P.5C.02

Heterogeneous data analysis for the prediction of food security indices

Simon Madec, Ulrich Mariel Ayena, Agnes Begue, Roberto Interdonato
UMR TETIS, CIRAD, Montpellier, France

Intro

Food insecurity in **Burkina Faso** and West Africa is one of the major development challenges of the region. The causes of this food insecurity are manifold: significant inter-annual variability in rainfall affecting agricultural production (mostly rain-fed), terrorist threats in border areas with Mali, Benin, and Niger, demographic explosion, and structural poverty affecting a large part of the population. Food security indices, such as the Food Consumption Score (FCS), measured by household survey can help to measure and quantify help needed to gauge these issues. However, surveys like EPA, CFSVA and LSMS demand significant resources and can not anticipate food insecurity, especially in conflict zones.

The objective of the study is to develop a methodological framework for predicting the **Food Consumption Score** (FCS) using **heterogeneous data** characterized by different types, **spatial**, and **temporal** scales, through the use of advanced **data science** and **machine learning** techniques.

Fig1 : Burkina Faso with their 352 disctricts where FCS are aggregated

Methodology

Food consumption score

We use data from the **2010-2020** period of the Permanent Agricultural Survey (**EPA**), a rural household survey conducted annually. It contains information on approximately 5,000 farm households per year, including data on food consumption.

The Food Consumption was derived from the survey. It is computed by considering the variety and frequency of the different food groups consumed over the past 7 days represented (Table 1). It is computed with the following equation:

$$FCS = \sum_{k=1}^{n} x^k . w^k$$

Where x^k is the frequency of consumption of each food group and w^k represents the weights reflecting the relative importance of each food group (Table 1).

Food Group	Weight
Cereal and tuber	2
Legume	3
Vegetables and leaves	1
Fruits	1
Animal Protein	4
Dairy Products	4
Sugars	0.5
Oils	0.5
Condiments	0

Table 1 : Food groups and associated weights

Machine learning framework

- We employ a three-branch model architecture.
- LSTM is used to extract features from the times series data
- High spatial resolution patches are processed with convolutional neural networks while conjunctural and spatial data are directly used as features. The different features are merged with a concatenation. Finally, a random forest is finally used to predict FCS scores (Fig. 3)

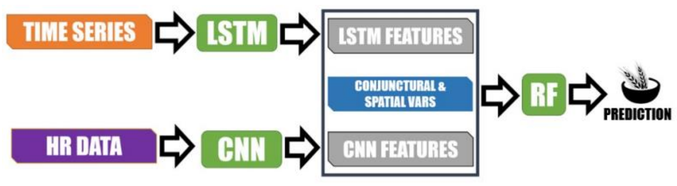

Fig 3 : Model architecture

Heterogeneous data (explanatory variables)

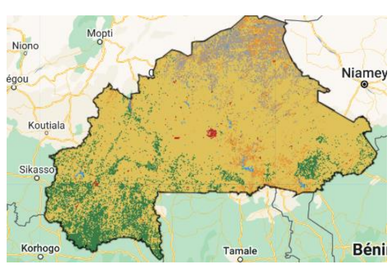

Fig 2 : Example of two rasters : Land use (left) and Population density (right) (High spatial resolution data)

- Different Publicly available heterogeneous datasets were fetched.
- They include rasters (e.g., population density, land use, soil quality), GPS points (hospitals, schools, violent events), line vectors (waterways), quantitative variables (economic variables, meteorological data) and time series...
- The data were processed and categorized into 4 groups (Table 2). This allows independent processing by different branches of the framework.

Time series data (monthly averaged)	
	Maximum and Minimum Temperature
	Total rainfall
	Smoothed brightness temperature
	Maize price
Conjunctural data [one value per year; one value per commune]	
	Meteorological Data (Sunlight, humidity, evoparation, annual precipitation)
	Population Density
	Vegetation Indices
	Soil Quality
Spatial data [one value per commune]	
	Hospitals and Schools
	Waterways (Number and total lenght)
	Altitude Data
	Elevation data
	Violent Events
High spatial resolution data [several values per commune]	
	Land Use (Crop, Forest, Built-up Areas...)
	Population Density

Table 2 : Summary of the variables used as features

Results and Discussion

- A random Train / Test splits was used for a quantative evaluation of the different branches of theframework (Table 3.) The R^2 are reported (Table 3)
- Figure 4 present an accuracy of the models when training with all years before the year N and Testing on year N for $N \in [2011 - 2020]$. For this test we report the accuracy for a classification of the FCS in three class.

Branch (s)	R^2
Time series	0.23
CS	0.41
Spatial	0.34
Time series + CS + Spatial	0.45

Table 3 : Performance of the framework for different features : Time Series, Conjonctural & Spatial variables (CS) and high spatial resolution data (Spatial).

The results highlight the complexity of the FCS, with up to 0.45 of the variation explained. Estimating FCS from publicly available datasets remains a challenge due to the relatively small survey datasets and the high dimensionality of the extracted features. Biases and noise in the reference dataset are also present (survey methodology).

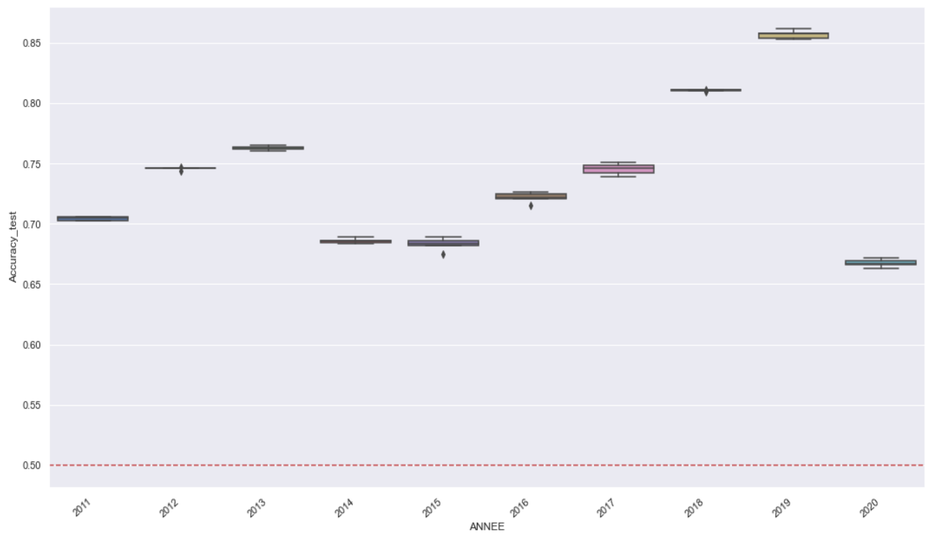

Figure 4: Accuracy of the model when training with year (N-1) and Testing on year N for $N \in [2011 - 2020]$

For evaluation robustness and benchmarking purposes, additional research is now being conducted with surveys in Rwanda and Tanzania (rHomiz, CFSVA, LSMS).

References

Deléglise, H., Bazié, Y.G., Bégué, A., Interdonato, R., Roche, M., Teisseire, M. and Maître d'Hôtel, E., 2022. Validity of household survey indicators to monitor food security in time and space: Burkina Faso case study. Agriculture & Food Security, 11(1), p.64.
Deléglise, H., Interdonato, R., Bégué, A., d'Hôtel, E.M., Teisseire, M. and Roche, M., 2022. Food security prediction from heterogeneous data combining machine and deep learning methods. Expert Systems with Applications, 190, p.116189.

cirad tetis TERRITOIRE ENVIRONNEMENT TÉLÉDÉTECTION INFORMATION SPATIALE