Food security prediction from heterogeneous data combining machine and deep learning methods

Hugo Deléglise, Roberto Interdonato, Agnès Bégué, Elodie Maître d'Hôtel, Maguelonne Teisseire, Maguelonne Teisseire, Mathieu Roche

Food security prediction from heterogeneous data combining machine and deep learning methods

Hugo Deléglise[a,c,*], Roberto Interdonato[a,c], Agnès Bégué[a,c], Elodie Maître d'Hôtel[b,d], Maguelonne Teisseire[a,e], Mathieu Roche[a,c]

[a] *TETIS, Univ Montpellier, AgroParisTech, CIRAD, CNRS, INRAE, Montpellier, France.*
[b] *MOISA, Univ Montpellier, CIHEAM-IAMM, CIRAD, INRAE, Institut Agro, Montpellier, France.*
[c] *CIRAD, UMR TETIS, F-34398 Montpellier, France.*
[d] *CIRAD, UMR MOISA, F-34398 Montpellier, France.*
[e] *INRAE, Montpellier, France.*

*Corresponding author

Email addresses: hugo.deleglise@cirad.fr (Hugo Deléglise), roberto.interdonato@cirad.fr (Roberto Interdonato), agnes.begue@cirad.fr (Agnès Bégué), elodie.maitredhotel@cirad.fr (Elodie Maître d'Hôtel), maguelonne.teisseire@inrae.fr (Maguelonne Teisseire), mathieu.roche@cirad.fr (Mathieu Roche)

Preprint submitted to Elsevier *September 20, 2021*

Abstract

After many years of decline, hunger in Africa is growing again. This represents a global societal issue that all disciplines concerned with data analysis are facing. The rapid and accurate identification of food insecurity situations is a complex challenge. Although a number of food security alert and monitoring systems exist in food insecure countries, the data and methodologies they are based on do not allow for comprehending food security in all its complexity. In this study, we focus on two key food security indicators: the food consumption score (FCS) and the household dietary diversity score ($HDDS$). Based on the observation that producing such indicators is expensive in terms of time and resources, we propose the $FSPHD$ (Food Security Prediction based on Heterogeneous Data) framework, based on state-of-the-art machine and deep learning models, to enable the estimation of FCS and $HDDS$ starting from publicly available heterogeneous data. We take into account the indicators estimated using data from the Permanent Agricultural Survey conducted by the Burkina Faso government from 2009 to 2018 as reference data. We produce our estimations starting from heterogeneous data that include rasters (e.g., population density, land use, soil quality), GPS points (hospitals, schools, violent events), line vectors (waterways), quantitative variables (maize prices, World Bank variables, meteorological data) and time series (Smoothed Brightness Temperature - SMT, rainfall estimates, maize prices). The experimental results show a promising performance of our framework, which outperforms competing methods, thus paving the way for the development of advanced food security prediction systems based on state-of-the-art data science technologies.

Key words: Food security, Machine learning, Deep learning, Heterogeneous data

1. Introduction

Hunger is well known to be one of the major problems in many African countries. While a generalized and solid solution to this situation is far from being reached, constant progress was made during the first fifteen years of the current century. For instance, taking into account the area of West Africa, the years after 2000 have seen a decrease in the prevalence of undernourishment of nearly 2% points, reaching a relatively low value of 10.4% in 2012, which remained nearly constant until 2014 (10.7%). Nevertheless, an inverted trend has been observed in recent years, with the same value reaching a peak of 15.2% in 2019 and an alarming projection of 23% for 2030 (FAO et al., 2020). The same troubling trend can be observed for similar indicators, e.g., the prevalence of severe food insecurity situations in West Africa has seen an increase of nearly 10% points from 2014 (20.7%) to 2017 (29.5%) (FAO and ECA, 2018).

Among West African countries, Burkina Faso shows one of the most critical situations, with a prevalence of undernourishment of 21.3% in the 2015–2017 period (FAO and ECA, 2018). Burkina Faso is also one of the most affected by what is generally called the "triple burden of malnutrition", i.e., the coexistence of overnutrition, undernutrition and micronutrient deficiencies (FAO et al., 2018). The reasons behind such situations are complex, multifactorial, and interdependent. Climate change has certainly had a major impact on food production and availability by causing an increase in the number of extreme weather events (Tapsoba et al., 2019). Another key factor is the increase in population movements, which, in turn, is caused by the increasing number of conflicts in the Sahel region. Such movements have a major impact on food production and distribution channels (Lacher, 2012).

As a consequence of several food crises that occurred in the 1970s and 1980s in different areas of the world, a certain number of food security warning and monitoring systems were established by governmental and nongovernmental organizations. The aim of these systems, which are still active today, is to prevent food crises and help countries plan food aid programs to optimize the production and distribution of food. Some examples are GIEWS (Global Information and Early Warning System), created by the Food & Agriculture Organisation (FAO), and FEWSNET (Famine Early Warning Systems Network), founded by the United States Agency for International Development. These systems also publicly provide periodical bulletins reporting on the food security (FS) situation at regional and national scales.

However, such systems are mainly based on paradigms that call for the manual combination and summarization of all the sources of information that are taken into account during the process according to a series of predefined rules. While the need to mainly rely on human intervention can be justified by the complexity of the task at hand, such an effort represents an obstacle to accurately predicting food crises in time. More specifically, the current processes behind these monitoring systems are extremely time-consuming and allow for limited complexity in the number and heterogeneity of information sources that can be taken into account.

Moreover, these systems take into account certain key FS indicators, such as the food consumption score (FCS) or the household dietary diversity score ($HDDS$), which require household surveys that are expensive in terms of time and resources (cf. Section 3.1 for a detailed description of these metrics). In this context, machine learning methods can be exploited to provide timely estimates of these indicators by using easily accessed publicly available data. While these early warning systems mainly integrate meteorological and remote sensing data, we believe that the integration of information from other fields is related to FS (e.g., commodity prices, violent events, population dynamics) and that other data types (e.g., time series, high-resolution images) will make it possible to describe the phenomenon more completely.

The main objective of this work is to use original and effective machine- and deep learning-based models able to estimate FCS and $HDDS$, starting from publicly available heterogeneous data. More specifically, we aim to answer the following research questions:

- **RQ1**: What types of publicly available data can be targeted to predict FS scores?
- **RQ2**: How can data that are heterogeneous in terms of thematic, structure, and spatiotemporal resolution be preprocessed to obtain consistent predictions of the FS for a given study site?
- **RQ3**: How can state-of-the-art machine and deep learning approaches be exploited and combined to treat such heterogeneous data?

To answer these questions, we propose a machine learning framework, namely, $FSPHD$ (Food Security Prediction based on Heterogeneous Data), able to exploit heterogeneous explanatory data to estimate two target FS indicators, i.e., FCS and $HDDS$. To this end, we take into account multisource heterogeneous data such as rasters (e.g., population densities, land use, soil quality), GPS points (hospitals, schools, violent events), line vectors (waterways), quantitative variables (maize prices, World Bank variables, meteorological data) and time series (Smoothed Brightness Temperature (SMT), rainfall estimates, maize prices). The proposed framework is based on an ensemble of state-of-the-art data science techniques such as random forest (RF) (Qi, 2012), convolutional neural networks (CNNs) (Huang et al., 2018) and long short-term memory (LSTM) (Song et al., 2020). We test different variants of the framework that differ in the number and type of input data and in how the ensemble methods are processed to obtain the final result. We carry out an extensive experimental evaluation centred on the study area of Burkina Faso, leveraging FCS and $HDDS$ indicators (calculated using data from the permanent agricultural survey conducted by the Burkinabe government) as ground truth. Our work continues and develops a preliminary study that we published in the Data Integration and Applications Workshop (DINA 2020) (Deléglise et al., 2020).

The rest of the paper is structured as follows: in Section 2, we discuss related work, in Section 3, we describe the main FS indicators and the data used in the study, in Section 4, we introduce the proposed framework, in Section 5, we discuss our experimental evaluation, and Section 6 concludes the study.

2. Related work

2.1. Machine learning for food security and related problems

Machine learning methods are increasingly used to extract relevant information in the context of FS-related problems. Several studies exploit classic machine learning techniques (e.g., support vector machines, K-nearest neighbour, decision trees, and naive Bayes) for the prediction of FS indicators (e.g., Household Food Insecurity Access Scale and dietary energy intake) (Okori and Obua, 2011; Barbosa and Nelson, 2016; Lukyamuzi et al., 2018). In recent years, deep learning techniques, which have proven to be particularly effective in analysing complex heterogeneous data (Valdés, 2018), have also been used for the analysis of several FS-related topics such as poverty (Shailesh et al., 2018), drought (Mumtaz et al., 2018), market prices (Min et al., 2019) and asset wealth (Yeh et al., 2020). However, the full potential of such techniques has not yet been expressed in the prediction of FS indicators, and very few studies have focused on this issue. Heisenberg et al. (2020) use a classical neural network method called the multilayer perceptron to predict an FS indicator called the integrated phase classifier obtained from a household survey conducted by the United States Agency for International Development (USAID) in the Horn of Africa between 2009 and 2017. They use explanatory data from various domains (e.g., NDVI, food prices, conflicts, and soil moisture). A small number of studies focus on variables directly related to the quantity and quality of the food consumed, such as FCS or $HDDS$ (cf. Section 3.1), that are crucial for the understanding of FS. The main example in this context is the framework developed by the VAM (vulnerability analysis and monitoring) team of the World Food Programme (WFP-VAM, 2019). Such a framework integrates machine and deep learning techniques on heterogeneous data (i.e., satellite images at different spatial resolutions and GPS points) to predict FCS and $HDDS$ in several countries. Nevertheless, the results strongly varied from country to country, and the predictions were generally not accurate enough to be used in operational cases. Lentz et al. (2019) also predict FCS and $HDDS$ with linear regressions using mainly open and free data as predictors. The response variables are from the Living Standards Measurement Surveys (LSMS) conducted in Malawi in 2010 and 2013. The data used are from diverse sources: meteorology, precipitation, market prices and soil quality. However, the quality of their resulting predictions is relatively low, confirming that the prediction of these FS indicators is a complex issue, mainly because of their multifactorial nature. In this article, we want to contribute to the understanding of this rarely studied problem. In addition, one of the objectives of the two studies mentioned above (i.e., (Lentz et al., 2019) and (WFP-VAM, 2019)) is to be able to use free and easily accessible data to obtain estimates of indicators that are not commonly available because their computation requires lengthy and expensive household surveys. This topical issue has been considered by other studies that used open data as explanatory variables to predict FS-related indicators (e.g., assets, nightlights and integrated phase classifiers) using machine learning techniques (Yeh et al., 2020; Jean et al., 2016; van der Heijden et al., 2018). The use of open data is also one of the objectives of this study.

2.2. Machine learning on heterogeneous data

Due to the inherent complexity of the phenomena of food insecurity, the prediction of FS indicators requires the use of explanatory variables of heterogeneous thematics, structures and scales. For this reason, to address this problem, there is a need for methodologies that can combine these explanatory variables such that each source of information contributes to the prediction of FS indicators. For these reasons, another body of work strictly related to our research questions involves the use of machine learning with heterogeneous data sources. Such approaches have been applied in several application domains. In biochemistry, P. Lewis et al. (2006) make inferences from large, heterogeneous sets of protein sequences and structures using support vector machine techniques. In medicine, Miotto et al. (2017) review recent literature on the application of deep learning technologies to acquire knowledge and practical insights from complex, large and heterogeneous biomedical data. In remote sensing, Benedetti et al. (2018) propose an end-to-end deep learning framework, named M3Fusion, able to simultaneously leverage the temporal knowledge contained in time series data and the fine spatial information available in very high spatial resolution data using recurrent and convolutional neural networks. In the road sector, Yuan et al. (2018) perform a comprehensive study on the traffic accident prediction problem using a convolutional long short-term memory neural network model to take into account the spatial heterogeneity and temporal autocorrelation of the environment at the same time. To address heterogeneous data with machine learning, it is necessary to apply data fusion methods because there is no generic machine learning method suited to all types of problems and data but rather methods that are better tailored to the specificities of each problem (e.g., classification, regression, clustering, and anomaly detection) and data source (e.g., quantitative variables, time series, images, and text) (Alzubi et al., 2018). Data fusion is the process of combining multiple data sources to produce more accurate, robust, and reliable and less redundant information than by considering each data source individually (Khaleghi et al., 2013; Chandrasekaran et al., 2017). Data can be merged at 3 levels (Brena et al., 2020; Hall and Llinas, 2017): 1) at the data level (early fusion) by simply concatenating the initial variables to obtain a dataset with a larger amount of information that can be input into a machine learning algorithm, 2) at the feature level (feature fusion) by extracting more relevant features from each data source and concatenating them to be input into a machine learning algorithm, and 3) at the decision level (late fusion) by aggregating the predictions from models associated with each data source into a global prediction. This principle of fusion of multiple data sources has been applied to machine learning at the feature level (Xue et al., 2017; Amin et al., 2018); for instance, Xue et al. (2017) propose DeepFusion, a multisensor deep learning framework for human activity recognition using feature extraction with convolutional and recurrent neural networks to learn informative representations of heterogeneous sensory data. Data fusion has also been applied to machine learning at the decision level (Peterson et al., 2018; Guo et al., 2019); for example, Guo et al. (2019) propose the iFusion framework for the classification of medical data, which uses CNNs to treat and combine real-time data and heterogeneous data at the decision level. They separately use each type of newly arrived data to train a new discrimination model and fuse the previously trained models to obtain the discrimination result. Most of the studies that use machine learning for FS or related fields use data-level fusion to deal with heterogeneous data, i.e., preprocess their data to jointly combine them by the same machine or deep learning algorithm, which is the case for the studies of Heisenberg et al. (2020) and Lentz et al. (2019) described above. However, fusion at the data level is the most naive and simple to implement and does not allow the extraction of information from each type of data in an optimized way (Hall and Llinas, 2017). Fusion at the feature and decision level has been barely exploited for problems related to FS, a multifactorial domain in which the prediction of indicators requires the use of explanatory data for highly heterogeneous thematics and spatiotemporal scales. The only such study is that of the WFP (WFP-VAM, 2019), which extracts features with a CNN and then inputs the features into a ridge regression. Nevertheless, the WFP study does not take into account the sequential aspect of time series, which is the case for our study. In this study, we use machine and deep learning approaches to address the complexity of our data, and we combine heterogeneous explanatory data by fusion at the three possible levels (data, features, decision) and compare the performance. To the best of our knowledge, this is the first study to propose such complete testing on different data fusion strategies for the prediction of FS scores. Additionally, this is the first study to take into account a large body of heterogeneous data sources. The existing study that uses the most complete set of information sources is the framework developed by the WFP (WFP-VAM, 2019) (cf. Section 2.1), which does not include several important data sources that we take into account in our study, i.e., maize price, population density, soil quality and time series.

3. Data and measures

3.1. Measuring food security

As noted in Section 1, the aspects that contribute to creating food insecurity situations in a country are manifold and often interdependent. This makes the measurement of FS a challenging problem. The scientific community has proposed a plethora of different indicators over the years (up to 450 different FS indicators exist, according to the study in Hoddinott (1999)) that address different aspects of the phenomenon and can thus be considered to some extent complementary to each other. For this reason, using multiple indicators can help obtain more objective results on the FS of a particular country. A well-known definition of FS is the one proposed by Shaw D.J. (2007), stating that FS holds "when all people have, at all times, physical and economic access to sufficient, safe and nutritious food". This definition entails four major aspects:

1. to what extent adequate amounts of food of an appropriate nature and quality are available to the population;
2. to what extent people have access to the resources necessary to acquire the food needed for a nutritious diet;
3. to what extent access to food is stable in cases of unforeseen events (e.g., natural and/or economic crises); and
4. to what extent appropriate use of food can be realized (e.g., storage, cooking and hygiene practices).

Such aspects can help analyse and measure food insecurity situations, which can be taken into account by using information from data sources at multiple scales (national, regional, household or even individual scale).

In this work, we focus on two indicators that are computed based on answers to household surveys: the food consumption score (FCS) and the household dietary diversity score ($HDDS$). These metrics, which are widely used in the scientific literature and by governmental and nongovernmental organizations (Jones et al., 2013; Maxwell et al., 2014; Vhurumuku, 2014), can be used to assess the frequency, quantity and quality of food in a certain area. Detailed definitions for the FCS and $HDDS$ indicators are reported as follows.

Food Consumption Score (FCS): The aim of this indicator is to estimate the cumulative frequency of the different food groups consumed over a period of 7 days within each household taken into account in the survey. FCS can then be considered a proxy of the quantity of nutrients and energy intake. Taking into account the weights reported in Table 1 for the different food groups, FCS can be computed using the following equation:

$$FCS = \sum_{i=1}^{9} x_i.p_i \tag{1}$$

where x_i is the frequency of consumption for each food group i and p_i is the weight of food group i, as reported in Table 1.

Food group	**Weighting**
Cereals and tubers	2
Pulses	3
Vegetables and leaves	1
Fruits	1
Animal proteins	4
Dairy products	4
Sugars	0.5
Oils	0.5
Condiments	0

Table 1: Food groups and their weights for the calculation of the food consumption score (FCS). *Source: (Wiesmann et al., 2009)*

Household Dietary Diversity Score ($HDDS$): The aim of this indicator is to measure food consumption frequency and diversity by focusing on the nutritional quality of the diet. More specifically, $HDDS$ is calculated based on the number of different food groups consumed in the last 24 hours. The categorization used to identify the food groups is not standard and may vary from case to case, depending on the context and on the available data. For instance, in some cases, the same categories taken into account for FCS are used (i.e., those in Table 1), as in the framework proposed in (WFP-VAM, 2019), while in other cases, a different classification including 12 food groups is taken into account, i.e., as proposed by the FAO (Kennedy et al., 2013). In this work, we use the FAO methodology to calculate the $HDDS$, using the categories reported in Table 2. $HDDS$ can then be computed using the following equation:

$$HDDS = \sum_{i=1}^{12} x_i \tag{2}$$

where $x_i = 0$ if food i has not been consumed during the past 24 hours and $x_i = 1$ if food i has been consumed during the past 24 hours.

Food group
Cereals
Roots and tubers
Vegetables
Fruits
Meat products
Eggs
Fish and seafood
Legumes, nuts and seeds
Milk and dairy products
Oils and fat
Sweets
Condiments, spices and drinks

Table 2: Food groups for the calculation of the household dietary diversity score ($HDDS$). *Source: (Kennedy et al., 2013)*

3.2. Study data

In this section, we present the data from which the response variables (i.e., FCS and $HDDS$) are derived and then detail the heterogeneous data used as explanatory variables.

3.2.1. Response data

The response variables, or FS indicators, are derived from the Permanent Agricultural Survey, which has been conducted annually as part of the routine of the Burkinabe Ministry of Agriculture since 1982 in Burkina Faso (Permanent Agricultural Survey, 2015). The survey was conducted using two-stage stratified sampling (villages and households). The sampling frame for the first stage was obtained from the agricultural module of the 2006 general population census. This frame yielded a list of villages (7,871 villages and areas) containing 1,219,241 agricultural households (in 2008, a total of 1,424,909 households were agricultural, representing 81.5 percent of households (Bureau central du recensement général de l'agriculture, 2011)). The sampling frame for farm households is created in each sampled village (selected with probability proportional to the number of farm households) from a household list drawn up each year by enumerating all farm households in the village. For this study, we take into account the data that are available from 2009 to 2018[1]. The resulting dataset contains information from 46,400 farm households, i.e., an average of 4,640 farm households per year distributed in 344 of the 351 communes illustrated in Figure 1. A farm household is defined as a household practising one of the following activities: temporary crops (rainfed and off-season crops), fruit growing, and animal husbandry. As in most surveys, the quality of data obtained from household surveys can be affected by biases. Nonobservation bias related to the failure to acquire information (coverage and sampling bias, nonconsent bias, nonre-

[1] These data are not public and were privately provided to us by the Burkinabe Ministry of Agriculture.

sponse bias) (Lepkowski, 2001) and measurement bias due to measurement error during data collection (linked to the interviewer, the respondent or the questionnaire) (Kasprzyk, 2001) are the two main types of bias. These biases can bring noise to the data and affect the performance of the machine learning algorithms applied to them. This is partly why studies (presented in Section 5) that have predicted FCS and HDDS via machine learning offer low performance. We used these data to build our ground truth by averaging the FCS and $HDDS$ indicators by commune, considering a time window from 2009 to 2018, resulting in 3,066 observations. The other studies that predict these indicators (presented in Section 2) are based on a 1-year time window. To the best of our knowledge, our study is the only one to date to be based on a 10-year time window, which makes it possible to establish decision rules based on interannual variations that are therefore more generalizable over time.

Figure 1: Spatial distribution of the 351 communes of Burkina Faso (background map: Google Maps).

3.2.2. Explanatory data

To address **RQ1** ("what types of publicly available data can be targeted to predict FS scores?"), there exist a large number of proxies indicators related to one or more components of FS that can be taken into account, e.g., vegetation indices, rainfall, food prices, local population densities, soil quality, and numbers of violent events, schools and hospitals (Fritz et al., 2018), that we take into account when selecting the explanatory data for our study.

The multifactorial aspect of FS implies the use of heterogeneous explanatory data to obtain as complete a picture of the situation as possible. FS proxies used as explanatory variables can be considered heterogeneous at three levels.

1. At the *thematic* level, they are related to domains such as remote sensing, meteorology, economy, demography, or soil quality and utilization. This implies having a complete vision of the factors of famine in the place under study.
2. At the *structural* level, there are various types of data: quantitative values, GPS points, line vectors, time series, and rasters. This requires the use of tools and methods appropriate for the processing of each data point.
3. At the *spatiotemporal scale* level, data can be available spatially by region, commune, station or pixel and temporally by decade, year, month or week. This point requires the use of techniques to extract pertinent information at a common scale on which they can be combined

This involves choosing the most suitable spatiotemporal scale, which also implies answering **RQ2** ("how can data that are heterogeneous in terms of thematic, structure, and spatiotemporal resolution be preprocessed to obtain consistent predictions of the FS for a given study site?"). First, FS proxies are preprocessed to extract relevant explanatory variables at the commune scale, which is the smallest administrative boundary for which the response variables are spatialized, enabling the number of examples for model learning to be maximized. Some proxies from raster data or GPS points have a finer granularity and must be aggregated by commune by summing (rainfall), averaging (minimum and maximum temperatures, soil quality), counting (number of hospitals, schools and violent events), taking the maximum (smoothed brightness temperature (SMT), elevation) or more complex aggregations (Gini coefficient, autocorrelation, and differential entropy of population rasters). Other proxies (meteorological data accessible by stations and commodity prices accessible by markets) are available at a coarser granularity and must be interpolated on every commune: in these cases, we chose to use K-nearest neighbour interpolation, which is an accurate and rapid technique. Population and land cover rasters are resampled to a 100 metre resolution, and 10x10 100 metre pixel patches are used as input to a CNN to predict FCS and $HDDS$, which returns predictions and features by commune. The scaling method for each variable is detailed in the "scaling up" column of Table 3. The normalized difference vegetation index (NDVI), a vegetation index, is treated by a culture mask to consider only NDVI in cropping areas[2]. Some variables are normalized by population (e.g., numbers of schools, hospitals, and violent events) or by area (number and length of waterways). Then, each explanatory variable is centred and reduced in relation to communes and years (consists of subtracting the mean and dividing it by the standard deviation). Finally, the explanatory variables obtained are selected by retaining only those that are significantly correlated with the response variable under consideration (p-value less than 0.05). The information on each dataset is available in Table 3; for more details on the variety of data used in the models, see Appendix A. To address the structural heterogeneity of the data, the selected explanatory variables are classified into 4 groups with a similar structure to treat each group with an appropriate machine learning method:

[2]Culture mask: S2 prototype land cover map at 20 m in Africa 2016

- **Time series** that have multiple values per year and one value per commune. They are aggregated into monthly time series (May to November of the year in which the FS indicator is collected and of the previous year)
- **Conjunctural data** that have one value per year and one value per commune.
- **Spatial data** that have one value per commune and are invariant per year.
- **High spatial resolution (HSR) data** that have multiple values per commune. The values are 10x10 100 m pixel patches extracted from each data source.

The aim of this categorization into 4 groups is to make the different categories of explanatory variables suitable for independent processing by different branches of the framework, each based on specific machine learning techniques (i.e., each branch will be based on the most suitable technique for the specific data type, as will be detailed in the next section).

Variable	Resolution	Frequency	Source	Scaling up
Time series [several values per year; one value per commune] [70 vars]				
Smoothed brightness temperature (SMT) [14 vars]	4 km	7 days	National Oceanic and Atmospheric Administration (NOAA)	Maximum
Rainfall [14 vars]	6 km	10 days	Tropical Rainfall Measuring Mission (TRMM)	Sum
Average minimum and maximum temperatures [2x14 vars]	21 km	1 month	WorldClim	Mean
Maize price [14 vars]	64 markets	1 month	Société Nationale de Gestion du Stock de Sécurité alimentaire (SONAGESS)	K-nearest neighbour interpolation
Conjunctural data [one value per year; one value per commune] [20 vars]				
Meteorological data [7 vars]	10 stations	1 year	Knoema platform	K-nearest neighbour interpolation
Population density [4 vars]	100 m	1 year	Afripop	Spatial autocorrelation 2 km and 5 km, Gini, entropy
Economic data [7 vars]	Country	1 year	World Bank	Country value
Normalized difference vegetation index [2 vars]	250 m	1 year	Modis	Mean
Spatial data [one value per commune] [13 vars]				
Hospitals, schools [2 vars]	Point vectors	2018	Open Street Map	Count
Violent events [4 vars]	Point vectors	2018	Armed Conflict Location & Event Data Project (ACLED)	Count
Soil quality [3 vars]	1 km	2008	Food and Agriculture Organization (FAO)	Mean
Waterways [2 vars]	Line vectors	2008	Digital Chart of the World (DCW)	Count, length
Elevation data [2 vars]	1 km	2018	Consultative Group on International Agricultural Research (CGIAR)	Maximum, variance
High spatial resolution data [several values per commune] [4 vars]				
Population density	100 m	1 year	Afripop	CNN
Land cover (crops, forests, building areas)	20 m	2016	European Space Agency	CNN

Table 3: Summary of the datasets

4. *FSPHD* Framework

To address **RQ3** ("how can state-of-the-art machine and deep learning approaches be exploited and combined to treat such heterogeneous data?"), in this section, we define our proposed machine learning framework, namely, Food Security Prediction based on Heterogeneous Data ($FSPHD$), devised to estimate FCS and $HDDS$. The aim of the proposed framework is to integrate several state-of-the-art machine and deep learning techniques able to exploit the full potential of the large body of heterogeneous data that are input. To this end, we propose two types of regression models (cf. Figure 2) to predict FCS and $HDDS$, which correspond to two different variants of the proposed framework, namely, model *(a)* and model *(b)*. Models *(a)* and *(b)* are inspired by machine learning studies using decision-level fusion (Peterson et al., 2018; Guo et al., 2019) and feature-level fusion (Xue et al., 2017; Amin et al., 2018), respectively, to process and combine heterogeneous explanatory variables. Both feature-level and decision-level fusion approaches have been proven to be effective in several application domains (e.g., medicine (Amin et al., 2018; Guo et al., 2019), human activity recognition (Xue et al., 2017), and chemistry (Peterson et al., 2018)). Therefore, given the exploratory nature of this work, we preferred to propose two variants of the framework rather than just choosing one of the available fusion approaches. Another key point of the framework design is that each branch is designed to integrate the most suitable method for each specific data type, i.e., according to what has been observed in the existing literature. The challenge is to extract, with each branch, complementary information on FS from each data source. Conjunctural and spatial (CS) data are classical numerical data (i.e., a single numerical value for each observation of the response variable) and are processed by an RF, which is one of the machine learning methods offering the best compromise between performance and interpretability when the data are not complex. Time series data are processed using an LSTM architecture, which is a proven machine learning method for sequential data processing owing to its feedback connections. (Song et al., 2020). The HSR data are input into a CNN, which is a neural network method tailored for the analysis of visual imagery (Huang et al., 2018). More specifically, model *(a)* and model *(b)* of $FSPHD$ are structured as follows:

- **Model (a)**: We apply a linear model (LM) with ridge regularization on the responses of the three machine and deep learning models: the response of the LSTM on the time series, the response of the CNN on the HSR data and the response of the RF on the CS data. This model is based on decision-level fusion, aggregating predictions obtained with different strategies to obtain a more robust overall prediction.

- **Model (b)**: We use an RF on the features extracted by the deep learning models. This model is based on feature-level fusion, which allows complex data (i.e., time series and HSR data) to obtain new representations better correlated to the response variable and more efficiently processable by an RF.

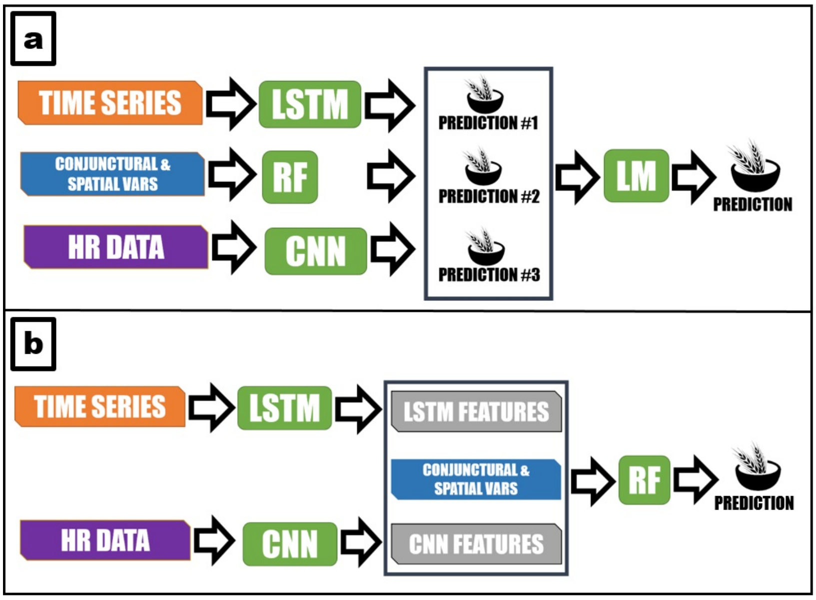

Figure 2: Visual schema of the $FSPHD$ framework, which combines heterogeneous explanatory data to predict the food consumption score and the household dietary diversity score. Models (a) and (b) are based on decision and feature fusion, respectively.

5. Experimental Evaluation

5.1. Competing methods and ablations

To evaluate the performance of the $FSPHD$ framework, we compare it to the performance of several baselines, ablations and competing methods.

We use the following two independent studies as our competing methods:

- As a first competing method, we select the WFP framework introduced in Section 2.1 (WFP-VAM, 2019). This study examines the regression of FCS and $HDDS$ obtained from a survey of 3,650 Burkinabe households. Households are aggregated in 567 geolocalized villages. They use data from various sources (Open Street Map, Google Maps, Sentinel 2, ACLED) as explanatory variables. Village-scale features are extracted from each data source with an appropriate method: images are processed by a CNN and the shortest distances from a village to a school and a hospital and the number of violent events at 10 km are extracted from GPS data. The features are input from a PCA, and the first 10 main components from a ridge regression are finally input for FCS and $HDDS$ predictions. We ran their framework (using the public code) with their data (personal communication) to obtain the results.

- The second competing method is that from a study conducted by (Lentz et al., 2019) (cf. Section 2.1). They predict FCS and $HDDS$ in Malawi using data obtained from the 2010 Living Standards Measurement Survey (LSMS) for training and using data obtained from the 2013 LSMS for testing. The data for 2010 (training) and 2013 (testing) contain 12,270 and 3,999 observations, respectively, which are aggregated in 768 and 204 villages. They use linear regressions and compare the performance of their models using only open and free data and adding data from the previous LSMS survey. In our study, we take into account the performance of models that incorporate only open data in accordance with our objectives. The data used are from diverse sources: meteorology, precipitation, market prices and soil quality.

We also define three baseline models to test simpler machine and deep learning techniques on different subsets of explanatory variables:

- Type (c) Model: We apply an RF directly on initial variables: time series only, CS variables only and both types of data. These models are based on data-level fusion.
- Type (d1) Model: We apply an LSTM suitable for the treatment of time series.
- Type (d2) Model: We apply a CNN suitable for HSR data.

Finally, to further investigate the contribution of each group of features to the final solution, we define three ablations of the type (b) model by applying it as follows:

- only on LSTM features
- only on CNN features
- only on CNN features and CS variables

5.2. Experimental Setting

The $FSPHD$ framework has been implemented using TensorFlow 1.15 under Python 3.7, and the code is publicly available on GitHub[3]. The RF is set up with 900 trees with a maximum depth of 20. The LSTM is parameterized with 2 layers of 64 neurons. The cost function used is the mean square error, and the optimizer is based on the FTRL algorithm. The CNN is configured with 3 convolution layers with 32, 64 and 128 filters. A max

[3]https://github.com/pipapou/FSPHD_Code

pooling of dimension 2 is placed after each convolution layer. The cost function used is the mean square error, and the optimizer is based on the Adam algorithm. For model (b) using feature fusion, 64 features are extracted from the LSTM, and 128 features are extracted from the CNN. LSTM and CNN are trained from scratch using a batch size of 250 as well as 1,000 and 100 epochs, respectively. To assess the performance, we randomly select 85% of the dataset for model learning and 15% for testing by repeating this procedure 5 times and calculating the average performances. We use Rˆ2 to evaluate the regression performance.

5.3. Quantitative Results

Table 4 reports the quantitative performance (Rˆ2) of the two variants of the $FSPHD$ framework and all the competing methods. It can be seen how $FSPHD$ outperforms all the competing methods and baselines, with model *(b)* (i.e., the feature fusion approach) outperforming model *(a)* (i.e., the decision-level fusion approach). Even though the predictions obtained with the $FSPHD$ framework are still not accurate enough to be used in operational contexts, the values of Rˆ2 obtained for FCS (0.469) and $HDDS$ (0.434) are undoubtedly statistically significant, outperforming all the competing methods and thus proving the benefits of the integration of different data science techniques for a large body of heterogeneous data.

Competing methods

The results obtained by the WFP framework in the Burkina Faso area (cf. Table 4) are relatively modest (0.34 for FCS and 0.30 for $HDDS$). As already stated in Section 2.1, the results of the WFP framework seem to be extremely data dependent, so the same framework may obtain comparable results in other countries (e.g., Senegal, Sierra Leone) (WFP-VAM, 2019). Regarding the study of Lentz et al., we note that the Rˆ2 associated with FCS and $HDDS$ is even lower, not exceeding 0.2.

Baseline models

We observe that the models of type (c) (only based on the use of RF directly on explanatory variables) give performances that are already significant and close to those of more sophisticated models; by integrating only the CS variables, we obtain Rˆ2 of 0.414 and 0.401 for FCS and $HDDS$, respectively. This validates the data source selection and preprocessing applied to the data in use. The type (d1) and (d2) models aim at processing data with complex structures (time series and HSR images) with a suitable deep learning method. The LSTM does not succeed in highlighting the sequential aspect of time series, offering lower performance than the RF on the same data (i.e., model (c) on the time series): 0.232 vs 0.241 for FCS and 0.223 vs 0.237 for $HDDS$. Our hypothesis is that in the present case, where we have noisy response variables, the LSTM capable of detecting complex patterns in sequential data overintercepts the noise despite our settings and therefore overfits the data. Future work should focus on other methods to improve the consideration of time series. The CNN on HSR data gives interesting performance results (= 0.34 for FCS and 0.392 for $HDDS$); these values are significantly improved by putting the features extracted from the CNN into an RF (0.434 for FCS and 0.418 for $HDDS$).

$FSPHD$ **framework**

Variants (a) and (b) of $FSPHD$ represent two strategies for combining different types of data. The model of type (a), which consists of aggregating the responses of the models with a linear model, has moderate performance for FCS (0.375) and greater performance for $HDDS$ (0.426), which shows the contribution of the type (a) model using decision-level fusion for the $HDDS$ prediction. The model of type (b) consists of aggregating the features of the models with an RF. This solution obtains the best performance by combining the features of the CNN with the CS variables (0.469 for FCS and 0.434 for $HDDS$). The performance is then significantly better than that when processing CNN features and CS variables separately with an RF, demonstrating that these two types of data provide additional information on FS and the ability of the type (b) model using feature-level fusion to extract this information.

Model	Model type	FCS	HDDS
Competing methods			
WFP study	Feature fusion	0.34	0.30
Lentz et al. study	Data fusion	0.16	0.18
Data-level fusion			
RF (Time series)	(c)	0.241	0.237
RF (CS vars)	(c)	0.414	0.401
RF (Time series + CS vars)	(c)	0.339	0.326
Suitable deep learning method			
LSTM (Time series)	(d1)	0.232	0.223
CNN (High spatial resolution data)	(d2)	0.34	0.392
$FSPHD$ ***Framework***			
Decision-level fusion			
LM (RF, LSTM and CNN responses)	(a)	0.375	0.426
Feature-level fusion			
RF (LSTM features)	(b)	0.194	0.181
RF (CNN features)	(b)	0.434	0.418
RF (CNN features + CS vars)	(b)	**0.469**	**0.434**
RF (CNN and LSTM features + CS vars)	(b)	0.455	0.43

Table 4: Performance (Rˆ2) of the $FSPHD$ framework, competing methods and ablations for food consumption score and household dietary diversity score prediction. The column “model type” designates the type of model used, according to the model categorization in Section 5.

5.4. Model interpretation

Due to the poor performance associated with the LSTM, it is not appropriate to use features importance techniques on it to deduce relevant and reliable information. Describing the complex spatial patterns of the CNN is also complicated because of its black box effect, which makes it also unsuitable for the use of classical important features techniques. We can simply say that the land use and population dynamics used at the CNN input seem to play an important role in FS, knowing the good CNN performance. In order to obtain information on the importance of the variables processed by both neural networks (i.e. CNN and LSTM), we considered the model of type (b) which combines by feature fusion the features obtained by the LSTM and the CNN and the CS variables. We computed on this model the permutation importance of all the variables processed by the RF (results available on GitHub[4]). The permutation importance is defined as the decrease in a model score when single feature values are randomly shuffled in the test set and is usually used in the literature to perform regressions in machine learning (Grömping, 2015). As expected, the most important variables are the CNN features while the features derived from LSTM and CS variables generally appear in lower positions in the ranking. Note that these features are the result of a complex representation learning process on the original input data, so that interpretability is limited to the importance of each model (i.e., by identifying which branch of the model the most important variables come from), and not of the original variables (e.g., conversely to what seen

[4]https://github.com/pipapou/Permutation-importance-for-FCS-and-HDDS

for RF in Table 5). The interpretability of neural networks, which is a current issue, should be the subject of future work. Several methods have been recently developed to explain the predictions of neural network models, by identifying the input variables (e.g., pixels, time series items) that have most contributed to the model decision (Montavon et al., 2018; Khormuji and Rostami, 2021), e.g., sensitivity analysis, occlusion analysis and layer-wise relevance propagation (LRP). However, these methods are not suitable for our experimental setting, since they are tailored to classification problems, while our study focuses on regression. For CS variables that are directly processed by the RF, the significance of variables can be approached by the permutation importance. The top 10 ranks of CS variables according to their permutation importance for FCS and $HDDS$ are shown in Table 5. We note that variables from multiple domains are included in these two top 10 lists: landscape structure (3 variables), population dynamics (2 variables), soil quality (2 variables), meteorological (2 variables), vegetation (1 variable), insecurity (1 variable), sanitary (1 variable) or economic (1 variable) variables, which confirms the importance of the combined use of data sources from multiple domains. Seven variables are included in the top 10 of both FCS and $HDDS$, and they seem essential for the prediction of FS in our case.

Among these variables, we find:

- The average NDVI of the year preceding the survey. The NDVI, which is an indicator of vegetation quality, is considered only for crop areas, so it is related to the quality of agricultural crops. An interesting fact is that the average NDVI of the previous year is more important than that of the same year, which is situated near these top 10.

- Three variables related to the landscape structure: the total length of the rivers, which allows us to evaluate the availability of water, and the maximum and variance of the altitude. The structure of the relief has an evident impact on agriculture (e.g., accessibility of cultivated areas to agricultural development and water, specific types of plantations at certain altitudes).

- Two variables express the population dynamics: the spatial autocorrelations at 2 km and the differential entropy associated with the populations; the population density and movements can create pressures for FS. This confirms, with the good performance of the CNN that takes land use and population data as input, the importance of these data sources.

- A soil quality variable: the nutrient retention capacity, which is directly linked to the availability of agricultural produce.

We also find that some variables are specific to a single FS indicator. The top 5 of the FCS contains 2 variables that are absent from the $HDDS$ top 10: the gross national expenditure of the country and the number of violent events, which seem to be more specific to the amount of nutrients consumed. Conversely, the average maximum temperature per day is present in the $HDDS$ top 5 and absent from the FCS top 10, so this variable seems to be specifically related to the quality and diversity of the diet. Our explanation is that crop diversification, and consequently diet diversification, is dependent on weather and temperature.

Rank	FCS	HDDS
1	Population entropy	Soil quality (Nutrient retention capacity)
2	Gross national expenditure	Average NDVI of previous year
3	Average NDVI of previous year	Maximum elevation
4	Maximum elevation	Average maximum temperature per day
5	Total violent events	Population entropy
6	Variance of elevation	Maximum relative humidity
7	Population spatial autocorrelation 2 km	Total length of waterways
8	Total length of waterways	Variance of elevation
9	Soil quality (nutrient retention capacity)	Number of hospitals
10	Soil quality (rooting conditions)	Population spatial autocorrelation 2 km

Table 5: Top 10 ranks of CS variables processed by RF according to their permutation importance for FCS and $HDDS$.

6. Conclusion

This study proposes the $FSPHD$ framework, which uses machine learning methods to predict by regressions two key FS indicators that are normally obtainable through costly and lengthy household surveys. To validate our models, we used as labels a database including households from across Burkina Faso from 2009 to 2018 using mostly global public data as explanatory variables to ensure the replicability and generalizability of our methods to other countries. We faced two scientific obstacles: 1) The multifactorial aspect of FS implies the choice of heterogeneous input data (at the thematic, structural and spatiotemporal scale levels) and suitable preprocessing to maximize the contribution of each data point. To take into account a maximum number of facets of FS, we integrated data from different topics (landscape structure, population dynamics, soil quality, meteorology, vegetation, insecurity, or economy), encoded in different types (quantitative values, GPS points, line vectors, time series, and images) and with different spatiotemporal granularities, and we had to perform suitable treatments to extract relevant information from these data (e.g., aggregations, interpolations, normalizations). 2) Machine learning methods suitable for each type of variable had to be chosen and combined. We have seen that the performance obtained by our models (Rˆ2) is not high, not exceeding 0.469 or 0.434 for Rˆ2 in the prediction of FCS and $HDDS$, respectively. However, the results of this study are superior to most of the rare works we were able to compare it with. These results indicate that the prediction of these FS indicators is a complex issue. This study adds another layer to this little-studied issue. We have also observed the modest but significant contribution of deep learning models (CNNs) to the processing of HSR data. The use of this type of data in the context of FS is therefore a relevant avenue for future work. However, the use of deep learning models (LSTM) for time series processing did not yield significant results, and future research will include better treatment of time series because the temporal aspect plays an important role in FS. Additionally, we have demonstrated the contribution of models combining different types of machine learning methods suited to each type of data, which confirms the contribution of this kind of approach. The models of type (b) based on feature-level fusion allow us to obtain the best performance. Finally, we have observed that the variables indicated by the models as being the most important for the prediction of FS indicators come from multiple domains, which confirms the need to link FS with a large spectrum of related domains to obtain the most comprehensive picture possible of this complex and multifactorial concept. Future work will consist of improving the architecture of deep learning models and integrating other types of data such as textual data (social networks, journals, etc.) to better take into account all factors of FS. We will also focus on the explanatory aspect of the models, made difficult by the black box effect of neural networks. For this, we are currently working on the integration of textual data with high explicative potential. We finally expect close collaborations with researchers from other disciplines to enable us to obtain a qualitative evaluation protocol. Data mining approaches only find their relevance in a close partnership with the various stakeholders, whether at the level of data production and preprocessing or at the level of validation of the knowledge that has been extracted. A project of this scale can only be fully realized in a close disciplinary partnership, in which scientists, researchers and users collaborate in a permanent mutual listening exercise.

Appendix A. Datasets

Time series data
Monthly smoothed brightness temperature (SMT) (May to November)
Monthly total rainfall (May to November)
Monthly average minimum temperature (°C) (May to November)
Monthly average maximum temperature (°C) (May to November)
Monthly maize price (May to November)
Meteorological data
Average sunlight duration per day
Maximum relative humidity
Minimum relative humidity
Average maximum temperature per day
Average minimum temperature per day
Evaporation (mm)
Annual precipitation (mm)
Population density data
2 km spatial autocorrelation
5 km spatial autocorrelation
Gini index
Differential entropy
World Bank Economic data
Foreign direct investment net inflows (% of GDP)
Foreign direct investment net outflows (% of GDP)
Gross national expenditure (% of GDP)
Households and NPISH final consumption expenditure (% of GDP)
Military expenditure (% of GDP)
Merchandise trade (% of GDP)
GDP per capita growth (% per annum)
Normalized difference vegetation index (NDVI)
Average NDVI from May to November of the year in which the response variable was collected
Average NDVI from May to November of the year preceding the collection of the response variable
Hospitals and schools
Number of hospitals per 1,000 inhabitants
Number of schools per 1,000 inhabitants
Violent events
Total number of violent events per 1,000 inhabitants
Number of protests per 1,000 inhabitants
Number of riots per 1,000 inhabitants
Number of violent events against civilians per 1000 inhabitants
Soil quality
Nutrient retention capacity
Rooting conditions
Oxygen availability to roots
Waterways
Number of waterways
Total length of waterways per km2
Elevation
Maximum elevation
Variance of elevation
High spatial resolution data
Population density 10x10 100 m patches
Land cover - crops 10x10 100 m patches
Land cover - forests 10x10 100 m patches
Land cover - building areas 10x10 100 m patches

Acknowledgements

This study was conducted with the help of the Ministry of Agriculture, Water Resources, Sanitation and Food Security of Burkina Faso, the World Food Program (WFP) and the Société Nationale de Gestion du Stock de Sécurité Alimentaire (SONAGESS), which provided data. This work was supported by the French National Research Agency under the Investments for the Future Program #DigitAg, referred to as ANR-16-CONV-0004.

References

Alzubi, J., Nayyar, A., Kumar, A., 2018. Machine learning from theory to algorithms: an overview, in: Journal of physics: conference series, IOP Publishing. p. 012012.

Amin, J., Sharif, M., Raza, M., Yasmin, M., 2018. Detection of Brain Tumor based on Features Fusion and Machine Learning. Journal of Ambient Intelligence and Humanized Computing .

Barbosa, R.M., Nelson, D.R., 2016. The Use of Support Vector Machine to Analyze Food Security in a Region of Brazil. Applied Artificial Intelligence 30, 318–330.

Benedetti, P., Ienco, D., Gaetano, R., Osé, K., Pensa, R., Dupuy, S., 2018. M3fusion: A deep learning architecture for multi-Scale/Modal/Temporal satellite data fusion. ArXiv.

Brena, R.F., Aguileta, A.A., Trejo, L.A., Molino-Minero-Re, E., Mayora, O., 2020. Choosing the Best Sensor Fusion Method: A Machine-Learning Approach. Sensor 20.

Bureau central du recensement général de l'agriculture, 2011. Rapport général du module tronc commun; Phase 2 RGA 2008. Ministère de l'agriculture et de l'hydraulique du Burkina Faso.

Chandrasekaran, B., Gangadhar, S., Conrad, J., 2017. A survey of multisensor fusion techniques, architectures and methodologies, in: SoutheastCon 2017, IEEE. pp. 1–8.

Deléglise, H., Bégué, A., Interdonato, R., Maître d'Hôtel, E., Roche, M., Teisseire, M., 2020. Linking heterogeneous data for food security prediction , in: Data Integration and Applications Workshop (DINA 2020), p. 12.

FAO, ECA, 2018. Addressing the threat from climate variability and extremes for food security and nutrition. Technical report.

FAO, FIDA, OMS, WFP, UNICEF, 2018. The State of Food Security and Nutrition in the World - Building climate resilience for food security and nutrition. Technical report.

FAO, FIDA, OMS, WFP, UNICEF, 2020. The State of Food Security and Nutrition in the World - Transforming Food Systems for Affordable Healthy Diets. Technical report.

Fritz, S., See, L., Laso Bayas, J.C., 2018. A comparison of global agricultural monitoring systems and current gaps. Agricultural Systems 168, 258 – 272.

Grömping, U., 2015. Variable importance in regression models. Wiley interdisciplinary reviews: Computational statistics 7, 137–152.

Guo, K., Xu, T., Kui, X., Zhang, R., Chi, T., 2019. iFusion: Towards efficient intelligence fusion for deep learning from real-time and heterogeneous data. Information fusion 51.

Hall, D., Llinas, J., 2017. Handbook of Multisensor Data Fusion: Theory and Practice. CRC Press.

van der Heijden, W., van den Homberg, M., Marijnis, M., de Graaff, M., Daniels, H., 2018. Combining Open Data and Machine Learning to predict Food Security in Ethiopia.

Heisenberg, G., da Rocha, L.F., Caspersen, L., Wöhrle, S., 2020. Deep learning approach for the prediction of food insecurity. 4th Global Food Security conference .

Hoddinott, J., 1999. Choosing Outcome Indicators Of Household Food Security. International Food Policy Research Institute.

Huang, B., Zhao, B., Song, Y., 2018. Urban land-use mapping using a deep convolutional neural network with high spatial resolution multispectral remote sensing imagery. Remote Sensing of Environment 214, 73 – 86.

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, 790–794.

Jones, A.D., M. Nguren, F., Pelto, G., Young, S.L., 2013. What Are We Assessing When We Measure Food Security? A Compendium and Review of Current Metrics. Advances in Nutrition 4, 481–505.

Kasprzyk, D., 2001. Measurement error in household surveys : sources and measurement. United States Federal Committee on Statistical Methodology.

Kennedy, G., Ballard, T., Dop, M.C., 2013. Guide pour mesurer la diversité alimentaire au niveau du ménage et de l'individu. FAO. Technical report.

Khaleghi, B., Khamis, A., Karray, F.O., Razavi, S.N., 2013. Multisensor data fusion: A review of the state-of-the-art. Inf. Fusion 14, 28–44.

Khormuji, H.B., Rostami, H., 2021. Fast multi-resolution occlusion: a method for explaining and understanding deep neural networks. Applied Intelligence 51, 2431–2455.

Lacher, W., 2012. Organized crime and conflict in the sahel-sahara region. Carnegie Endowment for International Peace. Technical report.

Lentz, E., Michelson, H., Baylis, K., Zhou, Y., 2019. A data-driven approach improves food insecurity crisis prediction. World Development 122, 399 – 409.

Lepkowski, J.M., 2001. Non-observation error in household surveys in developing countries. United States Federal Committee on Statistical Methodology.

Lukyamuzi, A., Ngubiri, J., Okori, W., 2018. Tracking food insecurity from tweets using data mining techniques, in: Proceedings of the 2018 International Conference on Software Engineering in Africa - SEiA '18, pp. 27–34.

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.

Min, W., Ping, L., Lingfei, Z., Yan, C., 2019. Stock market trend prediction using high-order information of time series. IEEE Access 7, 299–308.

Miotto, R., Wang, F., Wang, S., Jiang, X., 2017. Deep learning for healthcare: review, opportunities and challenges. Briefings in bioinformatics 19, 1236—-1246.

Montavon, G., Samek, W., Müller, K.R., 2018. Methods for interpreting and understanding deep neural networks. Digital Signal Processing 73, 1–15. doi:https://doi.org/10.1016/j.dsp.2017.10.011.

Mumtaz, A., Ravinesh, C.D., Nathan, J.D., Tek, M., 2018. Multi-stage committee based extreme learning machine model incorporating the influence of climate parameters and seasonality on drought forecasting. Computers and Electronics in Agriculture 152, 149–165.

Okori, W., Obua, J., 2011. Supervised Learning Algorithms For Famine Prediction. Applied Artificial Intelligence 25, 822–835.

P. Lewis, D., Jebara, T., Stafford Noble, W., 2006. Support vector machine learning from heterogeneous data: an empirical analysis using protein sequence and structure. Bioinformatics 22, 2753–2760.

Permanent Agricultural Survey, 2015. Résultats Défnitifs de la Campagne Agricole 2014/2015 et Perspectives de la Situation Alimentaire et Nutritionnelle. Ministère de l'Agriculture, des Ressources Hydrauliques, de l'Assainissement et la Sécurité Alimentaire. Technical report.

Peterson, K.T., Sagan, V., Sidike, P., Hasenmueller, E.A., Sloan, J.J., Knouft, J.H., 2018. Machine Learning-Based Ensemble Prediction of Water-quality Variables Using Feature-level and Decision-level Fusion with Proximal Remote Sensing. Photogrammetric Engineering & Remote Sensing 85, 269–280.

Qi, Y., 2012. Random forest for bioinformatics, in: Ensemble machine learning. Springer, pp. 307–323.

Shailesh, M.P., Tushar, A., Narayanan, C.K., 2018. Multi-Task Deep Learning for Predicting Poverty from Satellite Images (IAAI18). The Thirtieth AAAI Conference on Innovative Applications of Artificial Intelligence .

Shaw D.J., 2007. World Food Security: A History Since 1945. volume 1. Palgrave MacMillan.

Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., Jiang, L., Cheng, Z., 2020. Time-series well performance prediction based on long short-term memory (lstm) neural network model. Journal of Petroleum Science and Engineering 186.

Tapsoba, A., Combes Motel, P., Combes, J.l., 2019. Remittances , food security and climate variability : The case of Burkina Faso. Working Papers. HAL.

Valdés, J.J., 2018. Extreme learning machines with heterogeneous data types. Neurocomputing 277, 38–52.

Vhurumuku, E., 2014. Food security indicators - FAO. Integrating Nutrition and Food Security Programming for Emergency response workshop. Technical report.

WFP-VAM, 2019. Humanitarian High Resolution Mapping - Complementing Assessments with Remote Sensing Open Data. `https://wfp-vam.github.io/HRM/`. [Online; accessed 28-January-2021].

Wiesmann, D., Bassett, L., Benson, T., Hoddinott, J., 2009. Validation of the world food programme's food consumption score and alternative indicators of household food security. International Food Policy Research Institute (IFPRI). Technical report.

Xue, H., Jiang, W., Miao, C., Yao, Y.Y., Ma, F., Ma, X., Wang, Y., Yao, S., Xu, W., Zhang, A., Su, L., 2017. DeepFusion: A Deep Learning Framework for the Fusion of Heterogeneous Sensory Data. 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) .

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., Burke, M., 2020. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. Nature Communications 11.

Yuan, Z., Zhou, X., Yang, T., 2018. Hetero-convlstm: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data, in: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, Association for Computing Machinery, New York, NY, USA. p. 984–992.

- The analysis of Food Security related phenomena poses several research challenges
- We focus on the Food Consumption Score and Household Dietary Diversity Score indicators
- We propose the FSPHD machine learning framework for the prediction of such indicators
- We use a large set of input data heterogeneous in terms of format and domain
- The results show promising performances that outperform competing methods

Hugo Deléglise: Methodology, Software, Writing - Original Draft. **Roberto Interdonato:** Conceptualization, Formal analysis. **Agnès Bégué**: Investigation, Resources. **Elodie Maître d'Hôtel:** Methodology, Validation. **Maguelonne Teisseire:** Visualization, Supervision. **Mathieu Roche:** Writing - Reviewing and Editing, Supervision.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: