Linking heterogeneous data for strengthening food security systems

Hugo Deléglise [1,3], Roberto Interdonato [1,3], Agnès Bégué [1,3], Maguelonne Teisseire [1,5], Mathieu Roche [1,3] , Elodie Maître d'Hôtel [2,4]

[1]TETIS, Univ Montpellier, AgroParisTech, CIRAD, CNRS, INRAE, Montpellier, France. [2]MOISA, Univ Montpellier, CIHEAM-IAMM, CIRAD, INRAE, Institut Agro, Montpellier, France. [3]CIRAD, UMR TETIS, F-34398 Montpellier, France. [4]CIRAD, UMR MOISA, F-34398 Montpellier, France.
[5]INRAE, Montpellier, France.

Data

- **Hunger in Africa** is growing again after several years of decline
- **Burkina Faso** is one of the most food-insecure countries in West Africa
- **A number of reasons** account for this situation (climate, conflicts, economic downturn, etc.)
- Using **heterogeneous data** can help to adress this issue from a broader perspective

We use as **ground truth** two food security indicators: the Food Consumption Score (**FCS**) and the Household Dietary Diversity Score (**HDDS**) collected from 46400 households between 2009 and 2018 and aggregated in 344 of the 351 communes of Burkina Faso (3066 observations).

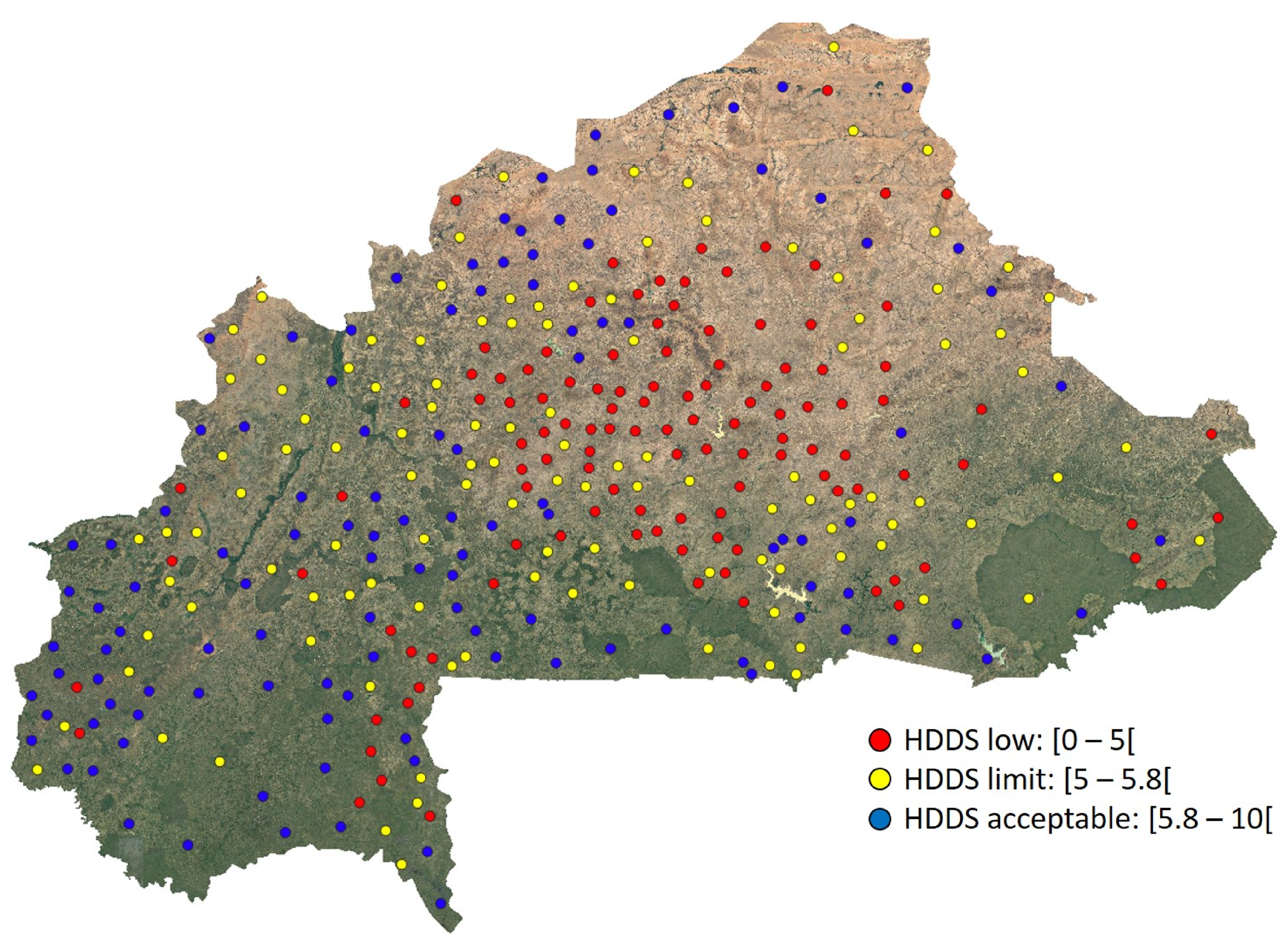

Figure 1:Spatial distribution of HDDS on average in 344 communes of Burkina Faso calculated on the 2009–2018 period (background map: Google Maps)

We use publicly available **explanatory variables** from several domains and structures to predict FCS and HDDS. We classify them into **4 groups**:

Data group	Spatio-temporal scale	Variables
Time series	Vary per month and per commune	Vegetation index, rainfall, ...
Conjunctural data	Vary per year and per commune	Meteo, violent events, ...
Structural data	Vary per commune	Soil quality, waterways, ...
High spatial resolution data	Vary per pixel	Land use, population density

Table 1:The four groups of data

Methods

- **Treating heterogeneous data** requires the use of suitable **Machine Learning** (ML) techniques
- We use 4 types of ML models suited to each data group to predict FCS and HDDS (Figure 2)
- We integrate in our models two **deep learning** methods: Long Short Term Memory (**LSTM**) and Convolutional Neural Network (**CNN**)
- We randomly select 85% of the dataset for model learning and 15% for testing, by repeating this procedure 5 times and calculate the average performances

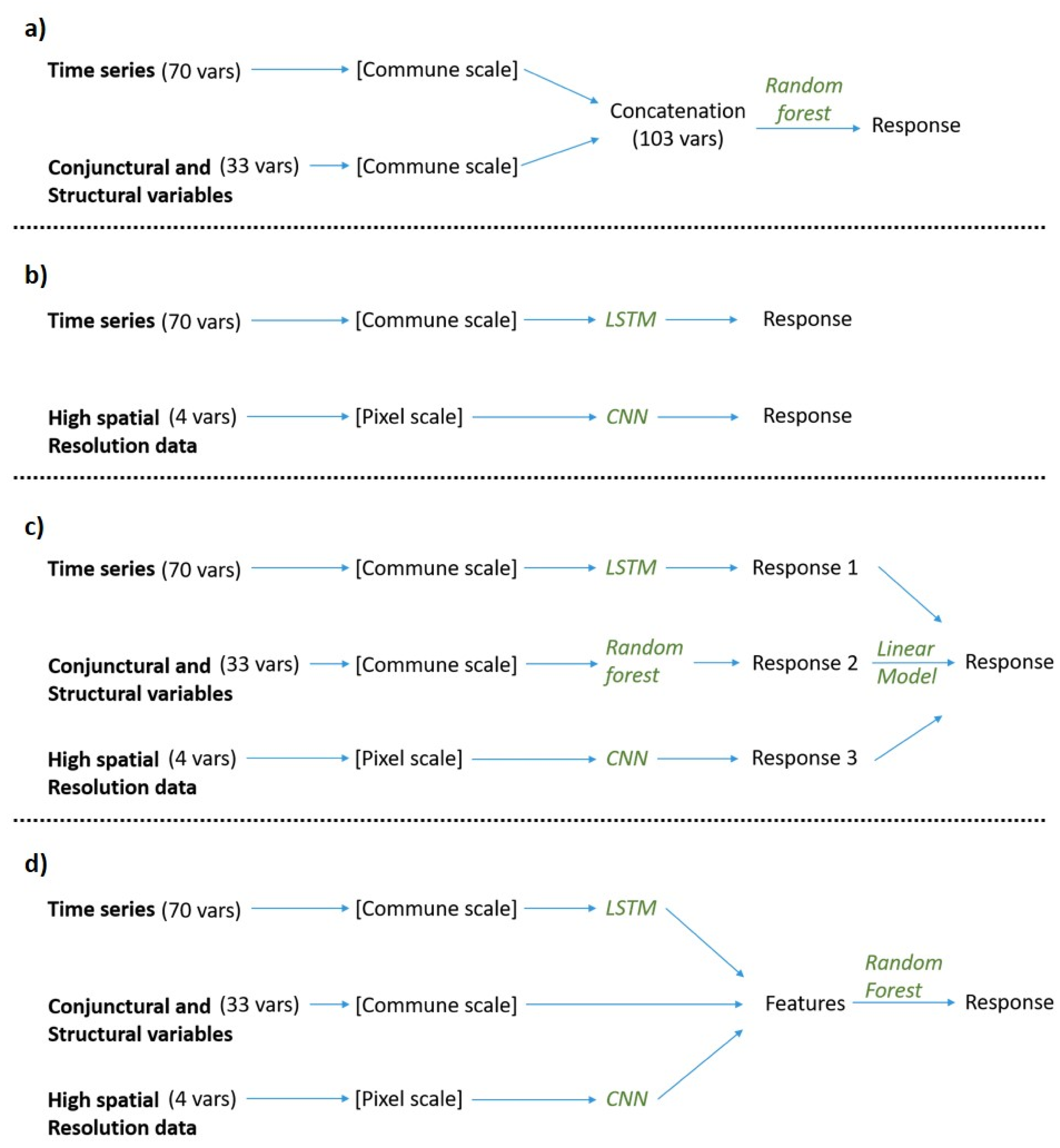

Figure 2:Architecture of the four machine learning models (a), (b), (c) and (d) used.

Results

Model	FCS	HDDS
Lentz & al study[1]	0.2	0.2
WFP study[2]	0.34	-
(a) Random forest on initial variables	0.339	0.326
(b) LSTM on time series	0.232	0.223
(b) CNN on high spatial resolution data	0.34	0.392
(c) Linear model on responses	0.375	0.426
(d) Random forest on features	**0.455**	**0.43**

Table 2:Performance (R^2) of the 4 types of models (a), (b), (c) and (d) and other works for FCS and HDDS prediction.

Key findings:

- The proposed approach **outperforms** competing methods (Table 2)
- **Type (d)** model which aggregate others models by feature fusion gives the best performance
- The most important variables come from **multiple domains** (population dynamics, soil quality, vegetation quality, meteo, etc.) which confirms the need to link food security with a large spectrum of data sources (Table 3)

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

Table 3:Top 5 ranks of variables according to their permutation importance for FCS and HDDS.

Future work:

- Improve our framework with the use of **textual data** (e.g., social networks, newspapers)
- **Refine models interpretability** (to bypass the black box effect of deep learning)
- **Develop an operational tool** to assist Food Security Monitoring Systems

Acknowledgment

This study was conducted with the help of Ministry of Agriculture and Hydro-agricultural Developments of Burkina Faso and Yves-Gerard Bazié which provided data. This work was supported by the French National Research Agency under the Investments for the Future Program #DigitAg, referred as ANR-16-CONV-0004.

References

[1] E. Lentz, H. Michelson, K. Baylis, and Y. Zhou, ``A data-driven approach improves food insecurity crisis prediction,'' *World Development*, 2019.

[2] World food programme, ``Humanitarian high resolution mapping; complementing assessments with remote sensing open data.'' https://wfp-vam.github.io/HRM/, 2019.

Email:hugo.deleglise@cirad.fr