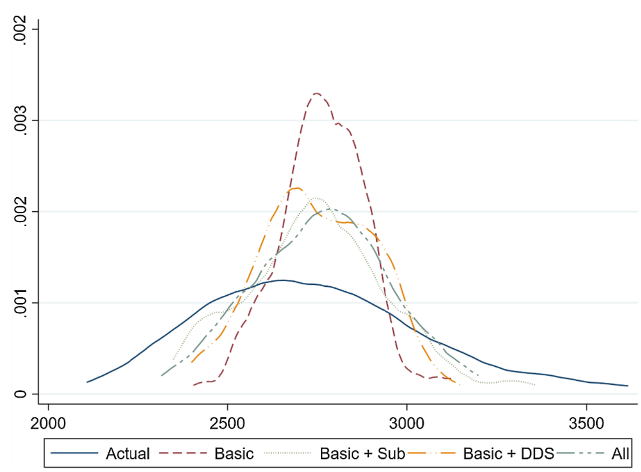

Fig. A6. Distribution of predicted average calorie intake at the community level by predictor sets (XGBoost method). Note: Total observation is 323, of which 70% are used to train the model and 30% are used as validation data. We use a ten-fold cross-validation technique in estimation. The benchmark set includes household proportions of non-Muslims, female heads, dwellings with bricked walls, floors, and roofs; electricity connection and access to sealed latrine; average education level; average household, livestock, and agricultural assets; and other land and homestead land values. At the community level, the benchmark set includes the number of households and road conditions in the community; number of private and government clinics, and pharmacies; number of informal lenders, NGOs, input dealers, and contract farming dealers. The DDS indicators include the average level of WDDS, ODDS, HDDS, and FCS. Finally, the subjective set includes the proportion of households with no food, no food at night, and no food at day and night.

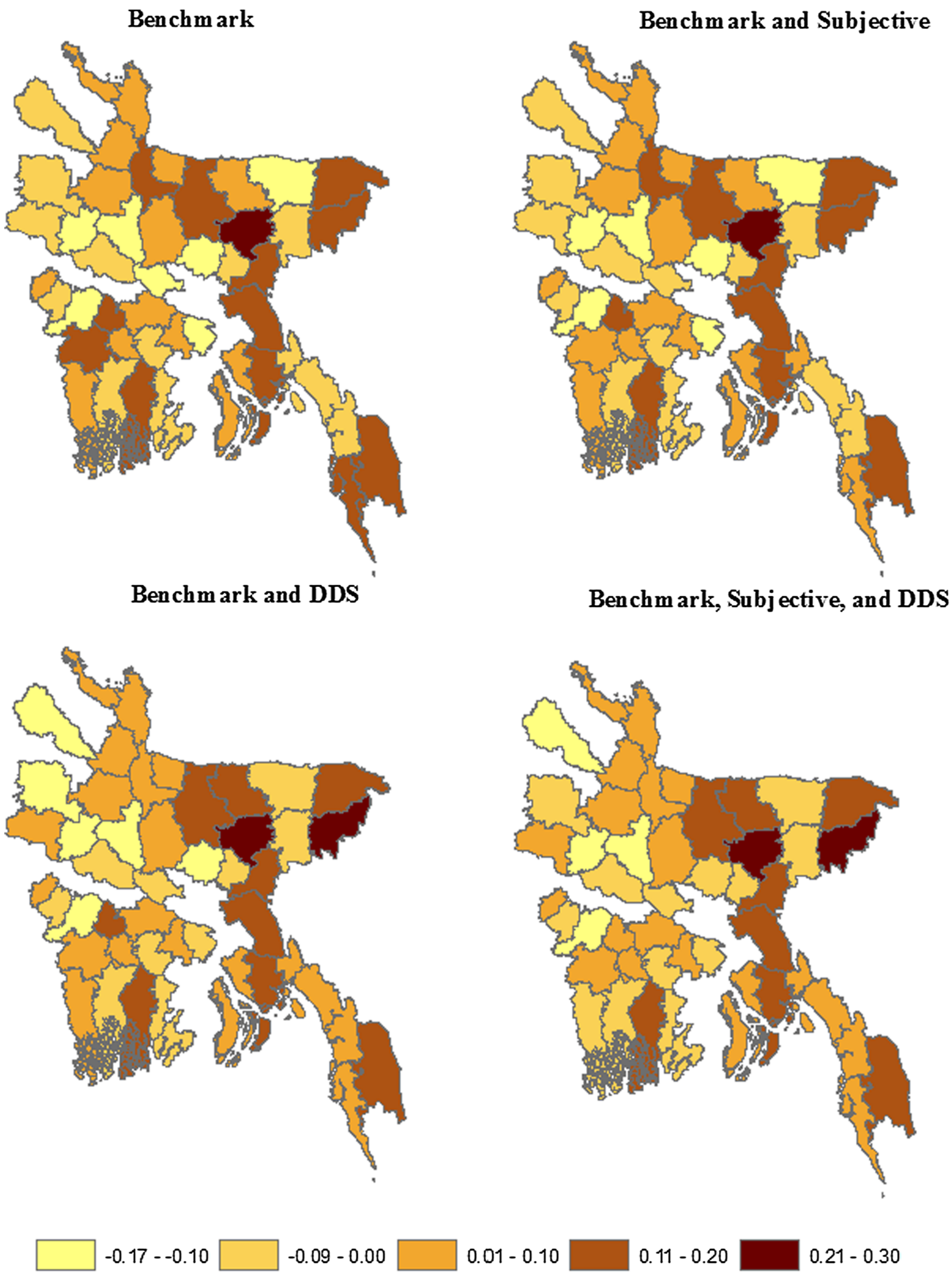

Fig. A7. Proportion of incorrectly targeted calorie poor households by District (RF method). Note: Each cell shows a district in Bangladesh. The average incorrectly targeted proportion of households is calculated from the community level analysis. Total number of districts is 64 of which information for 47 districts is shown in the graph. Online supplementary file contains an unedited district level map of Bangladesh. See note in Fig. A6 for details about the community level analysis.

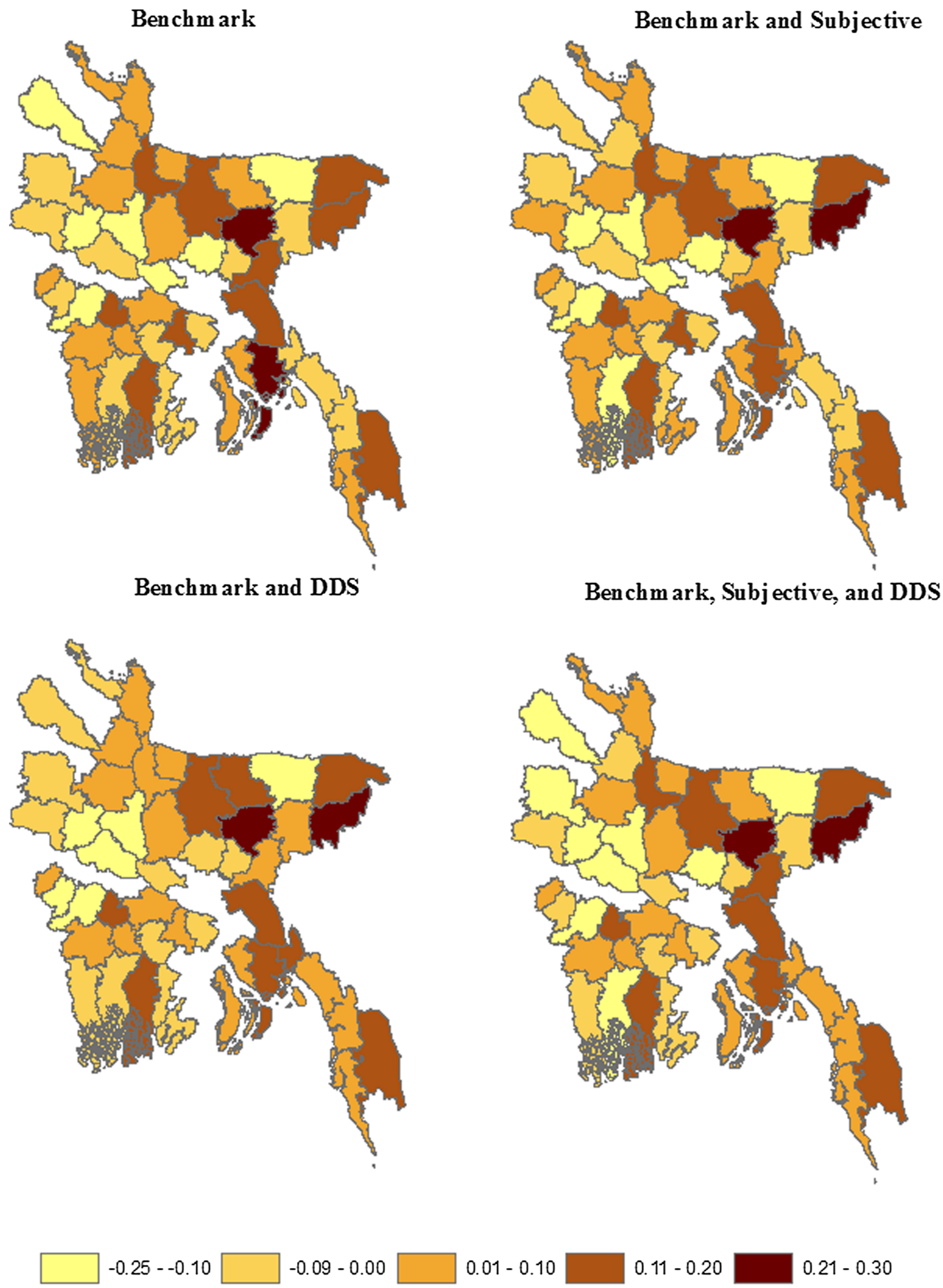

Fig. A8. Proportion of incorrectly targeted calorie poor households by District (XGBoost method). Note: Each cell shows a district in Bangladesh. The average incorrectly targeted proportion of households is calculated from the community level analysis. Total number of districts is 64 of which information for 47 districts is shown in the graph. Online supplementary file contains an unedited district level map of Bangladesh. See note in Fig. A6 for details about the community level analysis.

Appendix B. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodpol.2019.03.001.

References

Ahmed, A.U., Ahmad, K., Chou, V., Hernandez, R., Menon, P., Naeem, F., Hassan, Z., 2013. The Status of Food Security in the Feed the Future Zone and Other Regions of Bangladesh: Results from the 2011–2012 Bangladesh Integrated Household Survey. International Food Policy Research Institute, Washington, DC.

Anzid, K., Elhamdani, F.Z., Baali, A., Boëtsch, G., Levy-Desroches, S., Lôpez, P.M., Cherkaoui, M., 2009. The effect of socio-economic status and area of residence on household food variety in Morocco. Ann. Hum. Biol. 36 (6), 727–749.

Arimond, M., Ruel, M.T., 2004. Dietary diversity is associated with child nutritional status: evidence from 11 demographic and health surveys. J. Nutr. 134 (10), 2579–2585.

Ballard, T., Coates, J., Swindale, A., Deitchler, M., 2011. Household hunger scale: indicator definition and measurement guide. Food and Nutrition Technical Assistance (FANTA) 2.

Barrett, C.B., 2010. Measuring food insecurity. Science 327 (5967), 825–828.

Blumenstock, J., Cadamuro, G., On, R., 2015. Predicting poverty and wealth from mobile

phone metadata. Science 350 (6264), 1073–1076.
Carletto, C., Zezza, A., Banerjee, R., 2013. Towards better measurement of household food security: harmonizing indicators and the role of household surveys. Global Food Security 2 (1), 30–40.
Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority over-sampling technique. J. Artificial Intelligence Res. 16, 321–357.
Chung, K. (Ed.), 1997. Identifying the Food Insecure: The Application of Mixed-Method Approaches in India. International Food Policy Research Institute, Washington, DC.
Coates, J., Rogers, B.L., Webb, P., Maxwell, D., Houser, R., McDonald, C., 2007. Diet Diversity Study (No. 40). Friedman School of Nutrition Science and Policy.
Dahikar, S.S., Rode, S.V., 2014. Agricultural crop yield prediction using artificial neural network approach. Int. J. Innovative Res. Electr., Electron., Instrum. Control Eng. 2 (1), 683–686.
Deitchler, M., Ballard, T., Swindale, A., Coates, J., 2010. Validation of a Measure of Household Hunger for Cross-cultural use. USAID, Washington, DC.
Del Ninno, C., Dorosh, P.A., Smith, L.C., 2003. Public policy, markets and household coping strategies in Bangladesh: Avoiding a food security crisis following the 1998 floods. World Dev. 31 (7), 1221–1238.
Faisal, I.M., Parveen, S., 2004. Food security in the face of climate change, population growth, and resource constraints: implications for Bangladesh. Environ. Manage. 34 (4), 487–498.
Food and Agriculture Organization [FAO], 1996. Rome Declaration on World Food Security and World Food Summit Plan of Action. FAO, Rome.
Food and Agriculture Organization [FAO], 2014. The State of Food Insecurity in the World, 2013: The Multiple Dimensions of Food Security. FAO, Rome.
Glaeser, E.L., Kominers, S.D., Luca, M., Naik, N., 2018. Big data and big cities: the promises and limitations of improved measures of urban life. Econ. Inq. 56 (1), 114–137.
Hatløy, A., Hallund, J., Diarra, M.M., Oshaug, A., 2000. Food variety, socioeconomic status and nutritional status in urban and rural areas in Koutiala (Mali). Public Health Nutr. 3 (01), 57–65.
Headey, D.D., 2013. The impact of the global food crisis on self-assessed food security. The World Bank Econ. Rev. 27 (1), 1–27.
Headey, D., Ecker, O., 2013. Rethinking the measurement of food security: from first principles to best practice. Food Security 5 (3), 327–343.
Hoddinott, J., 1999. Choosing Outcome Indicators of Household Food Security. International Food Policy Research Institute, Washington, DC.
Hoddinott, J., Yohannes, Y., 2002. Dietary Diversity as a Food Security Indicator. International Food Policy Research Institute, Washington, DC.
Hossain, M., 2013. Food Security in Bangladesh: Achievement and challenges. The Daily Star. Retrieved from http://www.thedailystar.net/news/achievement-and-challenges.
Hyman, G., Larrea, C., Farrow, A., 2005. Methods, results and policy implications of poverty and food security mapping assessments. Food Policy 30 (5), 453–460.
Jamora, N., von Cramon-Taubadel, S., 2016. What world price? Appl. Econ. Perspect. Policy 39 (3), 479–498.
Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Science 353 (6301), 790–794.
Jones, A.D., Ngure, F.M., Pelto, G., Young, S.L., 2013. What are we assessing when we measure food security? A compendium and review of current metrics. Adv. Nutr. 4 (5), 481–505.
Knippenberg, E., Jensen, N., Constas, M., 2018. Resilience, shocks, and the dynamics of food insecurity: evidence from Malawi. Working paper, Cornell University, Ithaca, NY.
Kuhn, M., 2008. Building predictive models in R using the caret package. J. Stat. Softw. 28 (5), 1–26.
Maxwell, D., Ahiadeke, C., Levin, C., Armar-Klemesu, M., Zakariah, S., Lamptey, G.M., 1999. Alternative food-security indicators: revisiting the frequency and severity of coping strategies'. Food Policy 24 (4), 411–429.
Maxwell, D., Caldwell, R., Langworthy, M., 2008. Measuring food insecurity: Can an indicator based on localized coping behaviors be used to compare across contexts? Food Policy 33 (6), 533–540.
Maxwell, D., Vaitla, B., Coates, J., 2014. How do indicators of household food insecurity measure up? An empirical comparison from Ethiopia. Food Policy 47, 107–116.
Rashid, D.A., Smith, L.C., Rahman, T., 2011. Determinants of dietary quality: evidence from Bangladesh. World Dev. 39 (12), 2221–2231.
Reardon, T., Tschirley, D., Dolislager, M., Snyder, J., Hu, C., White, S., 2014. Urbanization, diet change, and transformation of food supply chains in Asia. Working paper, Michigan State University, East Lansing, MI.
Shaheen, N., Rahim, A.T.M., Mohiduzzaman, M., …, Stadlmayr, D., 2013. Food composition table for Bangladesh. Working paper, University of Dhaka, Dhaka, Bangladesh.
Steele, J.E., Sundsøy, P.R., Pezzulo, C., Alegana, V.A., Bird, T.J., Blumenstock, J., Hadiuzzaman, K.N., 2017. Mapping poverty using mobile phone and satellite data. J. R. Soc. Interface 14 (127), 20160690.
Suryanarayana, M.H., Silva, D., 2007. Is targeting the poor a penalty on the food insecure? Poverty and food insecurity in India. J. Hum. Dev. 8 (1), 89–107.
Swindale, A., Bilinsky, P., 2006. Household Dietary Diversity Score (HDDS) for Measurement of Household Food Access: Indicator Guide. USAID, Washington, DC.
Szabo, S., Hossain, M.S., Adger, W.N., Matthews, Z., Ahmed, S., Lázár, A.N., Ahmad, S., 2016. Soil salinity, household wealth and food insecurity in tropical deltas: evidence from south-west coast of Bangladesh. Sustain. Sci. 11 (3), 411–421.
Vaitla, B., Coates, J., Glaeser, L., Hillbruner, C., Biswal, P., Maxwell, D., 2017. 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.
Waid, J.L., Bogard, J.R., Thilsted, S.H., Gabrysch, S., 2017. Estimates of average energy requirements in Bangladesh: Adult Male Equivalent values for use in analyzing household consumption and expenditure surveys. Data Brief 14, 101–106.
Webb, P., Coates, J., Frongillo, E.A., Rogers, B.L., Swindale, A., Bilinsky, P., 2006. Measuring household food insecurity: why it's so important and yet so difficult to do. J. Nutr. 136 (5), 1404S–1408S.
Wiesmann, D., Bassett, L., Benson, L., Hoddinott, J., 2009. Validation of the World Food Programme's Food Consumption Score and Alternative Indicators of Household Food Security. International Food Research Policy Institute, Washington, DC.
World Food Program [WFP], 2008. Food Consumption Analysis: Calculation and Use of the Food Consumption Score in Food Security Analysis. World Food Programme, Rome.
Zhang, C., Ma, Y., 2012. Ensemble Machine Learning: Methods and Applications. Springer, New York.